



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**

**MODELO DE PRONÓSTICO DE VENTAS PARA POTENCIALES LOCALES
DE UNA CADENA DE MEJORAMIENTO DEL HOGAR**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

CAROLINA ELISA ÁNGELA MERINO VEYL

PROFESOR GUÍA:
RENÉ ESQUIVEL CABRERA

MIEMBROS DE LA COMISIÓN:
JUANITA GANA QUIROZ
LUIS ABURTO LAFOURCADE

SANTIAGO DE CHILE
2015

RESUMEN DE LA MEMORIA PARA
OPTAR AL TÍTULO DE:
Ingeniera Civil Industrial
POR: Carolina Merino Veyl
FECHA: 23/01/2015
PROFESOR GUÍA: René Esquivel C.

MODELO DE PRONÓSTICO DE VENTAS PARA POTENCIALES LOCALES DE UNA CADENA DE MEJORAMIENTO DEL HOGAR

En los últimos años, el mercado chileno de mejoramiento del hogar se ha desarrollado con gran rapidez en comparación al continente, generando una alta competencia entre las empresas participantes. De acuerdo a esta situación, es esencial que las compañías cuenten herramientas que apoyen su estrategia de expansión, como métodos que le permitan pronosticar las ventas de locales potenciales, donde resulta fundamental que aprovechen las oportunidades que se les presentan tomando decisiones acertadas.

De este modo, este trabajo tiene como objetivo crear un modelo de pronóstico de ventas, para apoyar la decisión de apertura de nuevas tiendas de una cadena de mejoramiento del hogar a lo largo de Chile.

Para lograr lo planteado, se caracteriza el método que actualmente utiliza la empresa, se revisan modelos y estudios vinculados con la descripción y estimación de demanda en el comercio minorista. En la actualidad existen varios modelos de estimación de ventas, que reúnen diversas características de los productos, mercado y población.

Luego, se seleccionan las variables que podrían influir en la predicción y que se considerarán. Para estimar los modelos, se consideran las variables de *superficie del local, distancia a la competencia, ubicación compartida, sector en que se ubican los locales, población en las zonas de influencia, caracterización de los hogares, barreras físicas y PIB*. Adicionalmente, se determina su forma funcional con las ventas de los locales, usando gráficos de dispersión y análisis de regresión.

Se estiman modelos mediante regresiones, utilizando el método de mínimos cuadrados y LASSO, considerando el valor promedio de las variables seleccionadas y los datos desagregados por local y año. Se compara el desempeño de estimación de los modelos realizados mediante los indicadores R^2 , AIC, BIC y MAPE.

El modelo seleccionado se estima con el método de mínimos cuadrados, tiene un coeficiente de determinación, R^2 , de 73% y un error de pronóstico del 14%. El ajuste de este modelo podría ser mejorado en el tiempo con la recopilación de nuevos registros para las variables utilizadas.

AGRADECIMIENTOS

Agradezco a mi familia, por su apoyo, confianza y amor incondicional durante toda mi vida. Especialmente a mi madre, que ha dado todo por mí, y a Patricio, que me escuchó y aconsejó durante todo el desarrollo de esta memoria.

A mi pololo, David Thackeray, por su amor y comprensión en los momentos críticos de mi vida universitaria.

A mis amigos y compañeros, con los que he crecido y vivido inolvidables momentos durante los seis años. Culis, las mejores amigas y compañeras de tareas. Alina, madrina incondicional, y Jorge, por ser mis partners memoristas. Deborah, amiga y guía en la empresa, por los alegres momentos vividos y ayuda.

A mis profesores, por su conocimiento y consejos que me entregaron para la realización de este trabajo.

A todos quienes me acompañaron y ayudaron en el proceso de desarrollo de esta memoria y transcurso de mi carrera.

TABLA DE CONTENIDO

I. FUNDAMENTOS	1
1. INTRODUCCIÓN	1
2. DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN.....	2
3. OBJETIVOS	4
3.1. OBJETIVO GENERAL	4
3.2. OBJETIVOS ESPECÍFICOS	4
4. ALCANCES	4
5. RESULTADOS ESPERADOS.....	4
6. MARCO CONCEPTUAL	5
7. METODOLOGÍA.....	8
II. DESARROLLO	17
1. ESPECIFICACIÓN DEL MODELO	17
1.1. DESCRIPCIÓN DE LAS VARIABLES	17
1.2. INFLUENCIA DE LAS VARIABLES PREDICTORAS.....	22
2. ESTIMACIÓN DEL MODELO.....	29
2.1. ESTIMACIÓN CON PROMEDIO DE DATOS ANUALES	30
2.2. ESTIMACIÓN CON DATOS DE LOCALES POR AÑO	37
3. EVALUACIÓN DE LA ESTIMACIÓN	50
4. EVALUACIÓN DE LA PREDICCIÓN.....	53
4.1. EVALUACIÓN DEL MODELO CON EJECUTIVOS DE LA EMPRESA	54
5. RECOMENDACIONES DE IMPLEMENTACIÓN DEL MODELO	55
III.CONCLUSIÓN Y COMENTARIOS FINALES.....	57
IV. BIBLIOGRAFÍA.....	59
V. ANEXOS	62
ANEXO 1: OTROS MODELOS DE ESTIMACIÓN DE VENTAS.....	62
ANEXO 2: SERIE IPC, ESTADÍSTICAS ECONÓMICAS DEL BANCO CENTRAL DE CHILE ...	67
ANEXO 3: ESTIMACIONES CURVILÍNEAS DE VENTA VS COMPA	67
ANEXO 4: ESTIMACIONES CURVILÍNEAS DE VENTA VS COMPB	68
ANEXO 5: ESTIMACIONES CURVILÍNEAS DE VENTA VS COMPE	68

ÍNDICE DE TABLAS

Tabla 1: Porcentaje de clientes por radio de distancia al local.....	19
Tabla 2: Quintiles de puntajes.....	20
Tabla 3: Tasas de consumo por GSE	21
Tabla 4: Mejores funciones para representar la distancia a las competencias	25
Tabla 5: Coeficientes del Modelo 1.....	31
Tabla 6: Coeficientes del modelo 3	32
Tabla 7: Ubicación compartida locales L01, L02, L04, L05, L11, L15, L18 y L25	33
Tabla 8: Coeficientes del modelo 5	35
Tabla 9: Resumen modelos estimados con promedio de datos anuales	37
Tabla 10: Coeficientes del modelo 6.....	38
Tabla 11: Coeficientes del modelo 7.....	39
Tabla 12: Coeficientes del modelo 8.....	40
Tabla 13: Características de ubicación de los locales L01, L02, L04, L05, L11, L15, L18 y L25	41
Tabla 14: Coeficientes del modelo 9.....	45
Tabla 15: Superficie de ventas de los locales L01, L02, L04, L05, L11, L15, L18 y L25	45
Tabla 16: Coeficientes del Modelo 10	47
Tabla 17: Resumen de los modelos estimados.....	50
Tabla 18: Indicadores de ajuste de los modelos.....	51
Tabla 19: Coeficientes modelo seleccionado	52
Tabla 20: Errores de predicción año 2013.....	54

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Promedio de ventas anuales 2006-2013 vs superficie del local	23
Ilustración 2: <i>Promedio de ventas anuales 2006-2013 vs distancia a compA</i>	24
Ilustración 3: Promedio de ventas anuales 2006-2013 vs distancia a compB	24
Ilustración 4: Promedio de ventas anuales 2006-2013 vs distancia a compE	24
Ilustración 5: Promedio venta anual 2006-2013 vs ubicación compartida...	26
Ilustración 6: Promedio de ventas 2006-2013 vs Sector	27
Ilustración 7: Promedio de ventas 2006-2013 vs Población	28
Ilustración 8: Promedio de ventas 2006-2013 vs Puntaje CSE ponderado	28
Ilustración 9: Variación PIB y Ventas promedio de locales con respecto al año anterior	29
Ilustración 10: Ventas vs Distancia a la competencia A (modelo 3)	34
Ilustración 11: Ventas vs Superficie (modelo 5)	36
Ilustración 12: Ventas vs Distancia a la competencia E (modelo 5)	36
Ilustración 13: Ventas vs Superficie (modelo 8)	41
Ilustración 14: Ventas vs Distancia a la competencia A (modelo 8)	42
Ilustración 15: Ventas vs Distancia a otros locales de la empresa (modelo 8)	43
Ilustración 16: Ventas vs Población a 5 Km de los locales (modelo 8).....	44
Ilustración 17: Ventas vs Distancia a otros locales de la empresa (modelo 9)	46
Ilustración 18: Ventas vs Distancia a la competencia A (modelo 10)	48
Ilustración 19: Venta vs Población a 5 Km (modelo 10)	49
Ilustración 20: Ventas pronosticadas vs Ventas reales 2006-2012	53

I. FUNDAMENTOS

1. INTRODUCCIÓN

En los inicios del retail en Chile, la industria contaba solamente con un par de tiendas que se encontraban en los sectores más céntricos de Santiago. La construcción del primer hipermercado en 1976 y el surgimiento de los primeros centros comerciales en los '80 marcaron un hito importante en su expansión.

El aumento progresivo de las utilidades permitió que las cadenas de retail pudieran expandirse más tarde por todo Santiago y llegar a lo largo de todo el país. Las cadenas comenzaron a diversificar su negocio, compitiendo en distintos mercados como: tiendas por departamento, supermercados en distintos formatos, sector inmobiliario, financiero, farmacias y mejoramiento del hogar, entre otros¹.

Chile ha alcanzado a ser el país líder dentro de Latinoamérica en la industria de mejoramiento del hogar, y ha sido el que se ha desarrollado con más rapidez. A pesar de que este mercado es considerado altamente fragmentado por los grandes operadores y ferreterías más pequeñas, su desarrollo se puede ver reflejado con una gran cantidad de puntos de venta, alcanzando 420² tiendas en el territorio nacional, y un volumen de venta de USD\$ 7.5 billones anuales [1]. En esta industria se pueden encontrar tiendas de grandes superficies pertenecientes a Sodimac, Easy y Construmart, asociaciones ferreteras como MTS y Chilemat, y otras ferreterías³.

Como es un mercado altamente competitivo, la evaluación de la ubicación de un nuevo local de una empresa es fundamental, ya que corresponde a un factor primordial en la elección de tiendas por parte del consumidor [2]. Jones menciona que las tres cosas más importantes para una tienda de retail son "ubicación, ubicación, ubicación" [3], por lo que la selección de una ubicación específica para la apertura de un nuevo local es una decisión estratégica fundamental, que puede ayudar a desarrollar una ventaja estratégica sostenible [2]. Además, para la empresa, la decisión de localización de sus tiendas envuelve una gran y casi irreversible cantidad de capital de inversión que, dado el comportamiento del consumidor, determina el área de influencia de cada tienda [4]. Durante los últimos 5 años, se han inaugurado 49⁴ tiendas correspondientes a las empresas Sodimac, Easy y Construmart, donde se calcula una inversión inicial de \$150.000 millones de pesos⁵ (2013).

¹ Ceret, Centro de Estudio de Retail. Santiago de Chile.

² Ubicaciones pertenecientes a tiendas de grandes superficies y asociaciones ferreteras a Agosto de 2014.

³ LAP Consultores Limitada, 2012.

⁴ Publicaciones de cada empresa en Press Release.

⁵ Dato entregado directamente por una de las empresas de mejoramiento del hogar.

A lo largo de las últimas décadas se ha tratado de identificar los aspectos que influyen en las ventas que las tiendas pueden alcanzar en las diferentes localidades, tales como las características de los potenciales clientes, de la propia tienda, condiciones económicas del sector y competencia [4], para escoger la ubicación que tenga asociada el mayor monto de ventas e inaugurar un nuevo local.

La empresa de mejoramiento del hogar para la que se realiza esta memoria, cuenta con 33 locales⁶ presentes en 10 regiones de Chile. Durante los últimos diez años, la empresa ha recibido diferentes ofertas de potenciales lugares en los que podría abrir nuevas tiendas, donde su decisión de instalación se ha basado en la demanda que podrían enfrentar en dicha ubicación, alcanzando una tasa de apertura de dos tiendas al año.

El proceso de evaluación de apertura de un nuevo local comienza con la propuesta de la división de inmobiliaria del conglomerado, la que define la ubicación y tamaño que tendría la nueva tienda. Con dicha información, el área de operaciones de la compañía realiza una visita al sector y analiza sus características, como accesibilidad, flujo de consumidores, acceso a las bodegas, entre otros factores.

Con una aprobación del área de operaciones, el jefe de control de gestión regional realiza una estimación de ventas para el posible local. Para realizarla, considera el gasto de los hogares de la zona (de acuerdo a la última Encuesta de Presupuesto Familiar), inversión en infraestructura y construcción de nuevas viviendas, definiendo de acuerdo a su experiencia y la situación del sector un potencial de mercado que podría alcanzar.

Si el jefe de control de gestión regional considera que hay un potencial de mercado atractivo, realiza un análisis de posible canibalización entre locales de acuerdo a la información que obtiene de los clientes fidelizados. Para finalizar el proceso, el formato que tendrá el nuevo local, de acuerdo a la estimación realizada, es definido junto a las gerencias comerciales de la compañía.

2. DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN

La división de inmobiliaria del conglomerado que apoya el presente proyecto, genera propuestas de ubicaciones para la apertura de nuevos locales de mejoramiento del hogar en Chile, que son enviadas al área de control de gestión de dicha unidad de negocio, donde deben responder rápidamente si desean iniciar un nuevo proyecto de expansión.

⁶ Número de locales a Enero 2015.

Según el jefe de control de gestión regional, en promedio mantienen un pool de 10 ubicaciones disponibles y pendientes de una respuesta, que por lo general no alcanzan a definir, ya que no poseen un método establecido para el pronóstico de ventas que podría alcanzar la nueva tienda. El proceso de estimación que ha realizado durante el último tiempo le consume una cantidad importante de tiempo, ya que debe pensar cada vez que evalúa un proyecto en los aspectos a considerar, y frecuentemente tiene destinado todo su tiempo a otros trabajos.

De acuerdo a la experiencia del jefe de control de gestión y al Gerente del área, un 60% de las propuestas que reciben se implementarían si ellos pudieran evaluar dichos proyectos. De esta forma, se está perdiendo la oportunidad de abrir 6 locales anuales⁷. Podríamos estimar que si estos locales se inauguraran y obtuvieran ventas iguales al promedio de la cadena, de \$13.136 millones de pesos anuales por local⁸ (2013), se deja de generar \$78.818 millones de pesos anuales aproximadamente. Este coste de oportunidad correspondería al 18,8% de las actuales ventas de la compañía en Chile.

Además, los últimos pronósticos realizados no han resultado efectivos, debido a que se han obtenido unas diferencias medias en los ingresos esperados de un 35%, alcanzando el 50%. En un caso particular, las ventas de un local fueron tan bajas que imposibilitaron su sustentabilidad, y fue necesario tomar la decisión de cerrarla. La apertura de esa tienda había involucrado una inversión inicial de \$2.750 millones de pesos. Esta situación se pudo haber previsto si se hubiese realizado una estimación de ventas confiable, ya que dicha tienda mantenía los indicadores de funcionamiento y calidad promedios de la cadena.

Esto ha causado la desconfianza en las proyecciones de ingresos que percibirán los locales que se abrirán prontamente, las que podrían variar en \$4.600 millones de pesos si se mantiene el error de estimación de 35% mencionado anteriormente. El escepticismo respecto de los pronósticos calculados ocasiona que los involucrados en el proceso de evaluación y decisión de apertura de locales no se quieran aventurar y arriesgar a tomar una resolución.

Actualmente, la empresa no cuenta con un método formal y definido de estimación de ventas para nuevos locales, donde el procedimiento que se utiliza para realizar el pronóstico queda a criterio del encargado de la evaluación. De acuerdo a esto, se obtiene una estimación que carece de objetividad y seguridad, generando incertidumbre y afectando nocivamente la estabilidad y sistematicidad del proceso de pronóstico.

⁷ Se considera el 60% del pool de ubicaciones que mantienen.

⁸ Memoria Anual Cencosud, 2013. Santiago, Chile.

Como la venta minorista que generan los locales de la cadena corresponde al 94% del total⁹, se desea proponer un modelo formal para el pronóstico de venta minorista para apoyar de modo oportuno la decisión de apertura de nuevos locales en las ubicaciones que ofrece la división inmobiliaria del conglomerado.

3. OBJETIVOS

3.1. Objetivo General

Modelar la venta minorista para potenciales tiendas de una cadena de mejoramiento del hogar, para facilitar el proceso de decisión sobre propuestas de apertura y disminuir la diferencia entre valores reales y estimados.

3.2. Objetivos Específicos

- Determinar las variables que influyen en el pronóstico de ventas del público minorista.
- Definir un modelo para el pronóstico de ventas que enfrenta la empresa.
- Generar un modelo que pueda ser utilizado en todas las regiones de Chile que le interesen a la cadena de mejoramiento del hogar.

4. ALCANCES

- El modelo se adapta a una empresa de mejoramiento del hogar en Chile, que entrega los datos de sus locales.
- El modelo será validado para su uso en todas las regiones de Chile, exceptuando la XII Región de Magallanes y de la Antártica chilena.
- La utilización del modelo creado en la decisión de apertura de una próxima tienda no está incluida en el proyecto.

5. RESULTADOS ESPERADOS

Con el desarrollo del trabajo de memoria se entregan los siguientes resultados:

- Los principales factores que influyen en las ventas de los locales de la cadena involucrada en el análisis.

⁹ Total de ventas reúne la venta minorista y mayorista.

- Modelo de pronóstico de demanda para venta retail de una potencial ubicación en Chile.
- Base de datos necesaria para utilizar el modelo en las diferentes regiones de Chile y sugerencias para su uso.

6. MARCO CONCEPTUAL

En esta sección se presentan los aspectos bibliográficos relevantes para el desarrollo del trabajo, describiendo los factores influyentes en las ventas de un comercio minorista.

Las ventas de los establecimientos minoristas han sido foco de estudio en todo el mundo y en los diferentes mercados de bienes. Entre los aspectos que han sido estudiados y pueden ser utilizados en la evaluación de potenciales locales destacan:

- a. *Distancia al establecimiento*: la distancia que separa a los clientes de los centros comerciales que se consideran en el estudio ha sido medida en diferentes unidades, como en *kilómetros* [11],[12], en *tiempo de conducción*¹⁰, *distancia perceptual* de los consumidores en minutos y en bloques de edificios [15].

Debido a este planteamiento, surge el uso de isócronas en el retail, que determina la distancia temporal que existe entre la tienda y su cliente objetivo, determinando su área de influencia y la densidad de distribución de los consumidores en la zona[20]. A través del uso de esta herramienta se puede calcular el área primaria de comercialización, donde generalmente provee entre el 45% y 75% de los clientes[21].

- b. *Competencia*: se ha considerado que locales de la competencia[21],[22], que se encuentran en la zona a evaluar, son un factor influyente en la probabilidad de que los consumidores compren en un determinado lugar debido a que se dividen el potencial de mercado del sector en que se encuentran.
- c. *Población*: un factor primordial al momento de evaluar el potencial de ventas que puede generar una tienda es la población y sus características[22], como su número, composición, densidad, crecimiento, hábitos de compra, ingresos y gastos[21]. Se debe mantener en cuenta que el crecimiento de la población en una zona puede atraer el desarrollo de autopistas o de la competencia en el sector.

¹⁰ [13], [14], [15], [16], [17], [18], [19]

Se ha estudiado que las *características demográficas*[31] de los consumidores influyen en la elección del punto de venta, en las que se han tenido en cuenta el género, edad, tipo de hogar, ingreso o salario[32]. Estas características podrían influir en los elementos que consideran relevantes cuando las personas configuran la *imagen* que tienen de una tienda[33].

- d. *Flujo de transeúntes*[42]: se considera que las personas que circulan por la zona en que se encuentra una tienda, aumentan la demanda que ésta percibe. La estimación del flujo de personas se puede realizar a partir de la presencia de diversos puntos de atracción ubicados en el sector, como locales comerciales, empresas, entre otros; o bien, puede ser estimada mediante estudios de mercado.
- e. *Atracción*: se han considerado diferentes variables para representar la atracción que ejerce una tienda en las personas, donde la más utilizada y validada por distintos autores corresponde a la *superficie*¹¹ de la tienda como metros cuadrados, puesto que a través de esta medida queda bien representada su oferta.

Otra variable usada es el *surtido*[13] o variedad de productos que se ofrecen, sin embargo resulta más difícil de medir que la superficie para diversos locales.

También se puede usar la *imagen*¹² del establecimiento, donde se considera la percepción que tienen los consumidores con respecto a la empresa o local, donde es fundamental determinar cuáles son los atributos esenciales que afectan a la imagen para los consumidores. Se propone que aspectos que afectan a la imagen de una tienda corresponden a la disponibilidad y atención por parte de los vendedores, precio de los productos o servicios, disponibilidad de crédito[21],[13], eventos especiales, valor recreativo[11] y publicidad[29]. Es esencial que los consumidores estén familiarizados con el establecimiento considerado para obtener una evaluación de éste[30].

La *accesibilidad*[12] de los clientes hacia la tienda también podría influir en la atracción hacia los consumidores, donde se deben considerar los ríos, vías de ferrocarriles, autopistas o carreteras cercanas a la ubicación que podrían dificultar la facilidad de acceso. La accesibilidad a la tienda puede verse reflejada en las isócronas considerando el tiempo que les requeriría a los habitantes de la zona en llegar a la ubicación que se analiza[21].

- f. *Ubicación compartida*[21]: se considera que si solo una tienda comparte con otras el sitio donde se ubica, el conjunto podría generar mayor poder

¹¹ [11], [14], [19], [23], [24], [25], [26]

¹² [27], [12], [15], [14], [28], [17], [19]

de atracción de clientes que una de aquellas tiendas aisladas[21],[22]. La localización dentro de un centro comercial representa este efecto [43]. Este poder puede variar dependiendo de la relación que haya entre los mercados y tipos de clientes objetivos de los locales.

- g. *Equilibrio económico*[21]: generalmente la estabilidad económica de una zona puede ser evaluada considerando su nivel de producción, renta y/o empleo, aunque obtener un pronóstico preciso es improbable.

El uso de indicadores económicos del área, como la producción industrial, puede ser útil para la medición de este factor. Se debe tener en cuenta que siempre es deseable contrastar esta variable con los hábitos de compra y calidad de tiendas disponibles en el sector.

La hipótesis de que los aspectos recién presentados influyen en el nivel de ventas que puede alcanzar una tienda minorista ha sido probada por diversos autores, donde se han generado varios modelos que permiten incluirlas y estimar las ventas de un local. Entre los modelos estudiados que permiten estimar la demanda de nuevas tiendas se encuentra la teoría de los lugares centrales, modelo de Huff, modelo multiplicativo de interacción competitiva, modelos logit, enfoque analógico y modelos de regresión.

La *teoría de los lugares centrales* postula que los clientes se dirigen a la tienda más cercana que posea los bienes deseados para realizar su compra. El *modelo de Huff* se basa en la premisa de que la probabilidad en la que un consumidor elegirá comprar en una tienda en particular se incrementa cuando el local incrementa su tamaño, y disminuye a medida que la distancia o tiempo de viaje que demora el consumidor en llegar a dicha tienda aumenta. El *modelo Multiplicativo de Interacción Competitiva* permite considerar una mayor cantidad de variables con el objetivo de alcanzar una mayor explicación en la atracción del modelo simple de Huff. En su formulación se pueden integrar todas las variables deseadas, definiendo la probabilidad de compra en un establecimiento como un cociente de utilidades, compuesta por los atributos de la tienda.

Los *modelos Logit* establecen que la probabilidad de que un individuo ubicado en un lugar determinado asista a un establecimiento definido puede ser expresado en función de su utilidad observable, que depende de las propiedades de la tienda como de las características del individuo. Cuando se utiliza el *enfoque analógico*, se describen las características de la tienda de los locales más exitosos, tratando de localizar uno nuevo con elementos similares. Los *modelos de regresión* describen la relación entre una variable dependiente, como las ventas de una tienda, y una o más variables explicativas o predictoras, como los componentes de atracción comercial mencionados en el punto anterior. La dependencia funcional de la variable dependiente respecto de las explicativas, pudiendo ser lineales o no lineales.

Los modelos mencionados¹³ permiten pronosticar las ventas de una nueva tienda considerando, en menor o mayor medida, los componentes de atracción comercial explicados.

Como se desea obtener un modelo con buen poder predictivo, pero manteniendo de manera sencilla la interpretación del efecto de sus variables, se podría optar por un modelo de regresión o modelo multiplicativo de interacción competitiva. Sin embargo, para la elección del modelo se debe tener en cuenta los datos que se tienen disponibles para la calibración de éste, donde solamente se posee información pública y privada, perteneciente a la empresa que apoya el proyecto.

Para el desarrollo del trabajo se considera utilizar el *Modelo de regresión*¹⁴, ya que:

- Permite la integración y evaluación de todos los componentes de atracción comercial presentados.
- El modelo se puede adaptar a las relaciones funcionales de las variables predictoras y dependiente.
- Posibilita equilibrar su capacidad explicativa y predictiva, minimizando el error de estimación y manteniendo la interpretabilidad del efecto de las variables.

7. METODOLOGÍA

La metodología a seguir durante el desarrollo del proyecto corresponde a la siguiente:

- a. Caracterización del método de pronóstico que actualmente se utiliza en una cadena de mejoramiento del hogar.

Se obtiene la información de la situación actual de la empresa mediante reuniones con los jefes de control de gestión regional, de Chile y Gerente del área, actores involucrados en el proceso y con experiencia en el mercado, donde describen el procedimiento actual, sus inquietudes y creencias con respecto al pronóstico de ventas de los locales.

- b. Revisión bibliográfica relacionada con métodos de estimación de demanda y de atributos que la podrían afectar.

Se buscan todos los modelos relacionados con los pronósticos de demanda, en particular, los utilizados en estimaciones de ventas a nivel agregado para

¹³ Para revisar en mayor detalle los modelos ver "Modelos de estimación" en los Anexos.

¹⁴ Una descripción más detallada de los modelos de regresión puede encontrarse en la sección 7.d., página 10.

nuevas tiendas, junto a los componentes que influyen en la atracción comercial de los establecimientos. Se considera la teoría de los lugares centrales, modelo de Huff, modelo multiplicativo de interacción competitiva, modelos logit multinomial y anidado, enfoque analógico y modelos de regresión.

c. Especificación del modelo.

Se crea una base de datos que reúne los factores a considerar en el modelo, basados en la revisión bibliográfica. Se utilizan datos publicados por el INE, Banco Central de Chile y empresas de mejoramiento del hogar, y privados entregados por la cadena que apoya el trabajo. Los atributos a considerar son:

- Superficie de venta: se considera la superficie de venta de cada local medida en metros cuadrados. Esta información es entregada por la empresa.
- Distancia a la competencia: se calcula la distancia euclidiana entre los locales de la empresa en estudio y su competencia usando las coordenadas de la ubicación de cada uno de ellos.
- Ubicación compartida: se clasifica el tipo de ubicación en que se encuentra cada local de la compañía, de acuerdo a si está aislado, comparte ubicación con un supermercado, o se encuentra dentro de un portal o mall.
- Sector en que se encuentra el local: se clasifica el sector en que se encuentran los locales de acuerdo a su uso principal y función que desempeñan en la ciudad. Se considera la clasificación en zonas residencial, comercial, industrial y empresarial, y mixturas de ellas.
- Población en las zonas de influencia: se contabilizan las personas que se encuentran a un radio de influencia de 3 y 5 Km de cada local, correspondientes al 50% y 70% de las ventas fidelizadas¹⁵, utilizando información obtenida desde la empresa y del Instituto Nacional de Estadísticas chileno (INE).
- Caracterización de los hogares: mediante el uso de la metodología de clasificación socioeconómica de los hogares chilenos, desarrollada por el INE, y datos del Censo 2002 se cataloga a cada hogar de Chile.

¹⁵ La justificación de elección del radio de influencia se encuentra en la página 19.

- Barreras físicas: se determina si existe acceso desde carreteras o autopistas al local utilizando Google StreetView como herramienta de visualización.
- PIB nacional: se utiliza el indicador económico publicado por el Banco Central de Chile para caracterizar la condición económica presente en el país.

Se considera la utilización de clústeres de locales por surtido de productos que usa la empresa. Sin embargo, la información no es facilitada para el desarrollo de este trabajo.

Utilizando la base de datos y un programa estadístico, SPSS¹⁶, se realizan gráficos de dispersión y regresiones para definir la relación funcional de las variables independientes especificadas en el punto anterior y las ventas anuales de los locales.

d. Estimación del modelo.

Para estimar los coeficientes que relacionan las variables especificadas en el punto anterior con las ventas de los locales, se utiliza el modelo de regresión.

Los *modelos de regresión* describen la relación entre una variable dependiente, como las ventas de una tienda, y una o más variables explicativas o predictoras, como los componentes de atracción comercial mencionados en el punto anterior. Los modelos de regresión pueden ser de varios tipos, según el número de variables independientes y de la función. Si solamente hay una variable, corresponde a un modelo de regresión simple, y si el número de variables es mayor, se realiza un modelo de regresión múltiple [9]. Un factor a tener en cuenta en la clasificación del modelo es la dependencia funcional de la variable dependiente respecto de las explicativas, pudiendo ser lineales o no lineales.

Modelo de regresión lineal, en el que las variables implicadas deben ser cuantitativas o continuas. Puede ser simple o múltiple, donde este último es una extensión del modelo de regresión simple. El valor de la variable dependiente viene dado por:

$$Y_i = \beta_0 + \beta_1 * X_1 + \dots + \beta_n * X_n + \varepsilon$$

Donde,

X_0, X_1, \dots, X_n son las variables explicativas o independientes,

$\beta_0, \beta_1, \dots, \beta_n$ son los parámetros que miden la influencia de las variables explicativas sobre Y ,

¹⁶ Software estadístico que permite realizar regresiones.

ε corresponde al error que recoge las diferencias encontradas entre la relación lineal y el valor de la variable, indicando la variabilidad de la variable dependiente debida a causas no controladas por el modelo.

En una formulación común mediante el modelo lineal, se asume simplemente que la participación de mercado o nivel de ventas de una empresa, se puede representar con una función lineal de las variables correspondientes al marketing mix y otras variables relevantes [39].

La asociación entre una variable dependiente y "n" variables independientes puede ser fuerte o débil. Para determinar la fuerza se utilizan parámetros de medición de la asociación, donde los más utilizados son el coeficiente de correlación de Pearson r y el coeficiente de determinación R^2 [9].

El *modelo de regresión no lineal* surge de la función no lineal entre las variables independientes y la dependiente, se puede expresar como

$$y = f(x, \theta) + \varepsilon$$

Donde,

$f(x, \theta)$ es una función no lineal respecto a algunos parámetros desconocidos θ ,

ε corresponde al error que recoge las diferencias encontradas entre la relación lineal y el valor de la variable.

Considerar la función no lineal permite "estimar modelos con relaciones arbitrarias entre las variables independientes y dependiente"¹⁷, donde no sólo se puede descubrir la medida en que las variables predictoras explican a la dependiente, sino que también, obtener predicciones de la dependiente a partir de las variables explicativas.

El modelo a evaluar puede ser ajustado al conjunto de datos a través de diferentes métodos, donde el más popular corresponde al *método de mínimos cuadrados*[34]. En este enfoque se eligen los coeficientes β que minimizan la suma de cuadrados residuales, representadas por la siguiente métrica:

$$RSS(\beta) = \sum_{i=1}^N (y_i - x_i^T * \beta)^2$$

Donde,

$RSS(\beta)$ es la suma de cuadrados residuales,

y_i corresponde al *i-ésimo* valor de la variable dependiente,

¹⁷ IBM, "IBM SPSS Regression 19".

x_i es el i -ésimo valor de la variable explicativa,
 β corresponde al parámetro que mide la influencia de la variable independiente.

Como $RSS(\beta)$ es una función cuadrática en sus parámetros, siempre posee un mínimo.

Existen diferentes métodos para la selección de variables del modelo. Los tres más comunes[38] son:

- Selección hacia adelante: se comienza con un modelo que contiene el intercepto sin predictores, y se agrega aquella variable que posee la menor suma de cuadrados residuales. Luego se agrega la variable que resulte con menor RSS para el nuevo modelo con dos variables. Este método continúa hasta que es detenido cuando una regla es satisfecha, como cuando no se genera una disminución en la RSS .
- Selección hacia atrás: se comienza con todas las variables en el modelo, y se remueve aquella variable estadísticamente menos significativa, con mayor p -valor. Este procedimiento continúa hasta que se alcanza la regla de detención, como cuando todas las variables poseen un p -valor menor a un umbral.
- Pasos sucesivos: es una combinación de los métodos de selección hacia adelante y hacia atrás. Se comienza con un modelo sin variables, y con la selección hacia adelante, se añade la que provea el mejor ajuste. Se continúan agregando nuevas variables hasta que alguna de ellas supere el umbral en su p -valor, que es removida. Se continúa con este método hasta que todas las variables en el modelo tengan un p -valor bajo, y todas las variables descartadas del modelo posean un p -valor superior al umbral si son añadidas a éste.

El modelo de regresión también puede ajustado utilizando el método LASSO [38], que estima los coeficientes β_λ , minimizando la siguiente métrica:

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j * x_{ij})^2 + \lambda * \sum_{j=1}^p |\beta_j| = RSS + \lambda * \sum_{j=1}^p |\beta_j|$$

Donde,

RSS es la suma de cuadrados residuales,

y_i corresponde al i -ésimo valor de la variable dependiente,

x_i es el i -ésimo valor de la variable explicativa,

β_j corresponde al parámetro que mide la influencia de la variable independiente j ,

λ es el parámetro de penalización de LASSO.

LASSO disminuye los coeficientes estimados hacia cero. Con el parámetro de penalización suficientemente grande los estima exactamente en cero, realizando una selección de las variables.

Para elegir el valor de λ , se realiza una cuadrícula con sus posibles valores, y se computa el error que se obtiene con validación cruzada, para cada valor de λ . Luego, se selecciona el λ con que se obtiene el mínimo error computado anteriormente.

El ajuste del modelo de regresión está representado por el coeficiente de determinación múltiple R^2 [35], que simboliza la calidad en que el modelo puede replicar los resultados, y la proporción de la varianza explicada por el mismo. Este coeficiente puede variar entre el rango $[0,1]$, en el que mientras más se acerque al valor 1, mejor será el poder explicativo que posee el modelo de regresión, y mientras más se acerque a 0, menor poder explicativo tendrá. Como este coeficiente no decrece cuando se van agregando un mayor número de variables al modelo, se puede utilizar la medida R^2 ajustado, que penaliza la mayor inclusión de variables y posee el mismo criterio de rango de ajuste mencionado.

En el caso de la estimación de demanda, se considera que los factores que afectan en las ventas de una tienda existente de una cadena tendrán el mismo impacto en la ubicación de nuevas tiendas [4].

Al crear el modelo se debe tener en cuenta el objetivo que se busca alcanzar; explicativo, predictivo o una combinación de ambas, para escoger el método de estimación de la función. Los modelos lineales permiten una interpretación simple, pero no siempre otorga predicciones tan precisas como otros modelos. En cambio, algunos de los modelos no lineales generan predicciones más exactas a expensas de una menor interpretabilidad [38].

Para estimar el modelo predictivo de ventas se utiliza el modelo de regresión, con el *método de mínimos cuadrados* con introducción de variables *pasos sucesivos*, porque puede ser utilizado si el número de variables a estimar supera el número de casos de la muestra, y al ir agregando nuevas variables, elimina aquéllas que pueden ser convertirse en redundantes en el proceso de ajuste. También se estima el modelo utilizando el método *LASSO*.

En una primera instancia se consideran las variables más utilizadas en otros modelos: superficie de venta del local, distancia a la competencia y población cercana al punto de venta. Posteriormente, se evalúa considerar nuevas variables en el modelo, de acuerdo a la estimación de los coeficientes resultantes. Utilizando la estimación de los coeficientes, se

interpretan los signos e influencia de las variables explicativas y se relacionan con la realidad.

Se escoge el nivel mínimo de significancia del 80%, ya que como se privilegia el pronóstico por sobre la explicación del fenómeno de ventas, se puede relajar la significación.

e. Evaluación de la estimación.

Se evalúa la significancia estadística de las variables del modelo, mediante su *p-valor* asociado.

Para comparar el desempeño de los modelos estimados en el procedimiento anterior [38] se utilizan los siguientes indicadores:

- *R² ajustado*, que simboliza la calidad en que el modelo puede replicar los resultados, y la proporción de la varianza explicada por el mismo. Este coeficiente puede variar entre el rango [0,1], en el que mientras más se acerque al valor 1, mejor será el poder explicativo que posee el modelo de regresión, y mientras más se acerque a 0, menor poder explicativo tendrá. Además, penaliza la mayor inclusión de variables en los modelos. Puede ser representado con la siguiente métrica:

$$R^2 = 1 - \frac{\sum_{j=1}^n (VtaReal_j - VtaEstim_j)}{\sum_{j=1}^n (VtaReal_j - \bar{VtaEstim})}$$

$$R^2 \text{ ajustado} = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1}$$

Donde,

$VtaReal_{i,t}$ representa la venta real asociada al local i en el año t ,

$VtaEstim_{i,t}$ es la estimación de ventas para el local i en el año t , obtenida con el modelo,

$n = \sum_{t=1}^7 m_t = 154$ es el número de casos utilizados para evaluar la estimación,

p representa el número de variables utilizadas en el modelo para estimar las ventas.

- *Criterio de información de Akaike (AIC)*, utilizado para medir la calidad relativa de un modelo estadístico para un conjunto de datos dados. Este indicador recompensa el ajuste que tiene un modelo a un conjunto de observaciones y penaliza el aumento en el número de parámetros utilizados. El modelo preferido será el que obtenga el mínimo valor en el indicador.

Puede ser calculado mediante la siguiente métrica:

$$AIC = n * \ln\left(\frac{RSS}{n}\right) + 2 * p,$$

Donde,

$n = \sum_{t=1}^7 m_t = 154$ es el número de casos utilizados para evaluar la estimación,

$RSS = \sum_{j=1}^n (VtaReal_j - VtaEstim_j)^2$ es la suma de cuadrados residuales, p representa el número de variables utilizadas en el modelo para estimar las ventas.

- El *Criterio de información bayesiano (BIC)* es similar al AIC, pero posee una mayor penalización a la inclusión de parámetros en los modelos. Puede ser calculado con la siguiente métrica:

$$BIC = n * \ln\left(\frac{RSS}{n}\right) + p * \ln(n)$$

Donde,

$n = \sum_{t=1}^7 m_t = 154$ es el número de casos utilizados para evaluar la estimación,

$RSS = \sum_{j=1}^n (VtaReal_j - VtaEstim_j)^2$ es la suma de cuadrados residuales, p representa el número de variables utilizadas en el modelo para estimar las ventas.

- *Error absoluto medio porcentual (MAPE)*, se usa para medir el error global de los modelos en términos porcentuales, lo que resulta un indicador fácil de interpretar. Se calcula con la siguiente métrica:

$$MAPE = \frac{100 * \sum_{t=1}^7 \sum_{i=1}^{m_t} \frac{|VtaReal_{i,t} - Pronóstico_{i,t}|}{VtaReal_{i,t}}}{n},$$

Donde,

$VtaReal_{i,t}$ representa la venta real asociada al local i en el año t ,

$Pronóstico_{i,t}$ es la estimación de ventas para el local i en el año t , obtenida con el modelo,

m_t corresponde al número de locales en el año t ,

$n = \sum_{t=1}^7 m_t = 154$ es el número de casos utilizados para evaluar la estimación.

Por lo tanto, se selecciona el modelo que posea un mejor ajuste, eso quiere decir, un menor MAPE, AIC y BIC, y mayor R^2 ajustado.

f. Evaluación de la predicción.

Para calcular la calidad de la predicción del modelo se estiman las ventas anuales de los locales de la cadena para el año 2013, y se comparan con las ventas reales realizadas ese mismo año, usando el error absoluto medio porcentual y error porcentual máximo del conjunto de locales.

II. DESARROLLO

1. ESPECIFICACIÓN DEL MODELO

En esta sección se describen las variables consideradas para la creación de la base de datos que se utiliza y su generación. Luego, para cada factor considerado, se explica la hipótesis de su influencia sobre las ventas y se determina la relación funcional con éstas.

1.1. Descripción de las variables

Para realizar el modelo se crea una base de datos que reúne la información anual de las ventas de cada local de la empresa de mejoramiento del hogar y de los siguientes componentes de atracción comercial:

- Superficie de venta: a cada local de la empresa se le asigna una superficie de venta en metros cuadrados. Esta información es entregada por la empresa que apoya el trabajo y se registra en la siguiente variable

Superf_i: Superficie en metros cuadrados del local *i*.

- Distancia a la competencia: se seleccionan los locales pertenecientes a las dos cadenas de tiendas de grandes superficies en Chile como competencia, puesto que la cadena pertenece a esta clasificación¹⁸. También, se consideran las tiendas que pertenecen a la cadena, permitiendo considerar la canibalización de ventas que podría existir.

Se busca la dirección de cada tienda en las páginas web de cada cadena y con su respectiva fecha de apertura. Esta dirección es ingresada en la página de Google Maps¹⁹, se verifica que la ubicación sea la correcta mediante la opción de Google StreetView y se registran las coordenadas geográficas de cada local.

Luego, para obtener la distancia euclidiana entre dos ubicaciones, se utiliza la siguiente fórmula²⁰ que aproxima esta distancia:

$$Distancia = R * \text{acos}(\text{sen}(\text{latit}_1 * \pi/180) * \text{sen}(\text{latit}_2 * \pi/180) + \text{cos}(\text{latit}_1 * \pi/180) * \text{cos}(\text{latit}_2 * \pi/180) * \text{cos}(\text{long}_1 * \pi/180 - \text{long}_2 * \pi/180))$$

Donde,

$R = 6372,795477598$, es el radio de la tierra en kilómetros.

$\text{latit}_{1,2}$ corresponde a la latitud de la ubicación 1 o 2.

¹⁸ LAP Consultores Limitada, 2012.

¹⁹ Página web de Google Maps: <https://maps.google.com>

²⁰ Mapanet, Geo Postal Codes Database: www.mapanet.eu/Resources/Script-Distance.htm, revisada en Agosto 2014.

$long_{1,2}$ es la longitud de la ubicación 1 o 2.

Se registra la menor distancia entre cada local de la empresa en estudio con su competencia en las siguientes variables:

$compA_{i,t}$: distancia entre el local i su competencia más cercana perteneciente a la cadena A , en el año t .

$compB_{i,t}$: distancia entre el local i su competencia más cercana perteneciente a la cadena B , en el año t .

$compE_{i,t}$: distancia entre el local i su competencia más cercana perteneciente a la cadena E , en el año t .

- Ubicación compartida: se clasifica el tipo de ubicación en que se encuentra cada local de la compañía, de acuerdo a si se encuentra dentro de un mall, de un portal o comparte el sitio con un supermercado. Esta información es entregada por la empresa y se registra en las siguientes variables:

$Mall_i = 1$ si el local i se encuentra dentro de un mall, 0 si no.

$Portal_i = 1$ si el local i se encuentra dentro de un portal, 0 si no.

$SupA_i = 1$ si el local i comparte su ubicación con el supermercado perteneciente a la cadena A , 0 si no.

$SupB_i = 1$ si el local i comparte su ubicación con el supermercado perteneciente a la cadena B , 0 si no.

$Solo_i = 1$ si el local i se encuentra solo en el sitio, 0 si no.

- Sector en que se encuentra el local: se clasifica el sector en que se encuentran los locales de acuerdo a su uso principal y función que desempeñan en la ciudad. Se pueden encontrar las siguientes zonas:
 - Habitacional: se caracteriza por su concentración de casas y edificios, que tienen como uso principal la residencia.
 - Comercial: zona dominada por la actividad de intercambio de bienes y servicios, a través de supermercados, farmacias, ferreterías, shopping centers, empresas de servicios, pequeños locales, entre otros.
 - Industrial: predominan los grandes centros de producción y distribución de productos, como fábricas, mataderos, vitivinícolas, entre otros.
 - Empresarial: sector que alberga una alta concentración de oficinas y empresas de servicios.

Se visualizan las zonas en que se encuentran los locales de la compañía utilizando Google StreetView y se clasifican en sectores, generando las siguientes variables:

$sRes_i=1$ si el local i se encuentra en una zona residencial, 0 si no,
 $sResCom_i=1$ si el local i se encuentra en una zona mixta residencial y comercial, 0 si no,
 $sResInd_i=1$ si el local i se encuentra en una zona mixta residencial e industrial, 0 si no,
 $sComEmp=1$ si el local i se encuentra en una zona mixta comercial y empresarial, 0 si no.

- Población del sector: la Gerencia de Clientes de la compañía cuenta con información de ubicación de los hogares de sus clientes fidelizados²¹, lo que permite obtener el porcentaje de clientes que viven dentro de distintas áreas circulares²². En la Tabla 1 se presenta el porcentaje de clientes que vive dentro de 10 áreas circulares, para distintos radios desde un local:

Tabla 1: Porcentaje de clientes por radio de distancia al local

% Vta 1.000 Metros 14,4%	% Vta 2.000 Metros 35,7%	% Vta 3.000 Metros 50,8%	% Vta 4.000 Metros 60,2%	% Vta 5.000 Metros 68,6%
% Vta 6.000 Metros 70,3%	% Vta 7.000 Metros 72,8%	% Vta 8.000 Metros 75,0%	% Vta 9.000 Metros 76,9%	% Vta 10.000 Metros 78,3%

Fuente: Elaboración propia.

Se obtiene que en promedio el 50% de sus clientes se ubican dentro del radio de 3 kilómetros de sus locales, y un 70% vive a 5 kilómetros. Estas áreas se consideran como las zonas de influencia primaria y secundaria[40], respectivamente, que se adecúa a la presentación realizada por Phillip Hordern²³ acerca de las isócronas en un retailer.

Luego de definir las zonas de influencia de los locales de la compañía, se utiliza la cartografía censal de Chile²⁴ visualizada en ArcMap²⁵, donde se ubican las tiendas de la cadena y se identifican las manzanas censales correspondientes a los radios de 3 y 5 kilómetros.

Con la obtención del código de las manzanas censales seleccionadas, se contabiliza la población que se vive en dichas manzanas según el Censo 2002. Esta población es ajustada según la tasa de crecimiento²⁶ anual de las comunas correspondientes. Los datos son guardados en las siguientes variables:

²¹ Se considera como clientes fidelizados a aquellos que dictan su RUT en caja al momento de realizar sus compras.

²² El área circular consiste en trazar un círculo en torno a la tienda, con un radio determinado.

²³ Phillip Hordern (2007), clase n°4, curso Administración de Retail, Diploma en Gestión de Retail, Escuela de Negocios de la Universidad Adolfo Ibáñez.

²⁴ Cartografía solicitada en el Instituto Nacional de Estadística chileno en Julio de 2014.

²⁵ Software sugerido por encargado de entregar la cartografía chilena del INE.

²⁶ Programa de proyecciones de la población, INE, Agosto de 2014.

$Pob3K_{i,t}$: número de personas que viven a 3 kilómetros de distancia del local i en el año t ,
 $Pob5K_{i,t}$: número de personas que viven a 5 kilómetros de distancia del local i en el año t .

- Caracterización de los hogares: para caracterizar a los hogares se utiliza la Metodología de Clasificación Socioeconómica de los Hogares Chilenos desarrollada por el INE, que "permite clasificar los hogares del Censo 2002 de acuerdo a un conjunto de variables socioeconómicas, como variables relacionadas con el hogar y la vivienda, relacionadas con la educación y ocupación del jefe de hogar, y otras que resulten de la composición de las anteriores"²⁷.

Utilizando la base de datos del Censo 2002, se le asigna un puntaje a cada hogar de acuerdo a la especificación en la metodología. Luego, se clasifican los hogares en *quintiles de puntajes*, donde se obtiene la siguiente división:

Tabla 2: Quintiles de puntajes

CSE Quintil	Puntaje	
	Mínimo	Máximo
1	5.606	6.966
2	6.966	7.205
3	7.205	7.376
4	7.376	7.563
5	7.563	8.270

Fuente: Elaboración propia

Se contabilizan los hogares de cada quintil de la clasificación socioeconómica (CSE) ubicados a 3 y 5 kilómetros de cada local de la cadena.

El número de hogares es ajustado de acuerdo al crecimiento de la población correspondiente al año en evaluación de acuerdo a las proyecciones publicadas por el INE para cada comuna, obteniendo las siguientes variables:

$cseN3K_{i,t}$: número de hogares pertenecientes a la CSE N a 3 kilómetros del local i en el año t .

$cseN5K_{i,t}$: número de hogares pertenecientes al CSE N a 5 kilómetros del local i en el año t .

²⁷ Metodología de clasificación socioeconómica de los hogares chilenos, Departamento de Metodología Estadística, Santiago, 19 de Diciembre de 2003.

Adicionalmente, para caracterizar la inversión en productos de mejoramiento del hogar que realizan los hogares chilenos se utilizan datos publicados en la VI y VII Encuesta de Presupuesto Familiar (EPF), donde se presenta el gasto que realiza un hogar en una diversa cantidad de productos y servicios.

Para obtener el monto asociado a productos que vende la cadena, se consideran todos los artículos que han vendido durante los últimos años. Se suma el gasto especificado en la EPF para los productos previamente seleccionados, obteniendo el gasto de los hogares pertenecientes a la Región Metropolitana y otro gasto asociado al resto de las regiones del país en este tipo de artículos. Este gasto es actualizado a precios de julio de 2014 según la variación del IPC²⁸.

Utilizando el consumo promedio de los clientes fidelizados por grupo socioeconómico (GSE) entregado por la empresa, se obtiene un ponderador del gasto asociado a cada agrupación, donde

CM : consumo promedio de los clientes fidelizados.

CM_N : gasto promedio de los clientes fidelizados pertenecientes al GSE N .

$TCM_N = \frac{CM_N}{CM}$: tasa de consumo del GSE N sobre el promedio de gasto de los clientes. Las tasas obtenidas se presentan en la Tabla 3: Tasas de consumo por GSE.

Tabla 3: Tasas de consumo por GSE

TCM ABC1	TCM C2	TCM C3	TCM D	TCM E
1,05	1,01	0,99	0,98	0,97

Fuente: Elaboración propia

El gasto en productos de mejoramiento del hogar por GSE puede ser calculado utilizando la siguiente expresión:

$GM_{N,t} = EPF * TCM_N$: gasto de los clientes del GSE N en productos de mejoramiento del hogar en el año t .

Es importante notar que los quintiles de clasificación socioeconómica (CSE) no son necesariamente equivalentes a los grupos socioeconómicos (GSE), ya que se utilizan diferentes metodologías para su clasificación. Sin embargo, se espera que exista una fuerte relación.

- Barreras físicas: para considerar la existencia de barreras físicas en el acceso a los locales se utiliza Google StreetView, revisando si existe

²⁸ Ver Anexo 2: Serie IPC, Estadísticas económicas del Banco Central de Chile, página 66.

algún obstáculo que dificulte el acceso a los puntos de venta, como la falta de acceso hacia la tienda desde una carretera o autopista.

Al analizar las ubicaciones en que se encuentran los locales de la empresa, no se identificaron barreras de acceso. Se solicitó información a la compañía para corroborar la observación, resultando válida. Debido a esto, este factor no aplica para el caso en estudio.

- PIB nacional: para caracterizar la situación económica en que se encuentra el país en un determinado año, se utiliza el PIB nacional²⁹ correspondiente a al período en evaluación. Este PIB es ajustado por el IPC a precio de julio de 2014, generando la siguiente variable:

PIB_t : PIB nacional correspondiente al año t .

- Venta anual: se solicita la información de las ventas anuales realizadas por cada local. La venta es ajustada por el IPC a julio de 2014, registrando el período 2006-2013 en la siguiente variable:

$Vta_{i,t}$: venta anual del local i en el año t , medida en millones de pesos.

Los valores de las variables descritas se registran por local para cada año del período 2006-2013, consolidando una base de datos de 183 registros.

1.2. Influencia de las variables predictoras

Utilizando la base de datos descrita en el punto anterior, se calcula el promedio de cada variable por local, para visualizar la posible relación funcional existente entre la variable dependiente, *Venta anual*, con las variables descritas anteriormente.

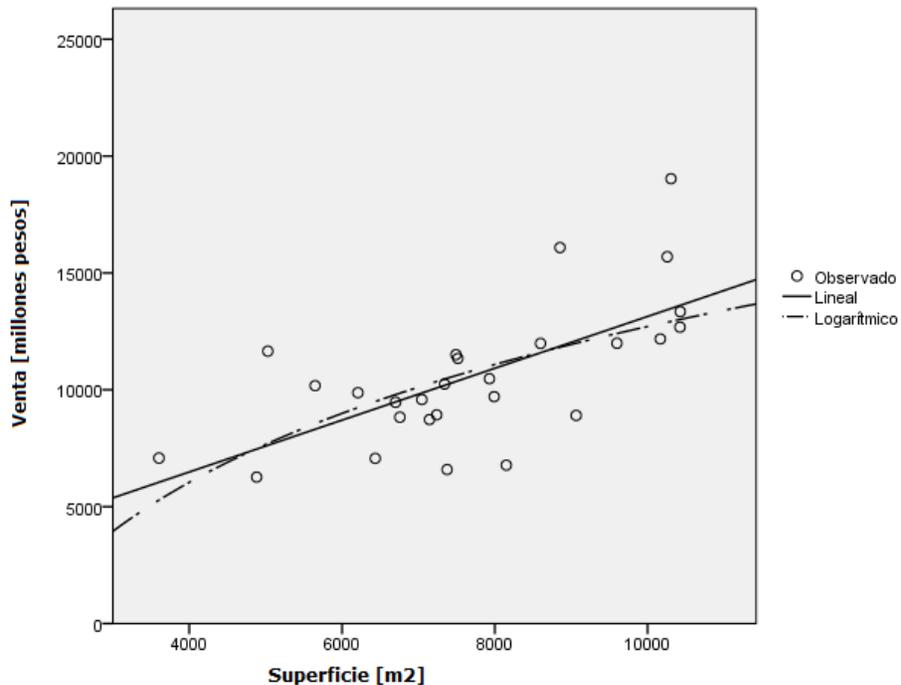
- Superficie de venta: como la superficie de venta de un local es utilizada para medir su atractivo, como representación de la variedad de productos ofrecidos, se esperaría que esta variable influyera positivamente en las ventas.

En la Ilustración 1 se presenta el gráfico de dispersión de ventas anuales versus superficie de venta para cada local. Se puede apreciar la relación positiva entre las variables mencionadas, donde a mayor superficie del local, se obtienen mayores ventas. Si bien se podría considerar que existe una relación lineal, resulta ilusorio pensar en que se podrían tener infinitas ventas si se define una superficie considerablemente mayor a las actuales, ya que debería existir un punto en que se alcance el

²⁹ PIB nacional publicado en las estadísticas económicas del Banco Central de Chile.

potencial de ventas en la zona. De acuerdo a esto, se considera una relación *logarítmica* entre las ventas y la superficie.

Ilustración 1: Promedio de ventas anuales 2006-2013 vs superficie del local



Fuente: Elaboración propia.

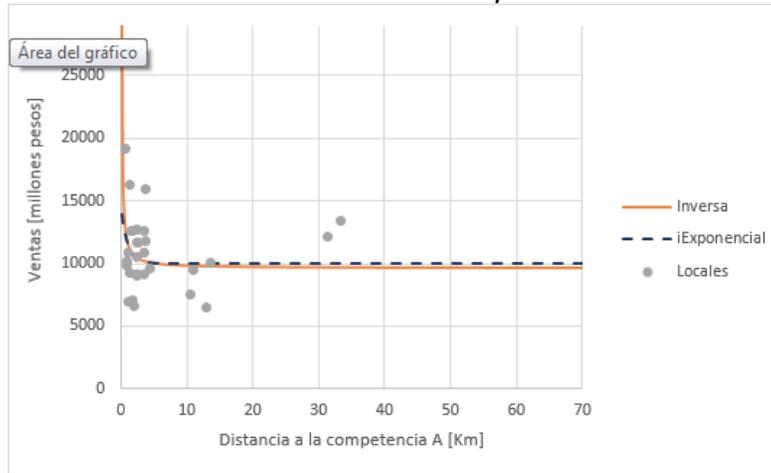
- Distancia a la competencia: se esperaría que la competencia afecte la venta de un local, mientras más cerca se encuentre la competencia, menores serían las ventas que generaría el local en cuestión.

En la *Ilustración 2*, *Ilustración 3* e *Ilustración 4*, se presentan los gráficos de dispersión de ventas versus distancia a la competencia A (*compA*), B (*compB*) y otras tiendas pertenecientes a la empresa (*compE*), donde cada punto representa un local.

Como no se ve un patrón de comportamiento de las ventas en función de la distancia a las distintas competencias, se realiza una regresión para las siguientes funciones que se podrían relacionar con los datos: lineal, logarítmica, inversa, potencia, compuesto, logístico, crecimiento e inversa de exponencial³⁰.

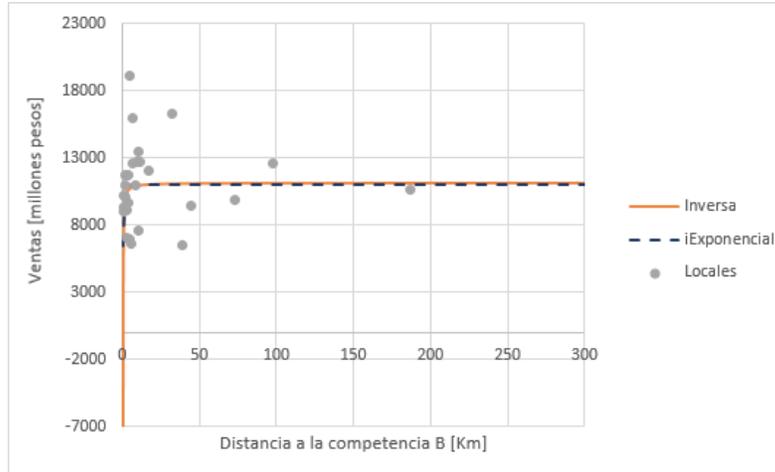
³⁰ En Anexo 3, Anexo 4 y Anexo 5 se puede ver el ajuste de cada función, páginas 66 y 67.

Ilustración 2: Promedio de ventas anuales 2006-2013 vs distancia a compA



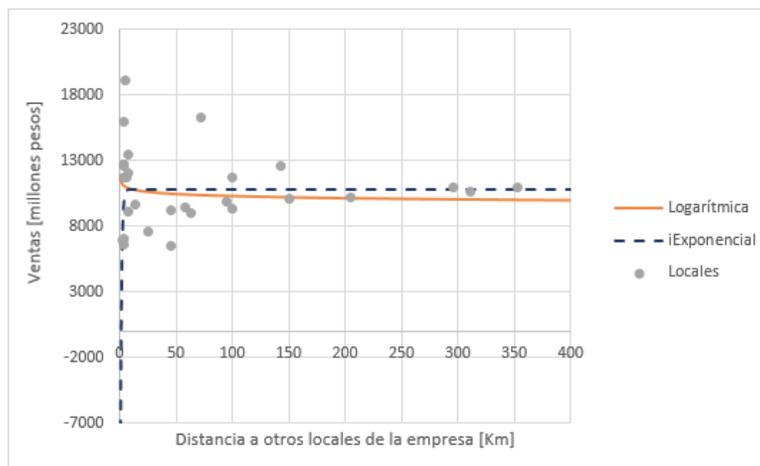
Fuente: Elaboración propia.

Ilustración 3: Promedio de ventas anuales 2006-2013 vs distancia a compB



Fuente: Elaboración propia.

Ilustración 4: Promedio de ventas anuales 2006-2013 vs distancia a compE



Fuente: Elaboración propia.

En la Tabla 4 se muestran las dos mejores funciones que se ajustaron a los datos.

Tabla 4: Mejores funciones para representar la distancia a las competencias

	Función	R2
compA	Inversa	0,064
	iExponencial	0,05
compB	Inversa	0,049
	iExponencial	0,047
compE	iExponencial	0,033
	logarítmica	0,017

Fuente: Elaboración propia.

Para las competencias A y B se tiene un mejor ajuste con la función inversa. Sin embargo, no se tiene un buen comportamiento cuando la competencia se encuentra a distancias menores de un kilómetro, donde las ventas se duplican cuando la distancia a la competencia varía en 200 metros, lo que no se acerca a la realidad. Debido a esto, se considera la función inversa de la exponencial para representar la influencia de la competencia en las ventas, representada por la siguiente métrica:

$$iExponencial(x) = \frac{b_1}{\exp(x)}$$

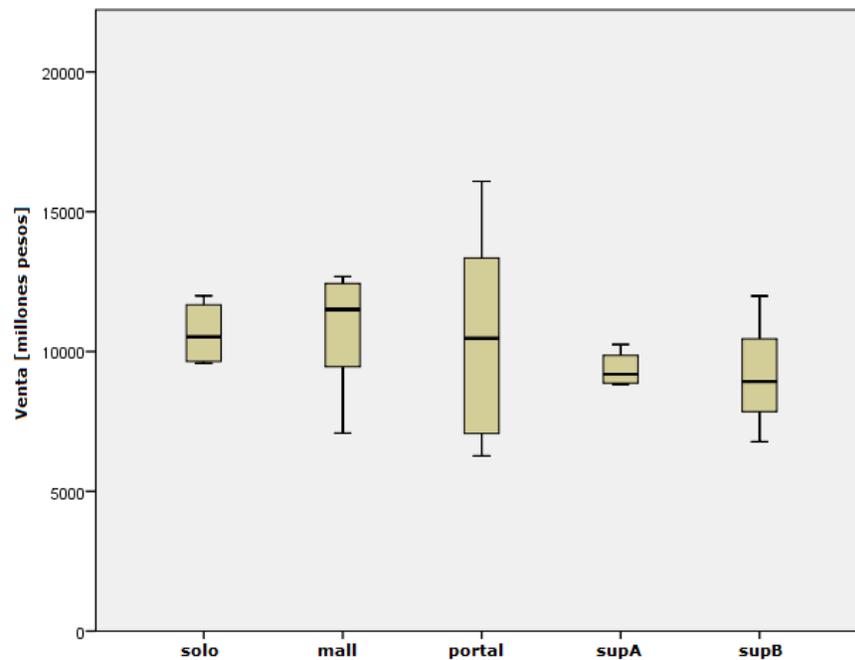
Si bien en la Ilustración 4 se ve que la función inversa de la exponencial no se comporta bien a distancias menores a un kilómetro de distancia, en la realidad ningún local perteneciente a la misma cadena se encuentra a dicha cercanía. Las tiendas que se encuentran más próximas están a 2,3 kilómetros, donde se tiene una buena representación, por lo que se cree apropiado utilizar dicha función.

Por otro lado, no se encuentra una relación negativa entre las ventas y la distancia a la competencia A, en la *Ilustración 2*. Esto se puede deber a que las ventas realizadas no sólo dependen de la competencia en cuestión, sino que existen otros factores que podrían disminuir su efecto negativo, como el tipo de ubicación y sector en que se encuentran ubicadas las tiendas.

- Ubicación compartida: se espera que los locales que compartan su sitio con otras tiendas comerciales tengan mayores ventas, debido a que se podría generar un mayor poder atractivo hacia las personas. En la Ilustración 5 se presentan el gráfico de promedio de ventas versus el

tipo de ubicación compartida, mall, portal, junto al supermercado A (*supA*) o B (*supB*).

Ilustración 5: Promedio venta anual 2006-2013 vs ubicación compartida



Fuente: Elaboración propia.

En la Ilustración 5 se ve que las ventas promedio de los locales ubicados en un mall o portal son mayores que las de las tiendas que comparten ubicación con un supermercado.

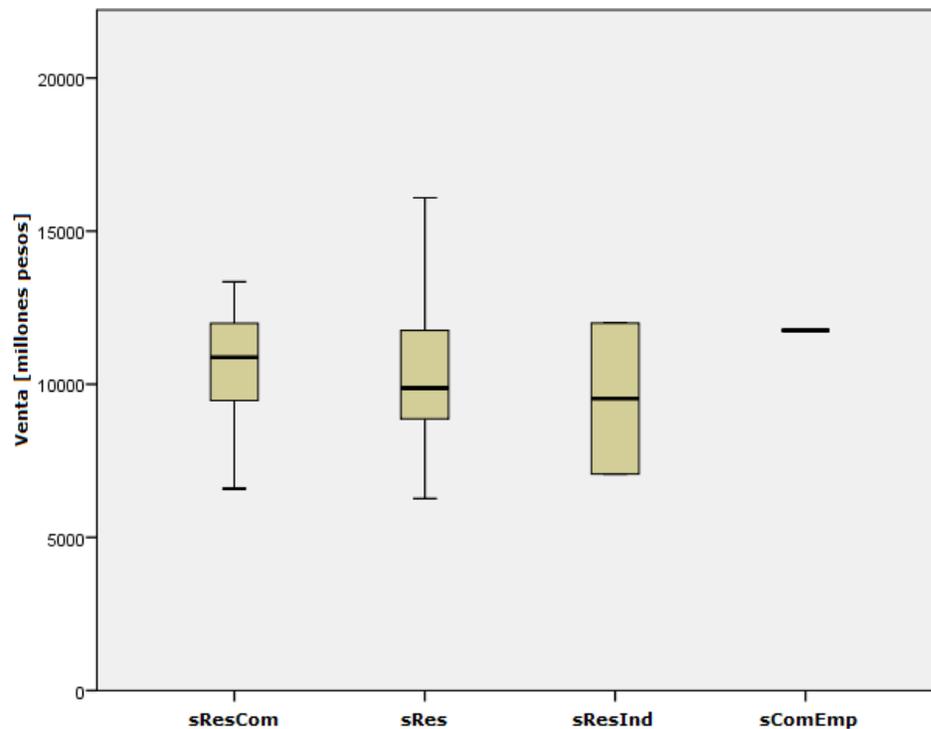
Las ventas promedio de los locales que comparten sitio con un supermercado son menores que la de tiendas que no comparten su ubicación. Esto se podría deber a que los supermercados generarían competencia para algunas categorías de productos de las tiendas. Sin embargo, los ejecutivos de la empresa aseguran que la presencia de estos supermercados incrementa el flujo de personas en sus locales, favoreciendo las ventas. Por lo que se cree que podría ser causa de la mayor presencia de su competencia en el sector en que se encuentran.

De acuerdo a lo planteado, se espera que la presencia de ubicación compartida afecte positivamente a las ventas variando su relevancia, donde estar situado en un mall o portal tenga mayor influencia en las ventas.

- Sector en que se encuentra el local: el sector en que se encuentra el local podría generar distinto nivel de flujo de personas, causando diferentes efectos en las ventas de un local. En la Ilustración 6 se

presenta el gráfico del promedio de ventas de los locales de acuerdo al sector en que se encuentran.

Ilustración 6: Promedio de ventas 2006-2013 vs Sector



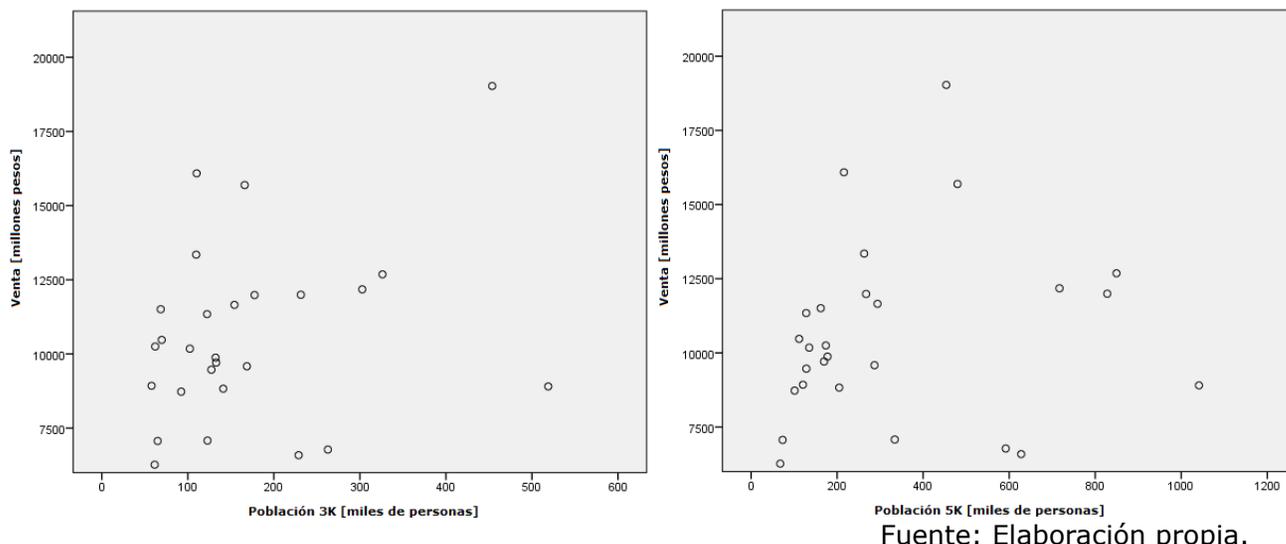
Fuente: Elaboración propia.

Se puede ver que el sector comercial-empresarial (sComEmp) posee la mayor venta, siendo importante notar que la cadena solamente cuenta con un local en dicha zona. El sector residencial-comercial (sResCom) tiene mayor venta promedio que los locales ubicados en una zona residencial (sRes) o residencial-industrial (sResInd).

De acuerdo a lo planteado, se espera que las zonas residencial, residencial-comercial y comercial-empresarial generen un efecto positivo en las ventas de los locales, y el sector residencial-industrial genere un efecto negativo.

- Población: la cantidad de personas que viven en las zonas de influencia de los locales de la empresa podrían representar la demanda de la zona en que se encuentran. En la Ilustración 7 se presenta el gráfico de dispersión del promedio de ventas versus el número de personas que viven a 3 y 5 kilómetros a la redonda para cada local.

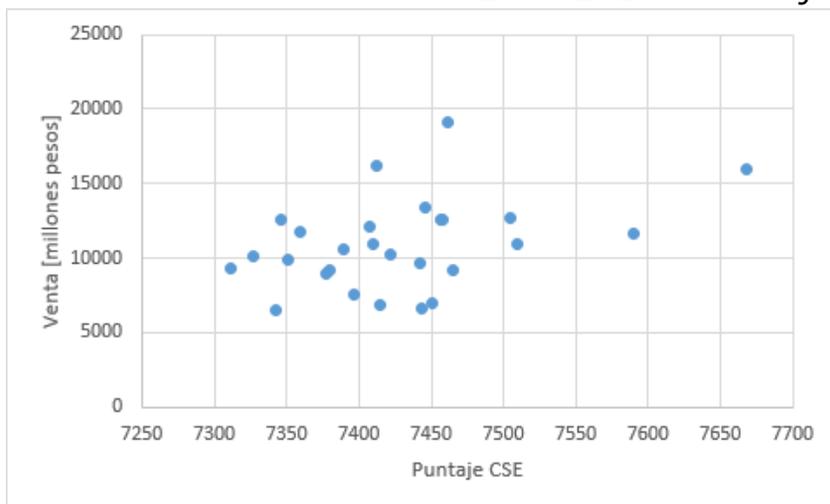
Ilustración 7: Promedio de ventas 2006-2013 vs Población



Se puede ver una relación positiva de la población en las ventas, por lo que se esperaría que las variables *Pob3k* y *Pob5k* tuvieran un efecto lineal favorecedor.

- Caracterización de los hogares: se podría esperar que al aumentar el número de hogares pertenecientes a una clasificación socioeconómica aumenten las ventas, como se describe para la población en el punto anterior. Sin embargo, como cada CSE tiene un gasto diferente en productos de mejoramiento del hogar, se cree que la influencia de cada CSE debe variar.

Ilustración 8: Promedio de ventas 2006-2013 vs Puntaje CSE ponderado



Fuente: Elaboración propia.

En la Ilustración 8 se presenta gráfico de dispersión del promedio de ventas anuales 2006-2013 versus el puntaje CSE ponderado para cada local. Este puntaje se calcula para cada local de la siguiente manera:

$$Puntaje\ ponderado_{local\ i} = \frac{\sum_{N=1}^5 Puntaje\ medio\ CSE_N * cseN_{local\ i}}{\sum_{N=1}^5 cseN_{local\ i}}$$

Donde,

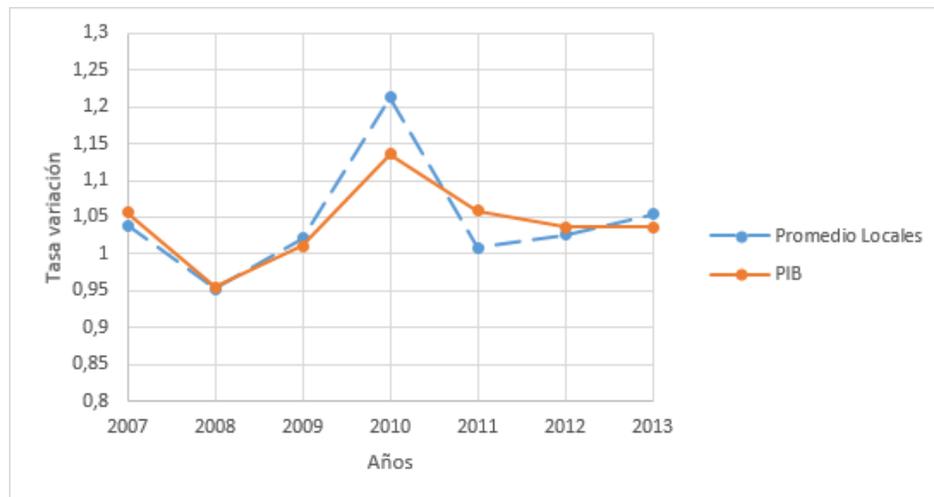
Puntaje medio CSE_N se obtiene al calcular el promedio de los puntajes mínimo y máximo de la CSE N³¹,

cseN_{local i} corresponde al promedio de número de hogares pertenecientes a la CSE N del local i durante los años 2006-2013.

Se puede ver una leve relación positiva entre el puntaje CSE y las ventas de cada local, la que resulta parecida a la presentada en la Tabla 3, donde la tasa de consumo va disminuyendo a medida que baja el GSE.

- PIB: el PIB nacional es un indicador que, entre otros aspectos, sirve para representar el bienestar económico en que se encuentra el país. De esta forma, se podría esperar que el PIB esté relacionado positivamente con las ventas de los locales, como se muestra en la Ilustración 9.

Ilustración 9: Variación PIB y Ventas promedio de locales con respecto al año anterior



Fuente: Elaboración propia.

2. ESTIMACIÓN DEL MODELO

La base de datos, con los valores de las variables por local y año, se separa en dos, base para calibración del modelo y base de testeo. Los datos correspondientes a los años 2006-2012 son utilizados para la calibración del modelo, y los registros correspondientes al año 2013 se destinan a su validación. Como las medidas de las variables consideradas son muy

³¹ Estos puntajes se presentaron en la Tabla 2, página 20.

diferentes, se normalizan para lograr que sean comparables en magnitud y dispersión.

Para la estimación del modelo se utilizan los métodos:

- Mínimos cuadrados ordinarios con selección de variables “pasos sucesivos”,
- LASSO.

El modelo a estimar tiene la siguiente forma:

$$Venta_{i,t} = \beta_0 + \sum_{j=1}^n \beta_j * x_{i,j,t}$$

Donde,

$Venta_{i,t}$ es la venta anual para el local i en el año t ,

$x_{i,j,t}$ representa la variable predictora j para el local i en el año t ,

β_0 es el intercepto del modelo,

β_j corresponde al coeficiente de la variable predictora j .

2.1. ESTIMACIÓN CON PROMEDIO DE DATOS ANUALES

Para la estimación del modelo, se utiliza el promedio de las variables escogidas por local, utilizando un registro de datos por local para comparar los factores que podrían influir en las ubicaciones en que se encuentran las tiendas. Al realizar esto, se obtienen 27 registros para la calibración del modelo.

Se comenzará estimando un modelo con MCO, considerando algunas variables para analizar su comportamiento, y posteriormente se agregarán más para ver su efecto.

2.1.1. Modelo 1

En este modelo se consideran las siguientes variables, siendo las más utilizadas en otros modelos de estimación de demanda, que representan el atractivo del local (*superficie*), potencial de mercado (*población*), rivalidad (*distancia a la competencia*) y situación económica del país (*PIB*):

$\ln(superf_i)$: logaritmo natural de la superficie de venta del local i ,

$ecompA_i = 1/\exp(compA)$: distancia a la competencia A más cercana al local i ,

$ecompB_i = 1/\exp(compB)$: distancia a la competencia B más cercana al local i ,

$ecompE_i = 1/\exp(compE)$: distancia al local de la misma empresa más cercano del local i ,

$pob3k_i$: población a 3 kilómetros a la redonda del local i ,

$pob5k_i$: población a 5 kilómetros a la redonda del local i ,

pib : PIB nacional.

Al realizar la estimación del modelo, considerando como variable dependiente la *Venta promedio* de los locales, se obtiene el resultado de la Tabla 5.

Tabla 5: Coeficientes del Modelo 1

Variables	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
<i>ln(superf)</i>	2038,626	423,352	,673	4,815	,000
PIB	-943,244	433,139	-,311	-2,178	,040
<i>ecompe</i>	-684,039	439,736	-,226	-1,556	,133

R cuadrado corregida
,206

Fuente: Elaboración propia.

El modelo 1 explica el 20,6% de las ventas anuales correspondientes a cada año del período de entrenamiento (2006-2012). Se tiene la relación esperada para la superficie de ventas (*ln(superf)*) y distancia a locales de la misma cadena (*ecompe*). Sin embargo, se tiene un coeficiente negativo para el PIB, indicando que las ventas de la empresa disminuirían con el crecimiento de la economía en el país.

La distancia a las competencias A y B, y la población no resultan significativas para el modelo, esto se puede deber a que faltan variables que expliquen el fenómeno.

2.1.2. Modelo 2

Como se cree que faltan variables que interpreten el fenómeno, se agregan las siguientes variables adicionales a la representación del atractivo los locales y sector en que se encuentran³²:

$Mall_i = 1$ si el local i se encuentra dentro de un mall, 0 si no.

$Portal_i = 1$ si el local i se encuentra dentro de un portal, 0 si no.

$SupA_i = 1$ si el local i comparte su ubicación con el supermercado perteneciente a la cadena A, 0 si no.

$SupB_i = 1$ si el local i comparte su ubicación con el supermercado perteneciente a la cadena B, 0 si no.

$Solo_i = 1$ si el local i se encuentra solo en el sitio, 0 si no.

$sRes_i = 1$ si el local i se encuentra en una zona residencial, 0 si no,

$sResCom_i = 1$ si el local i se encuentra en una zona residencial y comercial, 0 si no,

$sResInd_i = 1$ si el local i se encuentra en una zona residencial e industrial, 0 si no.

³² La variable *sComEmp*, que representa al sector comercial-empresarial, no se incluye en la estimación, ya que al 2012 no se encontraban locales inaugurados en este tipo de zona.

Al realizar la estimación del modelo se obtiene el mismo resultado entregado por el modelo 1, presentado en la Tabla 5. Esto se puede deber a que no se están considerando efectos relacionados entre el atractivo del local y la competencia o población de la zona.

2.1.3. Modelo 3

Para incluir efectos que podrían existir entre el atractivo de un local y sector en que se ubica con la competencia y población, se agregan interacciones entre las variables que miden estos factores.

La interacción entre las variables se representa con la siguiente métrica:

$$Interacción_{X,Y} = X * Y$$

Donde,

X, Y son las posibles variables relacionadas.

Para considerar el efecto que podría tener el atractivo de un local sobre la competencia se calcula la interacción entre las siguientes variables: $ln(superf_i)$, $Mall_i$, $Portal_i$, $supA_i$, $supB_i$, $solo_i$, $sRes_i$, $sResCom_i$, $sResInd_i$, con $ecompA_i$, $ecompB_i$ y $ecompE_i$. Formando 27 interacciones.

Para incluir el posible efecto que causa el sector junto a la población que circunvala al local, se calculan las interacciones entre $sRes_i$, $sResCom_i$, $sResInd_i$ con $pob3k_i$ y $pob5k_i$, generando 8 nuevas interacciones. Estas 35 relaciones complementan a las variables consideradas en el modelo 2.

Al realizar la estimación se obtienen los coeficientes mostrados en la Tabla 6. Al incluir efectos relacionados entre las variables predictoras, el modelo logra explicar el 47,6% de la varianza de las ventas anuales.

Tabla 6: Coeficientes del modelo 3

Variables	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
(Constante)	9568,036	431,664		22,165	,000
ln(superf)	1799,533	317,019	,594	5,676	,000
mall*ecompa	2127,520	379,300	,702	5,609	,000
sResCom*ecompe	-1822,010	405,835	-,602	-4,490	,000
sResCom	2787,770	835,355	,453	3,337	,003
portal*ecompa	1359,365	355,610	,449	3,823	,001
ecompa	-978,172	382,537	-,323	-2,557	,019

R cuadrado corregida
,476

Fuente: Elaboración propia.

Se obtiene el efecto esperado para la superficie de un local ($\ln(\text{supef})$), ubicación en un sector residencial-comercial y la distancia a otras tiendas pertenecientes a la misma empresa ($s\text{ResCom} * e\text{compE}$), pero esta última solamente influiría en las ventas si se encontrara en un sector residencial-comercial.

La cercanía a la competencia A ($e\text{compA}$) influye de diferentes maneras, de acuerdo al tipo de ubicación compartida que tiene el local. La relación entre el mall y la competencia A ($\text{mall} * e\text{compA}$) se puede interpretar como que situarse dentro de un mall obtiene mayor relevancia de atractivo a medida que se encuentra cerca una tienda perteneciente a la competencia A. Se tiene esta misma relación para la ubicación dentro de un portal ($\text{portal} * e\text{compA}$), pero con menor grado de influencia.

Para ver el efecto neto que genera la competencia en las ventas se consideran 8 locales, con tipos de ubicación compartida mostradas en la Tabla 7.

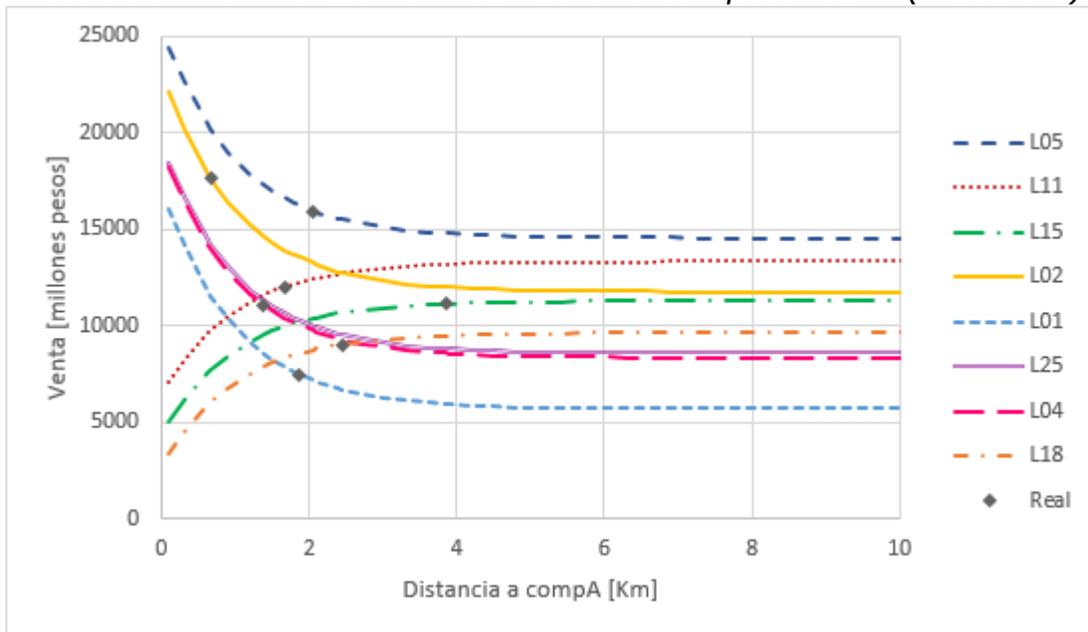
Tabla 7: Ubicación compartida locales L01, L02, L04, L05, L11, L15, L18 y L25

Locales	Ubicación compartida
L01	Portal
L02	Portal
L04	Mall
L05	Mall
L11	No Mall/Portal
L15	No Mall/Portal
L18	No Mall/Portal
L25	Portal

Fuente: Elaboración propia.

En la Ilustración 10 se representa la variación de las ventas con respecto a la distancia a la competencia A para los locales mencionados en la Tabla 7, manteniendo el resto de las variables constantes.

Ilustración 10: Ventas vs Distancia a la competencia A (modelo 3)



Fuente: Elaboración propia.

Se tiene un efecto negativo de la competencia A para los locales que no se encuentran ubicados en un Mall o Portal, y un efecto neto positivo para aquéllos que se encuentran en este tipo de centros comerciales. Esto se puede deber a que el atractivo que se genera con la reunión de diversas tiendas supera la influencia negativa de la competencia A en las ventas.

En el modelo no son significativas las variables que representan a la población que vive en las zonas de influencia de los locales, esto se podría deber a que no todas las personas tienen el mismo efecto en las ventas.

2.1.4. Modelo 4

Como el número de personas que se encuentra cerca de los locales no resulta ser significativo en los modelos anteriores, se analizará si el tipo de población tiene alguna influencia de acuerdo a su clasificación socioeconómica.

Se añaden las variables que representan la clasificación socioeconómica de la población:

$cseN3K_{i,t}$: número de hogares pertenecientes a la CSE N a 3 kilómetros del local i en el año t .

$cseN5K_{i,t}$: número de hogares pertenecientes al CSE N a 5 kilómetros del local i en el año t .

Además, se calculan las interacciones multiplicativas entre las variables $sRes_i$, $sResCom_i$, $sResInd_i$ y $solo_i$, con $cseN3k_i$ y $cseN5k_i$, para las 5 clasificaciones realizadas, agregando 30 relaciones.

Al realizar la estimación se obtiene el mismo resultado que en el modelo 3, mostrado en la Tabla 6. Esto se puede deber a que la clasificación socioeconómica utilizada no es representativa del comportamiento de gasto de las personas en productos de mejoramiento del hogar.

2.1.5. Modelo 5

Como se consideraron todas las variables en estudio en la estimación utilizando MCO, se estima un nuevo modelo con el método LASSO, para ver si se obtiene un mejor ajuste.

Para estimar el modelo 5, se usan todas variables con sus interacciones mencionadas en los modelos pasados, obteniendo los coeficientes de la Tabla 8.

Tabla 8: Coeficientes del modelo 5

Variables	Coeficientes
(Constante)	10600,543
ln(superf)	988,438
ln(superf)*ecompE	-15,789
mall*ecompA	484,529
PIB	-149,280

R cuadrado corregida
0,324

Lambda
794,534

Fuente: Elaboración propia.

El modelo explica un 32,4% de la varianza de las ventas anuales por local, seleccionando solamente 4 variables, con un parámetro de penalización de coeficientes de 794,534.

La superficie tiene el efecto positivo neto esperado en las ventas ceteris paribus, como se muestra en la

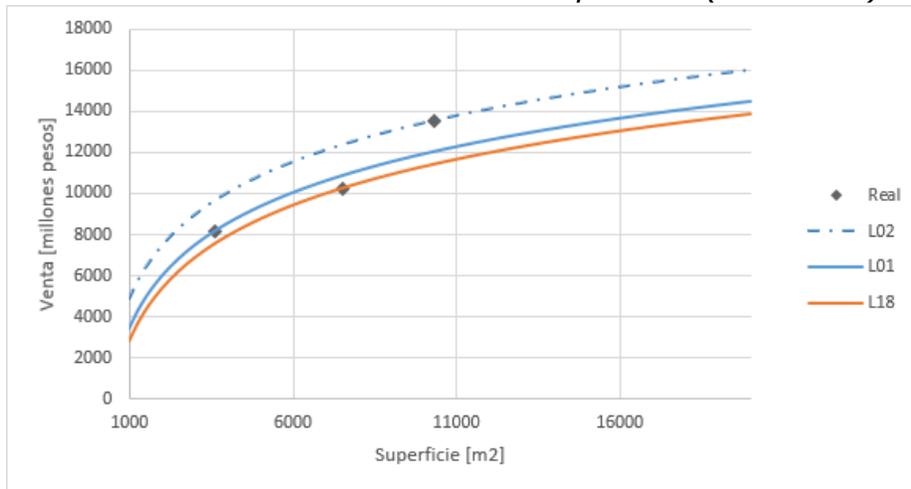
Cuando el local se encuentra dentro de un mall tiene una relación positiva con la cercanía de la competencia A. Como se mencionó anteriormente, esto se podría deber a que el atractivo de estar en un mall toma mayor relevancia cuando la competencia se acerca, o bien, a que el sector se vuelve más comercial y con mayor atractivo, aumentando el flujo de personas. Sin embargo, si el local no se encuentra en un mall, no se vería afectado por la competencia A.

Ilustración 11.

Cuando el local se encuentra dentro de un mall tiene una relación positiva con la cercanía de la competencia A. Como se mencionó anteriormente, esto se podría deber a que el atractivo de estar en un mall toma mayor relevancia cuando la competencia se acerca, o bien, a que el sector se vuelve más

comercial y con mayor atractivo, aumentando el flujo de personas. Sin embargo, si el local no se encuentra en un mall, no se vería afectado por la competencia A.

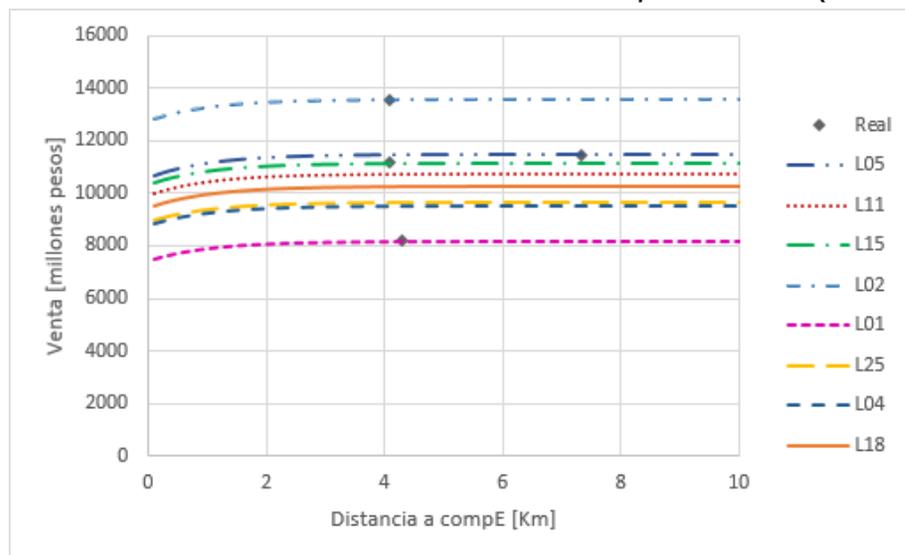
Ilustración 11: Ventas vs Superficie (modelo 5)



Fuente: Elaboración propia.

La cercanía a locales de la misma empresa tiene una incidencia negativa en las ventas ceteris paribus, como se muestra en la Ilustración 12.

Ilustración 12: Ventas vs Distancia a la competencia E (modelo 5)



Fuente: Elaboración propia.

Este efecto generado por la cercanía de los locales ($\ln(\text{superf}) * \text{ecompe}$) se incrementa cuando la tienda en evaluación tiene una mayor área de ventas. La interacción en este modelo indica que mientras menor es la superficie del local, disminuiría el efecto que provocan otros locales de la misma cadena en éste, donde las personas preferirían comprar en la tienda más pequeña de la empresa.

Por otro lado, el aumento del PIB tiene una relación negativa con las ventas, que va en contra de los resultados empíricos que ha obtenido la compañía durante el período en evaluación.

Los resultados mostrados se pueden deber a que falta la influencia de otras variables en el modelo que permitan explicar en mayor proporción las ventas, que se no se consideraron significativas. La selección de sólo 4 variables se podría deber a que LASSO tiene una mayor penalización en la estimación de coeficientes que MCO con pasos sucesivos.

En las estimaciones realizadas, utilizando MCO y LASSO, se obtuvo que las variables que representan a la población y PIB no son significativas. Esto se podría deber a que como se consideró un registro por local con valores promedios para los factores en estudio, no se recoge el cambio en la población y de la economía año a año, por lo que se considerará utilizar los datos de cada local por año de funcionamiento.

En la Tabla 9 se presenta un resumen de los 5 modelos estimados utilizando como variables el promedio de los datos anuales para los locales.

Tabla 9: Resumen modelos estimados con promedio de datos anuales

Modelos	Variables	R2 corregido
1	ln(superf), PIB, ecompE	21%
2	ln(superf), PIB, ecompE	21%
3	ln(superf), mall*ecompA, sResCom*compE, sResCom, portal*ecompA, ecompA	48%
4	ln(superf), mall*ecompA, sResCom*compE, sResCom, portal*ecompA, ecompA	48%
5	ln(superf), ln(superf)*ecompE, mall*ecompA, PIB	32%

Fuente: Elaboración propia.

2.2. ESTIMACIÓN CON DATOS DE LOCALES POR AÑO

Para que los cambios de la población y económicos en el tiempo se reflejen en la influencia de la variables del modelo, se considera utilizar la base de datos creada con los registros de los locales por año. De esta forma, se dispone de 154 registros para la estimación del modelo.

Como en las estimaciones pasadas, se comienza ajustando un modelo a través de MCO con pasos sucesivos, considerando un grupo de variables para ver el ajuste que tiene el modelo, para posteriormente ir integrando más para analizar sus efectos.

2.2.1. Modelo 6

Para ajustar el modelo se consideran las variables que representan el atractivo del local, rivalidad, potencial de mercado y situación económica del país:

$\ln(\text{superf}_i)$: logaritmo natural de la superficie de venta del local i ,
 $\text{ecompA}_i = 1/\exp(\text{compA})$: distancia a la competencia A más cercana al local i ,
 $\text{ecompB}_i = 1/\exp(\text{compB})$: distancia a la competencia B más cercana al local i ,
 $\text{ecompE}_i = 1/\exp(\text{compE})$: distancia al local de la misma empresa más cercano del local i ,
 pob3k_i : población a 3 kilómetros a la redonda del local i ,
 pob5k_i : población a 5 kilómetros a la redonda del local i ,
 pib : PIB nacional.

Al realizar la estimación, se obtienen los coeficientes de la Tabla 10.

Tabla 10: Coeficientes del modelo 6

Variables	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.	R cuadrado corregida
	B	Error típ.	Beta			
(Constante)	11180,492	174,885		63,931	,000	,476
$\ln(\text{superf})$	1853,580	200,653	,538	9,238	,000	
PIB	1007,535	177,681	,292	5,670	,000	
ecompE	-1134,814	190,427	-,329	-5,959	,000	
pob3k	745,980	222,481	,217	3,353	,001	
ecompA	257,843	187,402	,075	1,376	,171	

Fuente: Elaboración propia.

El modelo explica un 47,6% de la varianza de las ventas, a un nivel de significación del 82%. Se tiene el efecto positivo esperado de la superficie del local ($\ln(\text{superf})$), PIB y población a 3 kilómetros a la redonda (pob3k), y negativo de la cercanía de otros locales de la cadena (ecompE).

Sin embargo, se tiene un efecto positivo de la cercanía a la competencia A, lo que se podría deber a que no se están considerando todas las variables de atractivo que influyen en las ventas de los locales.

2.2.2. Modelo 7

Para considerar el atractivo de los locales y zona en que se encuentran, se agregan las variables que describen el tipo de ubicación compartida que poseen y el sector que los rodea:

$\text{Mall}_i = 1$ si el local i se encuentra dentro de un mall, 0 si no.

$\text{Portal}_i = 1$ si el local i se encuentra dentro de un portal, 0 si no.

$SupA_i = 1$ si el local i comparte su ubicación con el supermercado perteneciente a la cadena A , 0 si no.

$SupB_i = 1$ si el local i comparte su ubicación con el supermercado perteneciente a la cadena B , 0 si no.

$Solo_i = 1$ si el local i se encuentra solo en el sitio, 0 si no.

$sRes_i = 1$ si el local i se encuentra en una zona residencial, 0 si no,

$sResCom_i = 1$ si el local i se encuentra en una zona residencial y comercial, 0 si no,

$sResInd_i = 1$ si el local i se encuentra en una zona residencial e industrial, 0 si no.

La estimación de los coeficientes del modelo 7 se presentan en la Tabla 11.

Tabla 11: Coeficientes del modelo 7

Variables	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
(Constante)	9883,372	298,417		33,119	,000
$\ln(\text{superf})$	2060,802	189,126	,598	10,896	,000
PIB	1241,023	172,031	,360	7,214	,000
ecompe	-1234,080	175,416	-,358	-7,035	,000
mall	2330,982	452,398	,311	5,153	,000
portal	1816,858	413,043	,255	4,399	,000
ecompeA	509,492	189,676	,148	2,686	,008
pob3k	375,076	232,252	,109	1,615	,108
$sResInd$	-1256,070	741,948	-,083	-1,693	,092
ecompeB	-277,336	188,331	-,081	-1,473	,143

R cuadrado corregida
,607

Fuente: Elaboración propia.

Se obtiene un modelo que explica el 60,7% de la varianza de las ventas a un nivel de significación del 85%. Define la relación positiva esperada de la superficie ($\ln(\text{superf})$), PIB, ubicación en un mall o portal, y la población a 3 kilómetros de local (pob3k), e influencia negativa de la cercanía a locales pertenecientes a la misma empresa (compE) y a la competencia B (compB).

La relación negativa entre las ventas y la ubicación en un sector residencial-industrial ($sResInd$) se puede deber a que en ese tipo de zona no se genera tanto flujo de personas, por lo que disminuiría el atractivo del lugar.

Nuevamente se obtiene una relación positiva con la cercanía de la competencia A, lo que se puede deber a que no se están considerando todos los efectos de las variables.

2.2.3. Modelo 8

Para incluir nuevos efectos, se calculan interacciones entre las variables de atractivo del local y sector ($\ln(\text{superf}_i)$, Mall_i , Portal_i , supA_i , supB_i , solo_i , sRes_i , sResCom_i , sResInd_i) con las de competencia (ecompa_i , ecompb_i y ecompe_i), y las variables que representan el sector en que se encuentra el local (sRes_i , sResCom_i , sResInd_i) con la población (pob3k_i y pob5k_i). Estas 35 interacciones se adicionan a las variables consideradas en el modelo 7.

Al considerar las variables mencionadas, se obtiene el modelo de la Tabla 12.

Tabla 12: Coeficientes del modelo 8

Variables	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
(Constante)	9896,859	275,736		35,893	,000
$\ln(\text{superf})$	1668,261	243,024	,484	6,865	,000
$\text{mall}*\text{ecompa}$	1788,348	217,174	,519	8,235	,000
PIB	1216,591	135,963	,353	8,948	,000
ecompe	-810,685	145,906	-,235	-5,556	,000
portal	1139,974	456,219	,160	2,499	,013
$\text{sResCom}*\text{pob5k}$	-1530,222	254,316	-,444	-6,017	,000
sResCom	2803,739	538,013	,399	5,211	,000
sResInd	-4334,259	968,739	-,287	-4,474	,000
$\text{sResInd}*\text{pob5k}$	825,404	220,000	,240	3,752	,000
ecompa	-18076,915	7817,884	-5,248	-2,312	,022
$\text{portal}*\text{ecompa}$	529,804	223,247	,154	2,373	,019
$\ln(\text{superf})*\text{ecompa}$	17654,882	7881,990	5,125	2,240	,026

R cuadrado corregida
,734

Fuente: Elaboración propia.

Al integrar las interacciones entre las variables, se obtiene un modelo que explica el 73,4% de la varianza de las ventas a un nivel de significancia del 97%.

Para visualizar el efecto neto que generan las variables, se consideran los locales de la Tabla 13, que tiene características diferentes de ubicación compartida y sector en el que se encuentran.

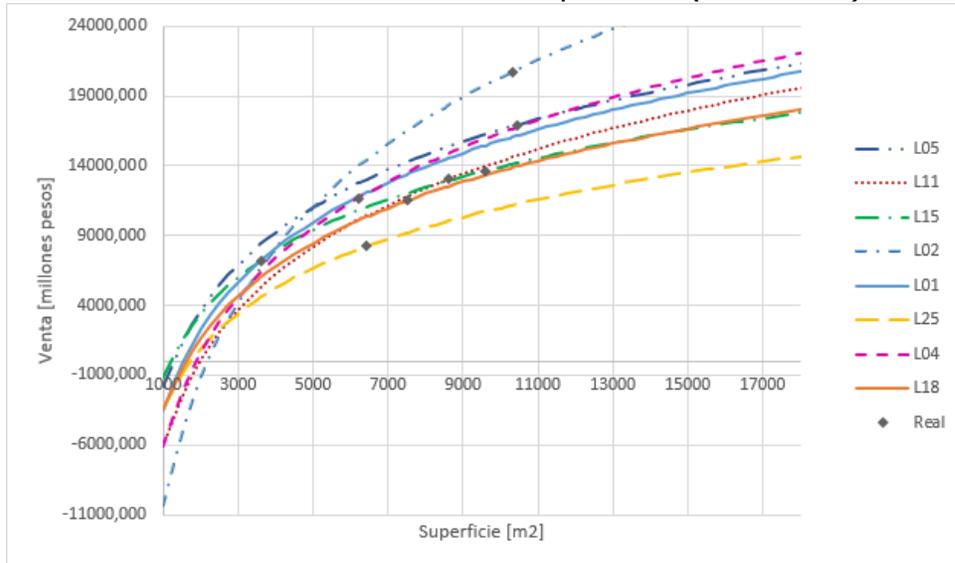
Tabla 13: Características de ubicación de los locales L01, L02, L04, L05, L11, L15, L18 y L25

Locales	Ubicación
L01	Portal; sResCom
L02	Portal; No sResCom/sResInd
L04	Mall; No sResCom/sResInd
L05	Mall; sResCom
L11	No Mall/Portal; sResCom
L15	No Mall/Portal; sResInd
L18	No Mall/Portal; No sResCom/sResInd
L25	Portal; sResInd

Fuente: Elaboración propia.

Se tiene la relación positiva esperada de la superficie de los locales, como se muestra en la Ilustración 13.

Ilustración 13: Ventas vs Superficie (modelo 8)



Fuente: Elaboración propia.

Se puede ver que el modelo no se comporta bien en estimar la venta de algunos locales para superficies menores de 1.500 metros cuadrados. Sin embargo, no representa un problema, ya que esas dimensiones están fuera del rango de evaluación de un local perteneciente a la clasificación de grandes tiendas de mejoramiento del hogar. La superficie promedio de sus locales es de 7.650 m², y la mínima es de 3.600 metros².

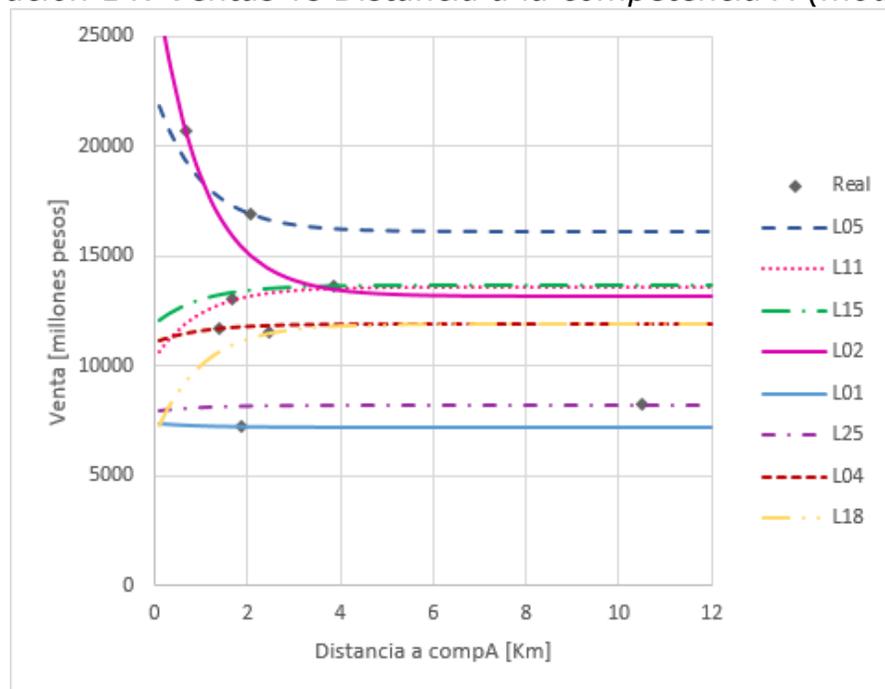
En el gráfico se muestra que se podrían obtener mayores ventas al aumentar la superficie del local ceteris paribus. Hay que tener en cuenta que la mayor cantidad de ventas no necesariamente generan más utilidad, ya que se debe considerar los costos asociados de poseer un área más grande. Además, como la superficie se considera como medida de atractivo, por surtido y disposición

de la tienda, se debería tener la capacidad de aumentar la variedad de productos y distribución.

La relación con la distancia a la competencia A se puede ver en la Ilustración 14.

La ubicación en un mall ($mall * ecompA$) o portal ($portal * ecompA$) y la superficie de ventas del local ($ln(superf) * ecompA$), se relaciona con un efecto positivo en el atractivo de la tienda, lo que indicaría que el atractivo de los locales toma mayor relevancia cuando la competencia A se acerca ceteris paribus, permitiendo disminuir el efecto negativo aislado que genera ($ecompA$). En unos casos, como en los locales L01, L02 y L05, se ve que el atractivo de éstos alcanza a superar la relación negativa con la competencia A, y en otros, como en los locales L04 y L25, logra hacerlo casi imperceptible.

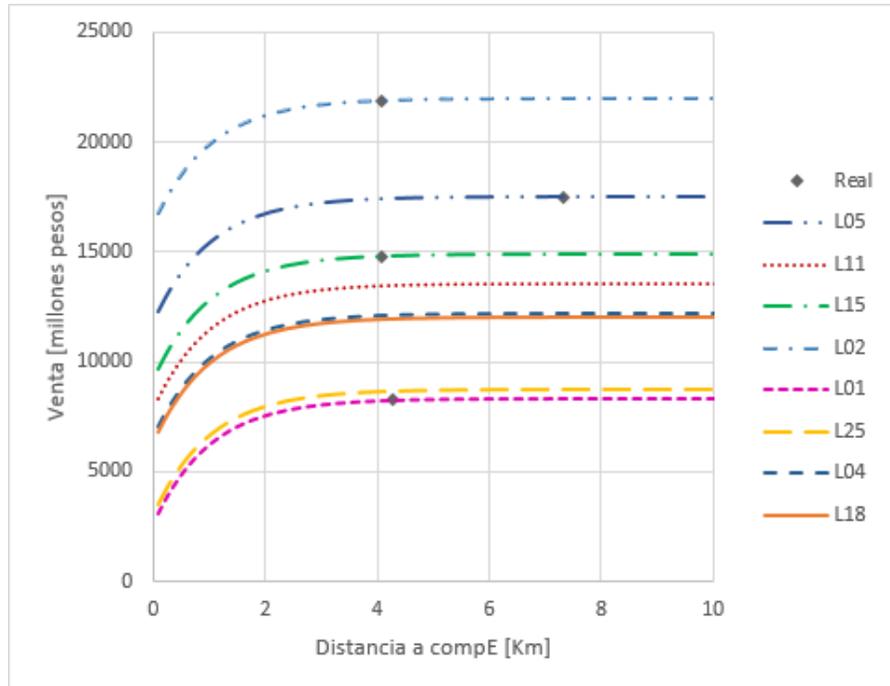
Ilustración 14: Ventas vs Distancia a la competencia A (modelo 8)



Fuente: Elaboración propia.

La relación con la cercanía a otros locales de la misma cadena ($ecompE$) es negativa, como se aprecia en la Ilustración 15. Se ve que alrededor de los 3 kilómetros se comenzaría a producir canibalización de ventas.

Ilustración 15: Ventas vs Distancia a otros locales de la empresa (modelo 8)



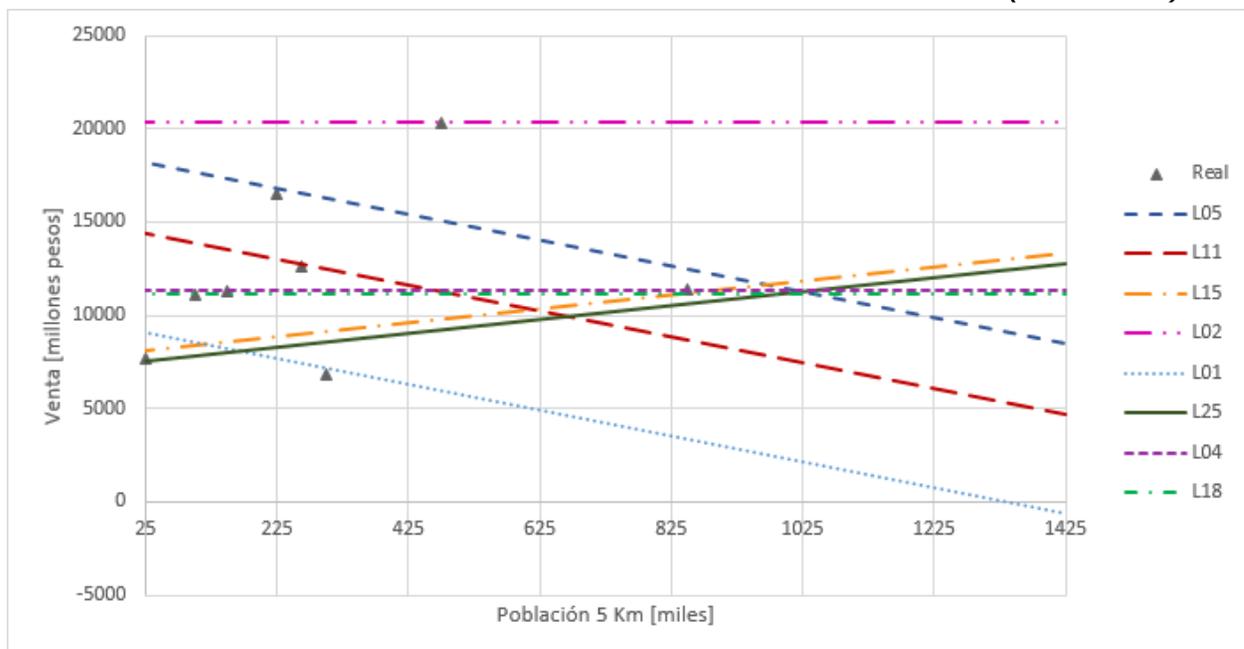
Fuente: Elaboración propia.

La ubicación en un sector residencial-comercial ($sResCom$) tiene una relación positiva, ya que podría generar mayor flujo de personas. Sin embargo, una mayor cantidad de población en este tipo de zona disminuiría su efecto positivo ($sResCom * pob5k$). Esto se puede deber a que si hay mayor población viviendo en el sector, se tendría una menor cantidad de locales comerciales.

La localización en un sector residencial-industrial tiene un efecto negativo en las ventas ($sResInd$), que se ve disminuido con la cantidad de personas que viven en esa zona ($sResInd * pob5k$). Este resultado tiene concordancia con la realidad, ya que en este tipo de zonas, el flujo de personas es menor que en los sectores residenciales o residenciales-comerciales.

El efecto neto de la población a 5 kilómetros a la redonda en las ventas de los locales se puede ver en la Ilustración 16.

Ilustración 16: Ventas vs Población a 5 Km de los locales (modelo 8)



Fuente: Elaboración propia.

De esta forma, la población tendría una relación negativa con los locales ubicados en un sector residencial-comercial (L01, L05 y L11), y una relación positiva con los que se encuentran en una zona residencial-industrial (L11 y L15). No se tiene una influencia en aquello que no se encuentran en este tipo de sectores, lo que se puede deber a que el tipo de población que vive en el sector tiene una relación diferente de acuerdo a su comportamiento de compra.

Por otro lado, el PIB tiene una relación positiva con las ventas anuales de los locales, resultado que se esperaba.

2.2.4. Modelo 9

Para considerar el posible comportamiento de compra en productos de mejoramiento del hogar diferente entre las personas, se consideran las variables de clasificación socioeconómica:

$cseN3K_{i,t}$: número de hogares pertenecientes a la CSE N a 3 kilómetros del local i en el año t .

$cseN5K_{i,t}$: número de hogares pertenecientes al CSE N a 5 kilómetros del local i en el año t .

Al ingresar estas variables se obtiene el resultado mostrado en la Tabla 14.

Tabla 14: Coeficientes del modelo 9

Variables	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
(Constante)	10497,341	282,244		37,192	,000
ln(superf)	1998,760	219,898	,580	9,089	,000
mall*ecompA	1931,185	186,262	,561	10,368	,000
PIB	1210,554	141,728	,351	8,541	,000
ln(superf)*ecompE	22009,615	12922,185	6,389	1,703	,090
portal*ecompA	1008,535	173,436	,293	5,815	,000
ecompA	-866,158	188,256	-,251	-4,601	,000
cse2,5K*sResCom	779,418	240,218	,226	3,245	,001
cse3,5K*sResCom	-1790,193	305,181	-,520	-5,866	,000
sResCom	2337,402	577,564	,332	4,047	,000
sResInd	-4327,629	1378,681	-,286	-3,139	,002
cse2,5K*sResInd	749,760	313,290	,218	2,393	,018
ecompE	-22789,910	12876,417	-6,616	-1,770	,079

R cuadrado corregida
,711

Fuente: Elaboración propia.

Se obtiene un modelo que explica el 71,1% de la varianza de las ventas a un nivel de significancia del 90%, con una selección y relación de variables similar al modelo 8.

La cercanía a locales de la misma empresa tiene una relación aislada negativa (*ecompE*), que resulta disminuida con la superficie de ventas del local en evaluación. La superficie de estos locales se presenta en la Tabla 15, donde el área de ventas promedio para el período 2006-2012 corresponde a 7.727 metros cuadrados.

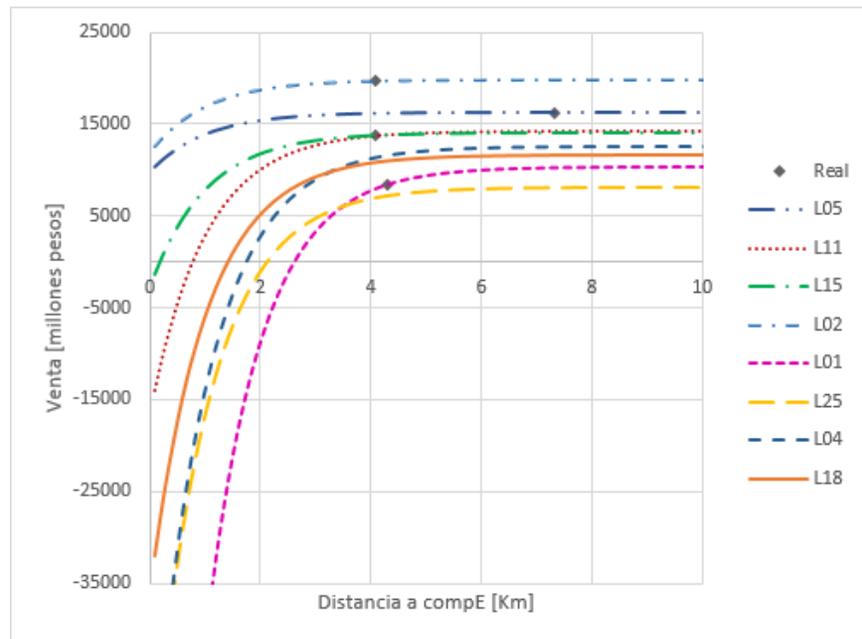
Tabla 15: Superficie de ventas de los locales L01, L02, L04, L05, L11, L15, L18 y L25

Local	Superficie
L01	3.604
L02	10.306
L04	6.208
L05	10.427
L11	8.598
L15	9.599
L18	7.516
L25	6.433

Fuente: Elaboración propia.

En la Ilustración 17 se presenta la variación de las ventas para los locales L01, L02, L04, L05, L11, L15, L18 y L25 en función de la distancia a otra tienda de la misma empresa, manteniendo el resto de las variables constantes.

Ilustración 17: Ventas vs Distancia a otros locales de la empresa (modelo 9)



Fuente: Elaboración propia.

Se ve que los locales más afectados por la cercanía a otras tiendas de la empresa son aquellos que tienen una superficie menor al promedio de la cadena, lo que resulta coherente si se considera la superficie del local como medida de su atractivo. Sin embargo, en este modelo no se considera el efecto que podría tener su ubicación en un mall o portal.

En el modelo se consideran significativa la relación con los hogares pertenecientes a las clasificaciones socioeconómicas 2 y 3 (*CSE2*, *CSE3*) a 5 kilómetros a la redonda del local ubicados en sectores residencial-comercial o residencial-industrial.

Los hogares pertenecientes a *CSE2* tienen influencia positiva en los dos tipos de zonas mencionadas (*CSE2,5K*sResCom* y *CSE2,5K*sResInd*), lo que se podría deber a que este grupo es el que más invierte en grandes tiendas de mejoramiento del hogar, específicamente en la empresa en estudio. La clasificación socioeconómica utilizada no sería equivalente a los grupos socioeconómicos (GSE) usados generalmente, donde se tiene que el ABC1 tiene un mayor monto de compra³³.

La *CSE3* tiene una relación negativa con las ventas de un local que se encuentra en un sector residencial-comercial (*CSE3,5K*sResCom*), que se puede relacionar con que a mayor número de hogares en dicha zona, se transforma en un lugar menos comercial, disminuyendo el flujo de personas.

³³ El detalle del gasto para los GSE se presentó en la Tabla 3: Tasas de consumo por GSE, página 21.

Nuevamente, los locales ubicados en otros sectores no se verían influenciados por la población que vive en la zona.

2.2.5. Modelo 10

Luego de evaluar las variables en estudio con el método de MCO con selección de variables *pasos sucesivos*, se estima un nuevo modelo considerando todos los factores utilizando el método LASSO. Se obtiene la estimación presente en la Tabla 16.

Tabla 16: Coeficientes del Modelo 10

Variables	Coeficientes	R2 corregido
(Constante)	10186,5388	0,711
ln(superf)	2109,0951	
mall*ecompA	1545,358	
portal*ecompA	561,6388	Lambda
sResCom	2888,4391	12,397
sResInd	-2607,4209	
sRes*pob5k	968,8211	
sResInd*pob5k	990,2976	
sResCom*ecompA	-866,0141	
sResInd*ecompA	114,17	
ecompA	-134,8306	
ecompE	-975,0814	
pob5k	-1060,0898	
PIB	1199,2545	

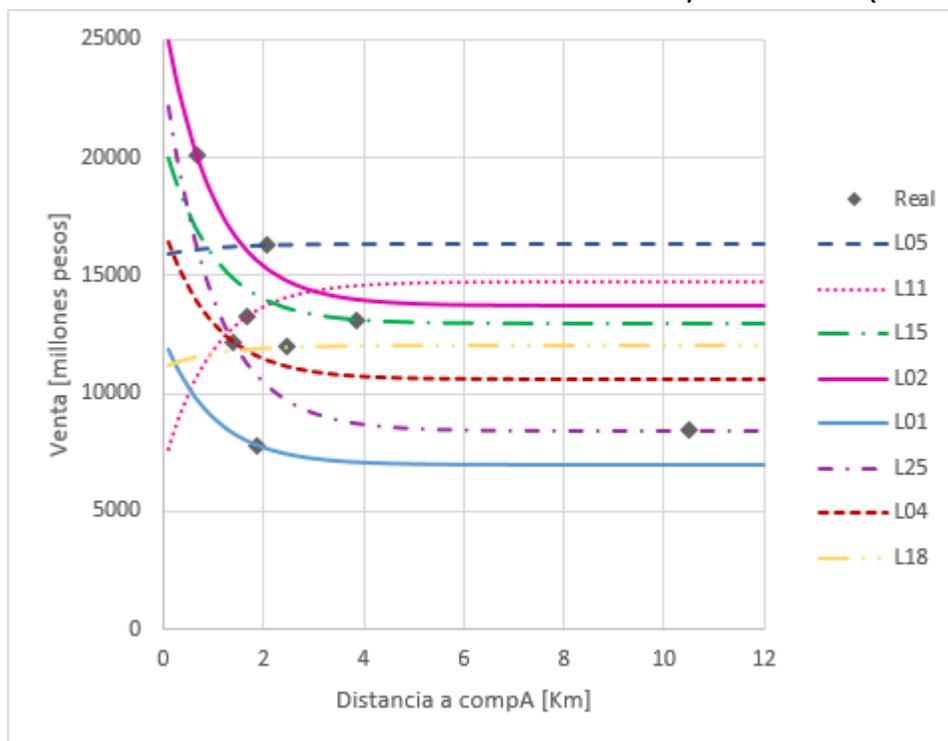
Fuente: Elaboración propia.

El modelo 10 explica el 71,1% de la varianza de las ventas, con un parámetro de penalización de los coeficientes de 12,397.

Este modelo tiene variables con sus efectos similares a los considerados en el modelo 8. Sin embargo, en éste se hace significativa la relación de la cercanía a la competencia A con el sector en que se ubica el local. Si se encuentra en un sector residencial-comercial, la competencia genera una influencia negativa en las ventas ($sResCom*ecompA$), que se puede deber a que ambas tiendas estarían influenciadas por el atractivo de la zona y se favorecerían del flujo de personas que se genera. Si el local está en un sector residencial-industrial, la competencia tendría una relación positiva en las ventas ($sResInd*ecompA$), que podría ser explicado por un aumento en el comercio, generando mayor circulación de personas en el sector. Para ver el efecto neto de la competencia A en las ventas, se analizan los locales L01, L02, L04, L05, L11, L15, L18 y

L25, con las características de ubicación expuestas en la Tabla 13³⁴, presentado en la Ilustración 18.

Ilustración 18: Ventas vs Distancia a la competencia A (modelo 10)



Fuente: Elaboración propia.

Los locales ubicados fuera de un sector residencial-comercial, no se verían afectados por la competencia A, donde sus características de ubicación favorecerían el atractivo que generaría ante la cercanía de la competencia.

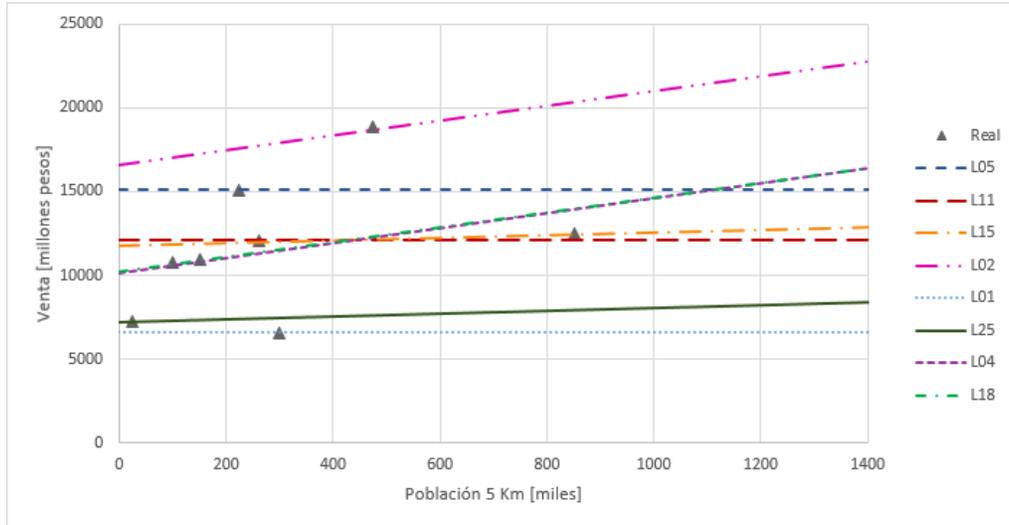
Se tiene un efecto negativo de la cercanía a la competencia A para el local L05, que se encuentra dentro de un portal en un sector residencial-comercial. Su atractivo vinculado a su ubicación en un centro comercial le permitiría disminuir en gran medida la rivalidad con la competencia ($portal * ecompA$), manteniendo una leve disminución de ventas para distancias menores a 2 Km. El local L18, ubicado fuera de un centro comercial y en un sector residencial, se vería más afectado que el local L05 con la cercanía a la competencia A.

Se ve que el local que más disminuiría sus ventas con la cercanía de una tienda de la competencia A es el L11, que se encuentra en un sector residencial-comercial, y no está dentro de un mall o portal. Esto se debería a que la competencia se ubicaría en el mismo sector, siendo favorecida por su atractivo, y como el local no tiene el atributo de encontrarse en un centro comercial, no podría disminuir el efecto negativo que se muestra.

³⁴ La Tabla 13: Características de ubicación de los locales L01, L02, L04, L05, L11, L15, L18 y L25 se encuentra en la página 40.

Este modelo considera una relación negativa con la población que vive a 5 kilómetros a la redonda de los locales (*pob5k*), que se puede ver aminorada en las tiendas que se encuentran en los sectores residencial ($sRes * pob5k$) o residencial-industrial ($sResInd * pob5k$). En la Ilustración 19 se presenta la influencia neta de la población para los locales descritos en la Tabla 13.

Ilustración 19: Venta vs Población a 5 Km (modelo 10)



Fuente: Elaboración propia.

La influencia positiva de la población en los sectores residencial y residencial-industrial se puede deber a que no existe una fuente de generación de flujo de personas externa al número de hogares, por lo que a mayor número de viviendas en la zona, habría más circulación de personas.

Por otro lado, en un sector residencial-comercial, las tiendas comerciales generarían la atracción que aumenta el movimiento en la zona. De esta forma, a mayor población, disminuiría el comercio, y con esto, el atractivo explicado. Luego de considerar todas las variables en estudio, con sus valores promedios y anuales por local, se tienen los 10 modelos expuestos en la Tabla 17.

Tabla 17: Resumen de los modelos estimados

Modelos	Tipo de datos	Variables	R2 corregido
1	Promedio datos anuales 2006-2012 para cada local	ln(superf), PIB, ecompE	21%
2	Promedio datos anuales 2006-2012 para cada local	ln(superf), PIB, ecompE	21%
3	Promedio datos anuales 2006-2012 para cada local	ln(superf), mall*ecompA, sResCom*compE, sResCom, portal*ecompA, ecompA	48%
4	Promedio datos anuales 2006-2012 para cada local	ln(superf), mall*ecompA, sResCom*compE, sResCom, portal*ecompA, ecompA	48%
5	Promedio datos anuales 2006-2012 para cada local	ln(superf), ln(superf)*ecompE, mall*ecompA, PIB	32%
6	Datos anuales 2006-2012 para cada local	ln(superf), PIB, ecompE, pob3k, ecompA	53%
7	Datos anuales 2006-2012 para cada local	ln(superf), PIB, ecompE, mall, portal, ecompA, pob3k, sResInd, ecompB	61%
8	Datos anuales 2006-2012 para cada local	ln(superf), mall*ecompA, PIB, ecompE, portal, sResCom*pob5k, sResCom, sResInd, sResInd*pob5k, ecompA, portal*ecompA, ln(superf)*ecompA	73%
9	Datos anuales 2006-2012 para cada local	ln(superf), mall*ecompA, PIB, ln(superf)*ecompE, portal*ecompA, ecompA, CSE2,5K*sResCom, CSE3,5K*sResCom, sResCom, sResInd, CSE2,5K*sResInd, ecompE	71%
10	Datos anuales 2006-2012 para cada local	ln(superf), mall*ecompA, portal*ecompA, sResCom, sResInd, sRes*pob5k, sResInd*pob5k, sResCom*ecompA, sResInd*ecompA, ecompA, ecompE, pob5k, PIB	71%

Fuente: Elaboración propia.

Cabe notar que en ninguno de los modelos se consideró como significativa la ubicación compartida con un supermercado. Esto se puede deber a que un 68% de los locales de la cadena se encuentra ubicado en un mall o portal, y un 18% comparte su sitio con un supermercado A o B, teniendo un número de casos insuficiente para obtener una relación entre las ventas y esta característica.

3. EVALUACIÓN DE LA ESTIMACIÓN

Luego de estimar los 10 modelos explicados en la sección anterior, se debe evaluar su ajuste para decidir cuál de ellos es el que mejor representa la realidad de la empresa. Esta evaluación se realiza mediante la comparación de la venta real con su pronóstico para cada una de las tiendas durante el período 2006-2012.

Para medir el desempeño de la estimación, se calcula el error absoluto medio porcentual (MAPE), AIC y BIC asociado a cada modelo, mediante las siguientes fórmulas:

$$MAPE = \frac{100 * \sum_{t=1}^7 \sum_{i=1}^{m_t} \frac{|VtaReal_{i,t} - Pronóstico_{i,t}|}{VtaReal_{i,t}}}{n}$$

$$AIC = n * \ln\left(\frac{RSS}{n}\right) + 2 * p,$$

$$BIC = n * \ln\left(\frac{RSS}{n}\right) + p * \ln(n)$$

Donde,

$VtaReal_{i,t}$ representa la venta real asociada al local i en el año t ,

$VtaEstim_{i,t}$ es la estimación de ventas para el local i en el año t , obtenida con el modelo,

m_t corresponde al número de locales en el año t ,

$n = \sum_{t=1}^7 m_t = 154$ es el número de casos utilizados para evaluar la estimación,

$RSS = \sum_{j=1}^n (VtaReal_j - VtaEstim_j)^2$ es la suma de cuadrados residuales,

p representa el número de variables utilizadas en el modelo para estimar las ventas.

Tras el cálculo de cada indicador, se obtienen los valores presentados en la Tabla 18.

Tabla 18: Indicadores de ajuste de los modelos

Modelos	R2 corregido	AIC	BIC	MAPE
1	21%	2471	2481	24%
2	21%	2471	2481	24%
3	48%	2410	2428	19%
4	48%	2410	2428	19%
5	32%	2448	2460	22%
6	53%	2847	2863	19%
7	61%	2818	2847	17%
8	73%	2749	2788	14%
9	71%	2765	2803	14%
10	71%	2765	2807	14%

Fuente: Elaboración propia.

Se escoge el modelo que tiene mejor ajuste, donde se tiene el menor error promedio de estimación (MAPE), AIC y BIC, y mayor R² corregido.

El menor error de estimación, 14%, se obtiene con los modelos 8, 9 y 10. Para decidir cuál de estos tres modelos tiene un mejor ajuste, se utilizan los criterios R² corregido, AIC y BIC.

El modelo 8 obtiene los mejores indicadores AIC, BIC y R² corregido, explicando un 73% de la varianza de las ventas anuales.

Debido a lo planteado, se decide utilizar el modelo 8 para estimar las ventas de los locales, representado por la siguiente métrica:

$$\begin{aligned}
 \text{Venta}_{i,t} = & \beta_0 + \beta_1 * \ln(\text{superf}_i) + \beta_2 * \text{mall}_i * \text{ecompa}_{i,t} + \beta_3 * \text{PIB}_t + \beta_4 * \text{ecompe}_{i,t} \\
 & + \beta_5 * \text{portal}_i + \beta_6 * \text{sResCom}_{i,t} * \text{pob5k}_{i,t} + \beta_7 * \text{sResCom}_{i,t} \\
 & + \beta_8 * \text{sResInd}_{i,t} + \beta_9 * \text{sResInd}_{i,t} * \text{pob5k}_{i,t} + \beta_{10} * \text{ecompa}_{i,t} + \beta_{11} * \text{portal}_i \\
 & * \text{ecompa}_{i,t} + \beta_{12} * \ln(\text{superf}_i) * \text{ecompa}_{i,t}
 \end{aligned}$$

Donde los coeficientes β_i asociados a cada variable se presentan en la Tabla 19.

Tabla 19: Coeficientes modelo seleccionado

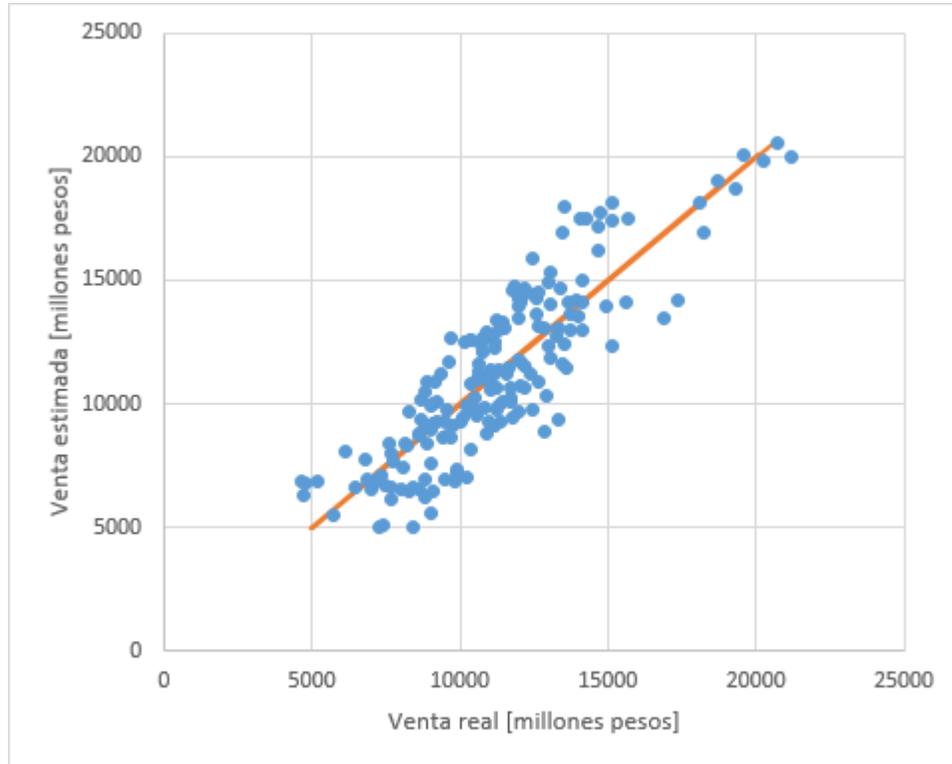
Variables	Coeficientes	Sig.
(Constante)	9896,859	,000
ln(superf)	1668,261	,000
mall*ecompa	1788,348	,000
PIB	1216,591	,000
ecompe	-810,685	,000
portal	1139,974	,013
sResCom*pob5k	-1530,222	,000
sResCom	2803,739	,000
sResInd	-4334,259	,000
sResInd*pob5k	825,404	,000
ecompa	-18076,915	,022
portal*ecompa	529,804	,019
ln(superf)*ecompa	17654,882	,026

Fuente: Elaboración propia.

Todas las variables resultan significativas a un nivel de confianza del 97%, que resulta satisfactorio respecto al nivel de confianza del 95%, utilizado típicamente [42].

Para visualizar gráficamente el ajuste de la estimación de ventas en el período 2006-2012, se realiza un gráfico entre las ventas pronosticadas y las ventas reales, que se presenta en la Ilustración 20.

Ilustración 20: Ventas pronosticadas vs Ventas reales 2006-2012



Fuente: Elaboración propia.

Cada punto en el gráfico representa un local por año, y la recta corresponde a los puntos en que las ventas reales son iguales a las pronosticadas. De este modo, mientras más se acerquen los puntos a la línea naranja, significa que existe un menor error de estimación. De esta forma, los puntos que se encuentran sobre la línea representan una sobreestimación de las ventas para el local en el año, y los puntos que se encuentran bajo la recta, muestran una subestimación de las ventas para el local en el año asociado.

4. EVALUACIÓN DE LA PREDICCIÓN

Para evaluar la capacidad de predicción del modelo se utilizan los datos asociados a los locales para el año 2013, que no fueron utilizados para estimar el modelo.

Se pronostican las ventas del año 2013 para cada local con el modelo seleccionado. Luego se calcula el error porcentual entre las ventas reales y las pronosticadas por el modelo, resultado que se presenta en la Tabla 20.

Tabla 20: Errores de predicción año 2013

Locales	Error %	Locales	Error %
L01	16%	L16	28%
L02	6%	L17	16%
L03	11%	L18	16%
L04	6%	L19	3%
L05	22%	L20	4%
L06	23%	L21	1%
L07	2%	L22	25%
L08	16%	L23	4%
L09	23%	L24	3%
L10	9%	L25	7%
L11	21%	L26	10%
L12	10%	L27	42%
L13	19%	L28	16%
L14	20%	L29	25%
L15	6%		

MAPE	14%
Error % máx	42%
Error % mín	1%

Fuente: Elaboración propia.

Se tiene un error de pronóstico medio del 14%, que es concordante con el error de entrenamiento mostrado en la sección pasada. Con este resultado se descarta la posibilidad de que exista sobreajuste del modelo.

Es importante notar que en este año se inauguran dos nuevos locales, L07 y L29, que tienen un error de pronóstico de 2% y 25%, respectivamente. Este suceso sirve para corroborar la predicción de ventas que entrega el modelo para nuevas ubicaciones, donde en promedio se obtiene un error del 14%.

4.1. EVALUACIÓN DEL MODELO CON EJECUTIVOS DE LA EMPRESA

El modelo escogido es presentado a los ejecutivos de la empresa que apoya el trabajo para su validación, donde se observan los errores de estimación de ventas que se obtienen para los locales. Los errores de pronóstico varían desde el 1 al 28%, con excepción del local L27, donde se sobreestiman sus ventas en un 42%. Este hecho llama la atención del jefe de control de gestión, donde menciona que dicha tienda posee una administración deficiente en comparación al promedio de la cadena, lo que puede explicar el alto error que se obtiene para este local frente al resto.

Debido a lo planteado, se podría considerar que la gestión que tiene un local es un factor que influye en las ventas que se generan, y que el modelo de pronóstico no considera. Integrar esta variable en el modelo no es posible, ya que la empresa no ha calificado la gestión de los administradores de las tiendas existentes.

Por otro lado, en la reunión se comenta que en la etapa de evaluación del proyecto asociado al local L07, sus ventas pronosticadas eran insuficientes para que la tienda fuera rentable. Debido a esto, deciden disminuir su superficie para generar un proyecto beneficioso para la compañía, y vuelven a estimar sus ventas. Transcurrido el primer año de funcionamiento del local, se calcula que el error en el pronóstico de ventas fue cercano al 50%, subestimando el potencial de la ubicación.

Si se hubiese utilizado el modelo planteado en este trabajo en la evaluación del proyecto final, se podría haber obtenido un pronóstico de ventas más preciso, con un error del 2%. O bien, se podría haber mantenido la superficie de local fijada inicialmente si la estimación de ventas hubiese generado un proyecto rentable, o se podría haber elegido otra área que permitiera maximizar las utilidades.

Como en la empresa no utilizan un método establecido para la estimación de ventas, y el encargado de evaluar el proyecto del local L07 ya no se encuentra trabajando en la compañía, no se tiene conocimiento acerca del procedimiento realizado para obtener su pronóstico, lo que permitiría estudiar los desaciertos pasados para tratar de mejorar las predicciones.

5. RECOMENDACIONES DE IMPLEMENTACIÓN DEL MODELO

Para que la empresa pueda utilizar el modelo para el pronóstico de ventas, se le entrega la base de datos creadas con la información necesaria para evaluar las ubicaciones en el territorio nacional.

Se debe tener en cuenta que la base de datos debe ser actualizada, considerando la apertura de locales de la competencia y pertenecientes a su misma cadena, para calcular la distancia a la que se encuentran. Los datos pueden ser actualizados por el ejecutivo que utilice el modelo, o bien, por algún analista del área.

También, se debe utilizar un software de georeferenciación, como ArcGis³⁵, para utilizar la cartografía de Chile desarrollada por el INE, de donde se extrae la información de las manzanas censales que se encuentran a un radio de 5 kilómetros de la ubicación a evaluar. De esta forma, se puede contabilizar a la población que se encuentra en dicha zona.

Luego, se debe actualizar la población al año en que se evalúa, de acuerdo a su tasa de crecimiento comunal, utilizando las proyecciones publicadas por el INE en su Programa de proyecciones de la población. El número de hogares y

³⁵ ArcGis, Esri: <http://www.esri.com/software/arcgis>

población en la zona deberían ser actualizados completamente con la publicación de un nuevo censo.

Por último, se debe actualizar el PIB del año que se va a evaluar, dato que se puede obtener de la base de datos estadísticos del Banco Central de Chile³⁶.

³⁶ Página web de la base de datos: <http://si3.bcentral.cl/Siete/secure/cuadros/home.aspx>

III. Conclusión y comentarios finales

En este trabajo se logra realizar un modelo de estimación de ventas de locales de una empresa de mejoramiento del hogar en Chile, que puede ser utilizado tanto para potenciales tiendas, como para las existentes.

Se concluye que la metodología utilizada para representar el fenómeno es satisfactoria, ya que se obtiene un modelo de pronóstico con coeficiente de determinación, R^2 , igual a 73% y un error de estimación del 14%. Sin embargo, si se tuvieran más datos para estimar el modelo, se podría obtener un mejor ajuste.

Mediante la estimación de los modelos realizando regresiones y usando, tanto valores promedios como desagregados para las variables, se concluye que el método de mínimos cuadrados es más simple y estima mejor que LASSO, superándolo por una leve diferencia cuando se utilizan los datos por local y por año.

Se concluye que los locales de una cadena de la competencia no siempre tienen la misma influencia en las ventas de las tiendas de la empresa para la que se realiza el estudio. Cuando los puntos de venta se encuentran dentro de un centro comercial, se logra disminuir el efecto negativo que causan locales rivales.

Que un local comparta su ubicación con un supermercado no resulta como factor significativo en las ventas que se podrían generar. Se cree que se puede deber a que no existen suficientes locales que no comparten ubicación para que el modelo logre calibrar su posible efecto.

La consideración del tipo de sector en que se ubica un punto de venta resulta ser una variable significativa de la medida de flujo de personas de la zona. Mientras más comercial sea el sector, mayores ventas puede generar. En cambio, mientras más industrial sea, las ventas se verían afectadas negativamente.

Se espera que el gasto que realizan las personas que viven en la zona de influencia de los locales tenga un efecto significativo en sus ventas. Dado que resultó no ser significativo a partir de la clasificación socioeconómica desarrollada por el INE, se concluye que se necesita una mejor medida de la inversión en productos de mejoramiento del hogar. Por esta razón, se sugiere realizar un estudio que evalúe características del comportamiento de compra en artículos de este rubro.

Como la única información de ubicación de los clientes fidelizados de la empresa es la distancia a su residencia, la zona de influencia más significativa encontrada corresponde a un área circular de radio de 5 kilómetros. Se

propone comparar otras zonas de influencia de los puntos de ventas, por ejemplo isócronas que consideren el tiempo de desplazamiento a las tiendas.

Por otro lado, se cree que la administración de los locales influiría en sus ventas, donde podría aumentar el error de pronóstico para aquellos que tengan una gestión particularmente superior o deficiente. Lamentablemente, esta variable no se pudo incluir en el modelo debido a la ausencia de información provista por la empresa.

IV. BIBLIOGRAFÍA

- [1] Cencosud, Memoria anual, 2011, pp. 42-48.
- [2] Michael Levy, Barton Weitz, "Retailing Management", 7th. ed. New York: McGraw-Hill, 2008.
- [3] Kenneth George Jones, James W. Simmons, Location, Location, Location: Analyzing the Retail Environment, Mathuen, 1987.
- [4] Michael Levy, Barton Weitz, "Retail Site Locations" in Retailing Management, 7th. ed. New York: McGraw-Hill, 2008.
- [5] Koutsoyiannis, A., Theory of econometrics: An introductory exposition of econometric methods, Great Britain: Macmillan Education Limited, 1973.
- [6] Graciela Kosiak de Gesualdo, Francisca Sánchez de Dusso, María Rosa Sánchez Rossi, Horacio Ramón Alesandria y María Belén Etchevarría, Atracción Comercial: El comportamiento de los consumidores en los centros comerciales de la ciudad de Santa Fe, 1ra ed. Argentina: Ediciones Universidad Nacional del Litoral, 2006.
- [7] Chasco, P., Modelos de Determinación de Áreas de Mercado del Comercio al Por Menor, Facultad de Ciencias Económicas y Empresariales, Universidad Autónoma de Madrid, Abril, 1997.
- [8] Chapman, Stephen, Planificación y control de la producción, 1ra ed. México: Perarson Educación, 2006.
- [9] Álvarez, Rafael, Estadística multivariante y no paramétrica con SPSS: aplicación a las ciencias de la salud, Madrid: Ediciones Díaz de Santos, 1994.
- [10] David W. Hosmer, Jr., Stanley Lemeshow, Rodney X. Sturdivant, Applied Logistic Regression, 3rd ed. USA: John Wiley & Sons, Inc, 2013.
- [11] Wee, Pearce, Patronage Behavior Toward Shopping Areas: A Proposed Model Based on Huff's Model of Retail Gravitation", Advances in Consumer Research, Provo, 1985.
- [12] Hauser, J. R., Koppelman, F. S., "Alternative Perceptual Mapping Techniques: Relative Accuracy and Usefulness," Journal of Marketing Research, 1979.
- [13] Gautschi, D. A., "Specification of Patronage Models for Retail Center Choice", Journal of Marketing Research, March 1981.
- [14] Nevin, Houston, "Image as a component of attraction to intraurban shopping areas", Journal of Retailing, Vol. 56, 1980
- [15] Howell, Rogers, "Research Into Shopping Mall Choice Behavior", en Monroe, K., Advances in Consumer Reseach, Vol. 8, Association for Consumer Research, Ed. Ann Arbor, 1980.
- [16] Weisbrod, Parcels, Kern, "A Disaggregate Model for Predicting Shopping Area Market Attraction," Journal of Retailing, Vol. 60, 1984.
- [17] Stoltman, Gentry, Anglin, "Shopping Choices: The case of Mall Choice", Advances in Consumer Research, Vol. 18, 1991.
- [18] Más, F., "Atracción de Centros Comerciales en el Contexto de la Elección Discreta Individual", Instituto Valenciano de Investigaciones Económicas, WP-EP, 1996.

- [19] Más, F., "La imagen de los Centros Comerciales Suburbanos en el Modelo de Huff: una aplicación empírica", AEDEM, Granada 1996.
- [20] Donthu, N., Rust, R., "Estimating geographic customer densities using Kernel Density Estimation", Marketing Science, Vol.8, U.S.A., 1989.
- [21] Cohen, Applebaum, "Evaluating Store Sites and Determining Store Rents", Economic Geography, Vol. 36, No. 1, Jan., 1960.
- [22] Burruezo García, Juan Carlos, "La gestión moderna del comercio minorista", Esic Editorial, Madrid, España, Enero 2003.
- [23] Huff, D. L., "Determination of Intra-Urban Retail Trade Areas", Real Estate Program, University of California, Los Angeles, 1962.
- [24] Huff, D. L., "Defining and Estimating a Trading Area," Journal of Marketing, Vol. 28, 1964.
- [25] Bucklin, L., "Trade Area Boundaries: Some Issues in Theory and Methodology," Journal of Marketing Research, Vol. 8, 1971.
- [26] Stanley y Sewall, "Image Inputs to a Probabilistic Model: Predicting Retail Potencial," Journal of Marketing, Vol. 40, 1976.
- [27] Gentry, J. W., Burns, A., "How Important are Evaluative Criteria in Shopping Centre Patronage", Journal of Retailing, Vol. 53, 1977-1978.
- [28] Wee, C. H., "Shopping Area Image: Its Factor Analytic Structure and Relationships with Shopping Trips and Expenditure Behavior", Advances in Consumer Research, Association for Consumer Research, Provo, 1986.
- [29] Stephenson, P. R., Identifying determinants of retail patronage, Journal of Marketing, vol. 33, July 1969.
- [30] Acito, F. y R.D. Anderson, The Influence of Learning on Store Image Differentiation: An Empirical Demonstration, Beckwith, N. et al, Educators' Proceedings, Ed. American Marketing Association, Chicago, 1979.
- [31] Spiggle, S., Sewal, M., "A Choice Sets Model of Retail Selection", Journal of Marketing, 1987.
- [32] Macik, R., Macik, D., Nalewajek, M., "Consumer preferences for retail format choice – Case of polish consumers", Active Citizenship by Knowledge Management & Innovation, Zadar, Croatia, June 2013.
- [33] Hawkins, Del, Pascale Quester, Roger Best, Consumer Behaviour: Implications for Marketing Strategy, 1st Australasian Edition, Sydney, 1994.
- [34] Hastie, T., Tibshirani, R., Friedman, J., The Elements of Statistical learning, Springer, 2009.
- [35] Colin Cameron, A., Windmeijer, Frank A.G., Gramajo, H., Cane, D., Khosla, C., "An R-squared measure of goodness of fit for some common nonlinear regression models", Journal of Econometrics, 1997.
- [36] Días Fernández, Monserrat, Costa Reparaz, Emilio, "Metodología de la investigación econométrica", Doc. 064, Facultad de Ciencias Económicas y Empresariales, Universidad de Oviedo, 1994.
- [37] Keat, Paul G., Young, Philip K. Y., Economía de Empresa, Pearson Educación, 2004.
- [38] James, G., Witten, D., Hastie, T., Tibshirani, R., An Introduction to Statistical Learning, Springer, 2013.

- [39] Nakanishi, M., Cooper, L., *Market Share Analysis*, Kluwer Academic Publishers, 1988.
- [40] Berman, B., Evans, J., *Retail Management: A Strategic Approach*, Prentice Hall, 2012.
- [41] Zar, J.H., *Biostatistical Analysis*, Prentice Hall International, New Jersey, 1984, pp 43–45.
- [42] Cabrerizo Elgueta, Montserrat, *Procesos de venta, Comercio y Marketing*, Editex.
- [43] Jean-Charles Chebat et. Al., "The Lost in a Mall, the Effects of Gender, Familiarity with the Shopping Mall and the Shopping Values on Shoppers' way finding processes", *Journal of Business Research*, 2005.

V. Anexos

Anexo 1: Otros modelos de estimación de ventas

Durante las últimas décadas se han desarrollado diversos métodos que tratan de representar la demanda que enfrentan las tiendas minoristas. Estos modelos han sido generados con la finalidad de explicar el comportamiento de dicho fenómeno, o bien, predecirlo.

Mediante los modelos explicativos[36] se intenta captar cuantitativamente el mecanismo de funcionamiento de la demanda de productos, obteniendo valores para los parámetros que lo definen. En los modelos predictivos, la finalidad es generar pronósticos de la evolución de determinadas variables. El pronóstico pone menor énfasis en la explicación de los cambios de la demanda generados por causas específicas, y mayor preocupación en la obtención de información relacionada a los niveles futuros de las ventas, dada las suposiciones más probables de las variables independientes [37].

Los métodos que permiten estimar las ventas agregadas de una nueva tienda, adicionales al modelo de regresión, corresponden a los siguientes:

- a. *Teoría de los lugares centrales*: modelo impulsado por Christaller (1935), postulando que los clientes se dirigen a la tienda más cercana que posea los bienes deseados para realizar su compra. Se propone el área de estudio como un plano isotrópico, donde no existen barreras de movimiento y la población se encuentra uniformemente distribuida, donde cada persona posee ingresos, gustos y preferencias idénticas.

Si bien parece razonable que la distancia existente entre las personas y la tienda tenga importancia en la pauta de viaje a ésta, la suposición de que los consumidores siempre escogen el local más cercano no se sustenta en la evidencia empírica. Ésta evoca que si el esfuerzo de viaje extra hacia una tienda más lejana es compensado por mejores oportunidades de compra, los consumidores preferirán realizarlo, como queda demostrado con el primer modelo gravitacional comercial de Reilly (1931).

- b. *Modelo de Huff*: la aceptación de que los modelos de elección espacial pueden incluir atributos diferentes a la distancia, introducido con el modelo de Reilly (1931), ha permitido el surgimiento de otros modelos que consideran factores como el tamaño de las tiendas. Es así como Huff (1962,1964) utiliza la distancia entre los clientes y las tiendas y la superficie de aquellas, para calcular la probabilidad de que un cliente compre en un establecimiento determinado.

Es el primer modelo probabilístico de estimación de demanda en función de la posición relativa de un local. Se basa en la premisa de que la probabilidad en la que un consumidor elegirá comprar en una tienda en particular se

incrementa cuando el local incrementa su tamaño, y disminuye a medida que la distancia o tiempo de viaje que demora el consumidor aumenta [4], pudiendo acudir hasta aquellos que se encuentran alejados.

La probabilidad de que un consumidor i visite la tienda j queda expresada mediante la siguiente métrica:

$$P_{ij} = \frac{T_j^\alpha}{\sum_j \frac{T_j^\alpha}{D_{ij}^\beta}}$$

Donde,

T_j^α es el tamaño del local j .

D_{ij}^β corresponde a la distancia entre el cliente i y la tienda j .

Las hipótesis del modelo corresponden a [6]:

- "Los consumidores pueden visitar más de un centro comercial".
- Las personas "valoran las diferentes opciones de selección de establecimientos en base a la utilidad total que les reporta cada una de ellas, y no solo a su localización".

Este modelo no toma en cuenta la proporción en el cambio de la probabilidad de elección de una tienda cuando una nueva entra en el mercado, repartiendo esta probabilidad equitativamente para cada local. Debido a esto, es aplicable a locales que poseen casi todas sus características semejantes, exceptuando su superficie y distancia a la población, por lo que Kotler (2001) "opina que la ecuación de Huff tiene un valor limitado para estimar las ventas potenciales de establecimientos individuales a menos que se incluyan variables adicionales" [6].

Se puede caracterizar la atracción de las tiendas con otro conjunto de variables, adicionales a su tamaño, como la imagen de la empresa como plantea Stanley y Sewall (1976), medidas de accesibilidad y atributos del transporte por Gautshi (1981), mejorando la capacidad predictiva. Sin embargo, se requiere de información de la situación de todos los locales en el mercado con respecto a los nuevos componentes que se tienen en cuenta.

- c. El modelo *Multiplicativo de Interacción Competitiva* de Nakanishi y Cooper (1974) permite considerar una mayor cantidad de variables con el objetivo de alcanzar una mayor explicación en la atracción del modelo simple de Huff. En su formulación se pueden integrar todas las variables deseadas, definiendo la probabilidad de compra en un establecimiento como un

cociente de utilidades, que está compuesta por los atributos de la tienda, quedando representada de la siguiente forma:

$$p_{ij} = \frac{\prod_{k=1}^K A_{kij}^{\beta_k} * \varepsilon_{ij}^*}{\sum_{j=1}^{m_i} \prod_{k=1}^K A_{kij}^{\beta_k} * \varepsilon_{ij}^*}$$

Donde,

p_{ij} es la probabilidad de que un consumidor residente en la situación de elección i seleccione el establecimiento j ,

A_{kij} es la k -ésima variable explicativa del establecimiento j en la situación de elección i ,

K corresponde al total de variables explicativas que incluyen en el modelo,

m_i representa al total de establecimientos j susceptibles de elección para i ,

β_k es el parámetro de sensibilidad de p_{ij} con respecto a la variable k ,

ε_{ij}^* es el término de error de especificación.

Según Nakanishi y Cooper (1982), los parámetros de sensibilidad mencionados pueden ser estimados transformando la igualdad mediante log-centering, que permite su calibración utilizando mínimos cuadrados ordinarios, mínimos cuadrados generalizados o redes neuronales. La transformación queda de la siguiente manera:

$$\ln \frac{p_{ij}}{\hat{p}_i} = \sum_{k=1}^K \beta_k * \ln \frac{A_{kij}}{\hat{A}_{ki}} + \ln \frac{\varepsilon_{ij}^*}{\widehat{\varepsilon}_{ki}^*}$$

Donde,

$\hat{p}_i = (\prod_{j=1}^m p_{ij})^{1/m_i}$ es la media geométrica de p_{ij} ,

$\hat{A}_{ki} = (\prod_{j=1}^m A_{kij})^{1/m_i}$ es la media geométrica de A_{kij} ,

$\widehat{\varepsilon}_{ki}^* = (\prod_{j=1}^m \varepsilon_{kij}^*)^{1/m_i}$ es la media geométrica de ε_{ij}^* .

Esta transformación facilitó su aplicación debido a la dificultad en la estimación de los parámetros. Sin embargo, González et al. (1998) comentan que las críticas que se le atribuyen con mayor frecuencia al modelo son [6]:

- Las complicaciones que surgen de haber calibrado el modelo en base a patrones de compras pasadas al intentar ubicar tiendas con nuevos formatos.
- "El contexto de dependencia del modelo con respecto a los consumidores y al número de opciones investigado, de lo que se deriva que los resultados del modelo no son extrapolables de una zona geográfica a otra".

d. *Modelos Logit* [7]: El modelo *Logit Multinomial* de Mac Fadden (1974) es aplicado en casos donde la variable endógena presenta varios estados, en oposición a la variable dicotómica propia de los modelos Logit Bivariantes. Como pueden existir más de dos estados, la probabilidad de elección de un individuo ubicado en el lugar i asista al establecimiento j puede ser expresado con la siguiente métrica:

$$p_{ij} = \frac{\exp(u_{ij})}{\sum_{k=1}^m \exp(u_{ik})}$$

Donde,

u_{ij} es la utilidad observable de que un individuo i asista al establecimiento j , función de las propiedades de la tienda como de las características del individuo.

Las características de este modelo son:

- Presupone que los individuos evaluarán y compararán simultáneamente a todos los establecimientos de la zona para elegir un determinado local comercial.
- La participación de mercado disminuye de igual forma entre todas las tiendas existentes cuando una nueva ingresa al mercado, en vez de que afecte más a las semejantes a la reciente.
- No se recoge el efecto de agrupación que puede generar el ingreso de un nuevo establecimiento al mercado, favoreciendo el flujo de clientela frente a las tiendas cercanas que se encontraban previamente.

MacFadden (1977) realiza el intento de solucionar los problemas del modelo Logit Multinomial y genera el modelo *Logit Anidado*. Los supuestos asociados son:

- Los individuos procesan jerárquicamente la información de mercado.
- Se conoce la manera en cómo se forma dicha jerarquía en la mente de los consumidores.
- Cada alternativa de elección pertenece a un clúster.

La probabilidad que una persona elija la alternativa M está representada por:

$$p_{iM} = \frac{\exp(u_{iM}) * [\sum_{j \in M} \exp(u_{ij})]^\sigma}{\sum_N \exp(u_{iN}) * [\sum_{j \in M} \exp(u_{ij})]^\sigma}$$

Donde,

$\sum_{j \in M} \exp(u_{ij})$ corresponde al valor inclusivo que describe el atractivo del clúster M, el cual es a su vez el resultado de las alternativas individuales incluidas en él,

σ es el parámetro que refleja el grado en que los individuos procesan la información de modo jerárquico.

Este modelo puede ser aplicado con relativa confianza en los casos en que la definición de los clúster de alternativas sea fácil, como en la elección no espacial (elección de bebidas colas o no colas). Sin embargo, en la elección espacial no siempre son conocidos los límites de cada clúster, puesto que el espacio es una variable continua, por lo que en este caso la aplicación de este modelo conllevaría una gran dosis de subjetividad y de probable error.

- e. *Enfoque Analógico*: corresponde a un método de pronóstico cualitativo basado en el criterio de la persona que pronostica, "con frecuencia estará sesgado con base en la posición potencialmente optimista o pesimista" de dicha persona [8]. Este tipo de técnica se utiliza cuando no se tiene información histórica. "Este tipo de método suele utilizarse para productos individuales o familiar de productos, y rara vez para mercados completos".

Se describen las características de la tienda de los locales más exitosos, tratando de localizar una nueva tienda con elementos similares [4]. Las principales desventajas asociadas a este enfoque corresponden a que no integra un método para la definición de las características de tiendas semejantes, quedando a juicio de cada analista. Además, no se considera la potencial competencia que podría variar entre las distintas ubicaciones.

Anexo 2: Serie IPC, Estadísticas económicas del Banco Central de Chile

Mes	Factor Act								
ene-06	1,348457779	sep-07	1,238990613	may-09	1,152999021	ene-11	1,118889279	sep-12	1,059501128
feb-06	1,349671641	oct-07	1,235149614	jun-09	1,149081175	feb-11	1,116498955	oct-12	1,053578873
mar-06	1,34180457	nov-07	1,225887852	jul-09	1,154040574	mar-11	1,107999059	nov-12	1,058330839
abr-06	1,333279444	dic-07	1,220308295	ago-09	1,15939359	abr-11	1,10448667	dic-12	1,058623169
may-06	1,330026935	ene-08	1,22075279	sep-09	1,146331629	may-11	1,100048434	ene-13	1,05677447
jun-06	1,322279146	feb-08	1,215855138	oct-09	1,14644593	jun-11	1,098157234	feb-13	1,055513287
jul-06	1,315152149	mar-08	1,205806642	nov-09	1,151036764	jul-11	1,096795405	mar-13	1,051459189
ago-06	1,311670227	abr-08	1,2012052	dic-09	1,155432242	ago-11	1,095019642	abr-13	1,056580246
sep-06	1,31145888	may-08	1,187497813	ene-10	1,149425796	sep-11	1,089623411	may-13	1,056677349
oct-06	1,314833468	jun-08	1,170072984	feb-10	1,14621735	oct-11	1,084382367	jun-13	1,049922951
nov-06	1,317052263	jul-08	1,156992574	mar-10	1,145303938	nov-11	1,080916258	jul-13	1,047245308
dic-06	1,315774782	ago-08	1,146365469	abr-10	1,139966908	dic-11	1,074351171	ago-13	1,044676198
ene-07	1,311760826	sep-08	1,134269636	may-10	1,135912491	ene-12	1,07344844	sep-13	1,039105851
feb-07	1,313984406	oct-08	1,124364573	jun-10	1,135912491	feb-12	1,069255672	oct-13	1,037699119
mar-07	1,308401971	nov-08	1,125830947	jul-10	1,128664596	mar-12	1,067567896	nov-13	1,033780456
abr-07	1,300918218	dic-08	1,139487277	ago-10	1,12977363	abr-12	1,066973481	dic-13	1,027682002
may-07	1,292888497	ene-09	1,13423165	sep-10	1,125240383	may-12	1,066676522		
jun-07	1,280927265	feb-09	1,149196026	oct-10	1,124140226	jun-12	1,069852632		
jul-07	1,266631774	mar-09	1,14873676	nov-10	1,123371396	jul-12	1,06995219		
ago-07	1,253037538	abr-09	1,1504609	dic-10	1,122055844	ago-12	1,067667029		

Fuente: Elaboración propia.

Anexo 3: Estimaciones curvilíneas de Venta vs compA

Ecuación	Resumen del modelo					Estimaciones de los parámetros	
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1
Lineal	,002	,056	1	25	,814	10510,765	14,906
Logarítmica	,014	,354	1	25	,557	10962,595	-332,440
Inversa	,064	1,708	1	25	,203	9601,181	1989,603
Compuesto	,004	,100	1	25	,754	10102,103	1,002
Potencia	,009	,231	1	25	,635	10490,213	-,025
Crecimiento	,004	,100	1	25	,754	9,220	,002
Logística	,004	,100	1	25	,754	9,899E-005	,998

Fuente: Elaboración propia.

Anexo 4: Estimaciones curvilíneas de Venta vs compB

Ecuación	Resumen del modelo					Estimaciones de los parámetros	
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1
Lineal	,000	,006	1	25	,937	10576,551	1,053
Logarítmica	,024	,619	1	25	,439	9964,811	324,287
Inversa	,049	1,295	1	25	,266	11161,425	-1880,371
Compuesto	,002	,041	1	25	,841	10156,686	1,000
Potencia	,020	,516	1	25	,479	9684,509	,027
Crecimiento	,002	,041	1	25	,841	9,226	,000
Logística	,002	,041	1	25	,841	9,846E-005	1,000

Fuente: Elaboración propia.

Anexo 5: Estimaciones curvilíneas de Venta vs compE

Ecuación	Resumen del modelo					Estimaciones de los parámetros	
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1
Lineal	,006	,148	1	25	,704	10771,443	-2,185
Logarítmica	,017	,432	1	25	,517	11347,141	-232,328
Inversa	,005	,120	1	25	,732	10409,427	1725,443
Compuesto	,000	,002	1	25	,966	10230,770	1,000
Potencia	,004	,110	1	25	,743	10573,081	-,011
Crecimiento	,000	,002	1	25	,966	9,233	-2,219E-005
Logística	,000	,002	1	25	,966	9,774E-005	1,000

Fuente: Elaboración propia.