

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

ESTUDIO DEL EFECTO DE CANIBALIZACIÓN EN UNA TIENDA DE RETAIL AL
INCORPORAR UNA NUEVA TIENDA DE LA MISMA CADENA UTILIZANDO
INFORMACIÓN TRANSACCIONAL

TESIS PARA OPTAR AL GRADO DE MAGISTER EN GESTIÓN DE OPERACIONES
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

ANDRÉS JAVIER HAZBUN RIUS

PROFESOR GUÍA:
RICARDO MONTOYA MOREIRA

MIEMBROS DE LA COMISIÓN:
MARCEL GOIC FIGUEROA
MARCELO OLIVARES ACUÑA

SANTIAGO DE CHILE
2014

Resumen

RESUMEN DE LA TESIS PARA OPTAR AL TITULO DE: Ingeniero Civil Industrial y grado de Magíster en Gestión de Operaciones

POR: Andrés Javier Hazbun Rius

FECHA: 24/11/2014

PROFESOR GUÍA: Ricardo Montoya Moreira

ESTUDIO DEL EFECTO DE CANIBALIZACIÓN EN UNA TIENDA DE RETAIL AL INCORPORAR UNA NUEVA TIENDA DE LA MISMA CADENA UTILIZANDO INFORMACIÓN TRANSACCIONAL

El presente trabajo de investigación estudia el efecto de canibalización, tanto en la incidencia de compra como en el monto de compra, al incorporar una nueva tienda de la misma cadena Falabella, en donde el foco principal está en analizar la variación en el comportamiento de compra de los clientes actuales de Falabella. Para cuantificar la canibalización se obtiene una caracterización de la decisión de compra de los clientes, en donde luego se realiza un comparativo entre las participaciones de mercado de las tiendas considerando los periodos “Antes” y “Después” de la nueva apertura.

Para la obtención de los valores relacionados a la incidencia de compra y la variación del monto de compra, se realizan modelos Mixed Logit y Regresión Lineal respectivamente, en donde para el primer modelo se obtienen parámetros específicos para cada cliente. A su vez, en ambos modelos se utiliza un enfoque bayesiano, en donde se aborda la incertidumbre con una distribución de probabilidad sobre los parámetros de la función de utilidad. Debido a la dificultad de resolver integrales múltiples en el enfoque bayesiano, los modelos son ejecutados por los métodos de muestreo Markov Chain Monte Carlo (MCMC) y Gibbs respectivamente, en donde no se maximiza ninguna función. Por otro lado, los datos que se utilizan para ejecutar los modelos son transaccionales, es decir se usa información específica de cada cliente e información de la tienda que sea percibida por el cliente en su experiencia de compra.

Los resultados obtenidos demuestran que el efecto de canibalización no supera el 4%, obteniendo 3.8% Plaza Norte, 0.8% Plaza Puente, 1.0% La Dehesa, 0% Manquehue, 1.4% Estación Central, siendo la distancia del cliente hacia la tienda y el formato de la misma las variables que tienen mayor influencia en esta decisión. Con respecto al monto de compra, el inaugurar una nueva tienda incrementa el monto promedio de los clientes, incrementándose la recaudación total de la cadena.

Agradecimientos

Quisiera agradecer a mis padres Concepción y Alexander, como también a mis hermanos Alexander, Francisco, Pablo, Cristian y a mi polola Carla por todo este tiempo de contante apoyo, paciencia, cariño y consejos. Ustedes han sido fundamentales en todos los desafíos que he emprendido a lo largo de mi vida y siempre son la luz que va mostrando el camino.

Además, quiero agradecer a mi profesor guía, Ricardo Montoya, por su gran ayuda en el desarrollo de esta tesis, entregándome siempre el conocimiento y la paciencia necesaria para finalizar este trabajo de buena forma.

Por último, un saludo especial a todos mis amigos que hicieron de esta experiencia universitaria, un recuerdo inolvidable. Estoy convencido que esta amistad perdurarán para siempre.

Muchas gracias.

Tabla de contenido

Resumen.....	2
Agradecimientos.....	3
Índice Tablas.....	6
Índice Ilustraciones.....	8
1) Contexto.....	9
2) Problema a Investigar.....	12
3) Revisión de la literatura.....	14
4) Objetivos.....	16
4.1) Objetivo General.....	16
4.2) Objetivos Específicos.....	16
5) Hipótesis.....	17
6) Marco Teórico.....	18
6.1) Teoría de Utilidad Aleatoria.....	18
6.2) Teoría de Elección Discreta.....	20
6.3) Logit Multinomial.....	20
6.4) Logit Mixed.....	22
6.5) Método Máxima Verosimilitud.....	25
6.6) Enfoque Bayesiano.....	27
6.7) Markov Chain Monte Carlo (MCMC).....	28
6.8) Regresión Lineal Bayesiano.....	30
6.9) Muestreo de Gibbs.....	32
7) Base de Datos.....	35
7.1) Muestra.....	42
8) Modelo Incidencia de Compra.....	47
8.1) Metodología.....	48
8.2) Formato Base de Datos Utilizados.....	49
8.3) Formulación Modelo Logit Mixed.....	53
8.4) Resultados y Análisis.....	56
8.4.1) Variables Descriptivas y Parámetros.....	57
8.4.2) Preferencia de los Consumidores.....	60
8.4.3) Canibalización.....	68

8.4.4) Hit Rate	71
9) Modelo Monto de Compra.....	73
9.1) Metodología	73
9.2) Formulación Modelo de Regresión Lineal Bayesiano	74
9.3) Formato Base de Datos Utilizados	75
9.4) Resultados y Análisis.....	76
10) Conclusiones Generales	79
11) Simulación tienda Costanera Center v/s Parque Padre Hurtado.....	81
12) Bibliografía.....	84
13) Anexos	87

Índice Tablas

Tabla 1: Participación de Mercado según ventas, Tiendas por Departamento, Año 2012.	100
Tabla 2: Distribución de transacciones base de datos original.	36
Tabla 3: Distribución de transacciones base de datos final.	39
Tabla 4: Participación poblacional desde 2002 hasta 2009.	53
Tabla 5: Codificación alternativas (tiendas Falabella).	544
Tabla 6: Datos estadísticos variables explicativas Mixed Logit.	577
Tabla 7: Parámetros variables explicativas por zonas RM Mixed Logit.	5959
Tabla 8: Comunas por zona RM.	600
Tabla 9: Parámetros variables descriptivas por genero Mixed Logit.	600
Tabla 10: Interceptos tiendas Falabella.	622
Tabla 11: Interceptos tiendas Falabella zona norte.	633
Tabla 12: Interceptos tiendas Falabella zona poniente.	64
Tabla 13: Interceptos tiendas Falabella zona sur.	655
Tabla 14: Interceptos tiendas Falabella zona centro.	666
Tabla 15: Interceptos tiendas Falabella zona oriente.	677
Tabla 16: Interceptos tiendas Falabella zona sub urbana.	688
Tabla 17: Canibalización tiendas Falabella.	6969
Tabla 18: Distancia Euclidiana entre tiendas Falabella.	700
Tabla 19: Hit Rate.	711
Tabla 20: Hipótesis signos variables explicativas modelo monto.	755
Tabla 21: Abstracto base de datos modelo Monto.	755
Tabla 22: Parámetros variables explicativas modelo Monto.	766
Tabla 23: Estadístico t-probability modelo Monto.	766
Tabla 24: Signos variables explicativas modelo Monto.	777
Tabla 25: Comparativo resultados v/s base de datos (monto promedio).	788

Tabla 26: Comparativo resultados v/s base de datos (monto total).	788
Tabla 27: Datos variables para simulación Costanera Center v/s La Reina.....	822
Tabla 28: Datos variables modelo monto para Costanera Center.	833

Índice Ilustraciones

Ilustración 1: Distribución principales tiendas Retail dentro Región Metropolitana.....	10
Ilustración 2: Línea de Tiempo Nuevas Tiendas Falabella.....	11
Ilustración 3: Gráfico histograma Transacciones Mensuales Base Datos Final.....	38
Ilustración 4: Gráfico Distribución de montos por tienda Falabella.....	40
Ilustración 5: Gráfico Número de tiendas por cada consumidor.....	41
Ilustración 6: Gráfico Distribución de compras por comuna (principales 12).....	41
Ilustración 7: Gráfico Número compras realizadas por cliente.....	43
Ilustración 8: Gráfico Participación tiendas.....	44
Ilustración 9: Gráfico Número de tiendas por cada consumidor.....	44
Ilustración 10: Gráfico Distribución de compras por comuna (principales 10).....	45
Ilustración 11: Gráfico Distribución de compras y montos desde Noviembre 2002 hasta Junio 2009.....	46

1) Contexto

El Retail en Chile, durante los últimos años ha tomado un dinamismo importante, el cual se ve reflejado en las expansiones internacionales que han realizado los distintos integrantes de esta industria. La razón de este crecimiento se debe a que estas empresas cubren una gran diversidad de las necesidades que tienen las personas, tales como alimentación, vestuario y calzado, equipamiento para el hogar y artículos de cuidado personal. Complementando la razón anterior, el país ha logrado tener una estabilidad económica que ha fomentado el consumo interno y el desarrollo de créditos de consumo, por ende la industria del Retail se ha convertido en un sector altamente competitivo siendo supermercados, multitiendas, tiendas por departamento, ferreterías y farmacias las principales facetas que adquieren estos actores.

Tiendas por Departamento

Las tiendas por departamento son establecimientos de grandes dimensiones que ofertan una variedad de productos enfocados a cubrir una amplia gama de necesidades, las cuales están ubicadas principalmente en Santiago y ciudades importantes del país debido a su tamaño poblacional y desarrollo económico.

Al momento de ver los beneficios de estas empresas, se aprecia que tienen una significativa ventaja en costos (economías de escala) pues al tener una elevada cantidad de ventas, les proporciona un poder de negociación relevante con los proveedores.

Competencia

El sector de tiendas por departamento tiene una elevada concentración, dominada por 3 actores principales, los cuales compiten en distintos formatos y países en donde mantienen presencia. Dentro de Chile, estos actores son: Falabella (37%), Cencosud (27%), Ripley (23%) y La Polar (7%)¹.

¹ Fuente: Sector Retail Latinoamericano, Reporte Especial. Fitch Ratings. Febrero 2013.

Tabla 1: Participación de Mercado según ventas, Tiendas por Departamento, Año 2012.

Tiendas	N° Tiendas	M2/Tienda	Ventas (Miles de USD)	Participación
Falabella	41	6.110	2.204.000	37%
Ripley	40	6.384	1.342.477	23%
Cencosud	77	4.846	1.611.315	27%
La Polar	41	4.128	415.463	7%
Hites	14	6.214	336.404	6%

Por último, se tiene que la distribución espacial dentro de la región metropolitana es muy homogénea, es decir al considerar las principales compañías, la ubicación de las tiendas están en las mismas comunas lo que representa un aumento significativo en la competitividad del sector y una centralización de la oferta.

Ilustración 1: Distribución principales tiendas Retail dentro Región Metropolitana.



Falabella

Falabella (S.A.C.I. Falabella) es un holding chileno, el cual comercializa variados productos bajo el concepto de tiendas por departamento, homecenters, supermercados e hipermercados. A su vez, tiene presencia en Chile, Argentina,

Perú y Colombia, siendo una sociedad anónima cerrada hasta 1996 cuando abre sus transacciones en la Bolsa de Comercio de Santiago².

A su vez, Falabella es un actor principal dentro del país contando con 41 tiendas por departamento, 67 tiendas de mejoramiento del hogar, 31 supermercados, 11 malls, 4 power centers rentas Falabella, 2 shopping centers rentas Falabella, 2,2 millones de cuentas CMR activas, 48.000 empedados, 6.593 USD millones de ingreso de explotación y USD\$ 3.458 millones de colocaciones brutas³.

Ahora bien, al analizar el posicionamiento que ha logrado tener Falabella durante los últimos 10 años, se aprecia que las instalaciones de nuevas tiendas dentro de la región metropolitana se produjeron en mayor cantidad hasta el 2000, en donde con el transcurso del tiempo paulatinamente se han ido agregando otros locales:

- 2003: Plaza Norte.
- 2005: Plaza Puente y La Dehesa.
- 2008: Manquehue y Estación Central.

En la Ilustración N° 2 queda de manifiesto la velocidad de incorporaciones de nuevas tiendas.

Ilustración 2: Línea de Tiempo Nuevas Tiendas Falabella.



² Fuente: Memoria Falabella 2010.

³ Fuente: Memoria Falabella 2010.

2) Problema a Investigar

Como se señala en el contexto actual del mercado de tiendas por departamento, los participantes de la industria tienen una fuerte rivalidad por conseguir que los consumidores vayan a sus tiendas y no a la competencia, en donde la principal iniciativa para lograrlo es la apertura de nuevas instalaciones que ayuden a este objetivo. Adicionalmente a lo anterior, otro beneficio de ingresar nuevas tiendas es impedir que instalaciones de la competencia se ubiquen en lugares cercanos a las tiendas actuales de la cadena de retail impidiendo que éstas se vean perjudicadas en sus ventas.

Para el caso particular de Falabella, hasta hace unos años atrás se tenía una política de localización de tiendas desagregada, es decir no colocaba dos tiendas dentro de un perímetro cercano (influencia) pues se buscaba abarcar la mayor cantidad de comunas posibles dentro de la región metropolitana. En este momento, la situación ha cambiado porque se busca modificar la política actual de instalación de tiendas a través de instaurar una estrategia de aperturas agregada, es decir abrir nuevas tiendas en las mismas zona de influencia de aquellas tiendas de Falabella ya establecidas.

La finalidad de instalar una nueva tienda es incrementar los ingresos y ganancias de la cadena en su conjunto, a través de generar nuevas ventas. Las principales razones que producen nuevas ventas son:

- Aumentar la participación de mercado, captando nuevos clientes que habitualmente compran en tiendas de la competencia.
- Aumentar la penetración de mercado, incrementando las ventas de los actuales clientes.

Con esta nueva visión de localización de tiendas agregada, existe la posibilidad de que algunos clientes antiguos trasladen sus compras desde su tienda de Falabella habitual hacia la nueva tienda de la misma cadena, produciéndose un efecto llamado “canibalización”⁴.

⁴ Efecto Canibalización: Traslado de compras de un consumidor de una tienda antigua hacia una tienda nueva de la misma cadena, que se encuentra dentro de la misma zona de influencia.

Este efecto de canibalización se produce tanto en la incidencia de compra, es decir en la elección de dónde el cliente va a comprar, como en la variación de los montos de las transacciones al dividir el dinero gastado en más de una tienda. Esta situación produce que los ingresos canibalizados desde la nueva tienda no proporcionen mejoras en los ingresos totales de la cadena porque se obtiene una suma cero de ganancias al considerar todo el conjunto de tiendas, debido a que esa “compra” ya se realizaba de igual forma antes de la nueva apertura.

Cabe señalar, que los ingresos a través de la canibalización pueden ser beneficiosos para la cadena Falabella siempre y cuando el monto de compra habitual de aquellos clientes antiguos aumente en la nueva tienda. Por esta razón adquiere importancia no sólo la incidencia de compra sino también la variación en el monto de la misma.

Considerando la información planteada en los párrafos anteriores, el problema que se aborda en el presente trabajo de tesis es identificar el cambio en el comportamiento de compra de los clientes actuales de Falabella al ingresar una nueva tienda de la misma cadena, lo cual se realiza a través de dos puntos principales: En primer lugar, medir el efecto de canibalización asociado a la incidencia de compra a través de identificar los factores que determinan la decisión de compra de los cliente al tener un conjunto de elección finito y que varía con el tiempo, es decir obteniendo una descripción y caracterización de la decisión de compra que se produce al incorporar una nueva tienda en un lugar cercano a otra de la misma cadena, situada en el mismo radio de influencia. En segundo lugar, se busca analizar la influencia de estas nuevas incorporaciones (tiendas) en el monto de compra que realizan los clientes. Para el trabajo se considera sólo el canal físico de venta y no otras vías de venta, tales como canales virtuales o por catálogo.

Un valor agregado que se incorpora en este trabajo es la utilización de información transaccional de los clientes, es decir se busca añadir datos desagregados de ellos, incorporando la información obtenida de las mismas boletas de venta de las personas para realizar análisis que ayuden a la toma de decisiones estratégicas para la cadena, como es la localización de una nueva tienda. En otras palabras, basándose en que se quiere obtener una descripción y caracterización del comportamiento de los clientes asociado a la decisión de compra, se hace imperiosa la utilización de datos individuales de los clientes, y no sólo datos agregados de la tienda, para determinar el comportamiento de éstos. Además, queriendo enfatizar con mayor fuerza este concepto, se utilizan modelos bayesianos con el afán de obtener parámetros para cada uno de los clientes y no parámetros globales para toda la muestra poblacional.

3) Revisión de la literatura

Al momento de revisar la literatura relacionada al efecto de canibalización, tanto en incidencia como monto de compra, se aprecia que no es un tema abarcado como punto principal de los estudios sino más bien como una arista adicional a otros fenómenos. Lo anterior puede explicarse debido a que la canibalización producida al agregar una nueva tienda de la misma cadena en el mismo lugar de influencia de las antiguas tiendas es un tema que ha sido tomando en cuenta sólo en los últimos años, ya que con anterioridad se tomaban decisiones sobre dónde colocar la nueva instalación basado en pronósticos de demanda o asesoramiento de expertos relacionados a este tema, sin realizar estimaciones de la canibalización a través de estudios específicos para este fenómeno.

Al tener foco sobre el tipo de información que se utiliza en los estudios realizados hasta el día de hoy, se encuentran trabajos basados principalmente en información agregada de las empresas o locales (Koen y Neslin 2006, Gupta 2008, Kumar, Pancras y Sriram 2009), tales como ventas totales de las tiendas, montos promedios, por nombrar algunos.

Ahora bien, considerando la literatura asociada a la incidencia de compra producida con el ingreso de nuevas instalaciones de venta, los primeros modelos utilizados fueron auto-regresivos, los cuales utilizan información histórica de las cadenas. Luego se dio paso a modelos basados en series temporales (Hanssens, Parsons y Schults, 2001). Otra línea de modelos matemáticos que fueron usados para ver de forma anexa la canibalización, fueron aquellos modelos de elección discreta para estudiar preferencias en los consumidores (Craig 1984; Fotheringham 1991, Rust y Donthu 1995).

Tomando en cuenta los resultados que obtuvieron los distintos investigadores al analizar la canibalización de incidencia de compra, específicamente Caravella, Deighton, Jill y Steenburgh (2008), se tiene un consenso general en que los valores de este fenómeno no son elevados, teniendo cifras inferiores al 12% en las distintas estimaciones.

Al revisar los trabajos realizados sobre la influencia que tiene la apertura de una nueva tienda de la misma cadena sobre los montos de compra de los clientes actuales, los modelos de monto en su mayoría de las veces son desarrollados a través de regresiones lineales con enfoque frecuentista a través de variables exógenas para describir la función de utilidad (Ansari, Mela y Neslin, 2008,

Boatwright, Borle, Kadame, Nunes y Shmueli 2005, Neslin, Henderson y Quelch 1985, Paull 1978).

Como se mencionó inicialmente, el tipo de dato que se ha utilizado hasta el momento para realizar los estudios, han tenido una dirección clara hacia los datos agregados de las cadenas o industrias analizadas, usándose principalmente la variable de monto promedio (Gupta 1988). Otra información muy relacionada al monto de compra es el tiempo que transcurre entre una compra y la siguiente, tomándose en cuenta 2 acciones: Primero que al ingresar una nueva tienda, la frecuencia de compra aumenta (Neslin y Pauwels, 2006), es decir el tiempo entre compras sucesivas es menor, y en segundo lugar al tener intervalos de tiempo menores entre cada compra consecutiva, el monto de compra disminuye (Boatwright, Borle, Kadame, Nunes y Shmueli 2005).

Habiendo revisado la literatura y estudios relacionados a incidencia de compra, monto de compra y canibalización de forma anexa, se aprecia que no se han realizado estudios exhaustivos en donde se desarrollen modelos específicos para este fenómeno y tampoco se ha utilizado información más detallada de las industrias utilizadas. Este escenario genera una posibilidad importante para confeccionar modelos matemáticos específicos y con información desagregada (datos individuales de los clientes) que pongan en un foco principal la canibalización, especialmente para la canibalización de incidencia de compra.

4) Objetivos

4.1) Objetivo General

- Cuantificar el efecto de canibalización al ingresar 5 nuevas tiendas de Falabella en la región metropolitana dentro del mismo radio de influencia de las tiendas antiguas de la cadena.
- Cuantificar la variación en los montos de compra al ingresar 5 nuevas tiendas de Falabella dentro de la región metropolitana en mismo radio de influencia de las tiendas antiguas de la cadena.

4.2) Objetivos Específicos

- Describir y caracterizar la decisión de compra que se produce al incorporar una nueva tienda, determinando los factores influyentes en el efecto de canibalización asociada a la incidencia de compra.
- Identificar las preferencias (atractivo) de los consumidores con respecto a las tiendas de Falabella, tanto antiguas como nuevas.
- Identificar los factores que determinan el monto de compra de los consumidores.
- Aplicar un enfoque bayesiano para los modelos de Incidencia de Compra (Logit Mixed) y Monto de Compra (Regresión Lineal) para obtener parámetros individuales de los clientes.

5) Hipótesis

Teniendo claridad sobre el problema a investigar y habiendo revisado los estudios previos que se han realizado sobre el efecto de canibalización en la incidencia y monto de compra, a continuación se elaboran las hipótesis que se validarán o refutarán a través de los resultados entregados por el trabajo desarrollado en esta tesis.

- El efecto de canibalización en la incidencia de compra para Falabella no supera el 12%, (según la bibliografía y opinión de expertos).
- El efecto de canibalización no es constante en el tiempo y varía según cada nueva apertura.
- La menor distancia geográfica entre las tiendas antiguas de la cadena y la nueva tienda instalada tiene un efecto negativo sobre el efecto de canibalización, es decir una ubicación cercana entre las tiendas de la misma cadena disminuye la canibalización.
- La distancia entre el cliente y la tienda, afecta negativamente a la incidencia de compra.
- El consumidor no realiza variaciones importantes en el monto de compra a lo largo del tiempo, manteniendo un monto relativamente constante.
- Con el ingreso de una nueva tienda al conjunto de elección del cliente, el monto promedio de compra aumenta.

6) Marco Teórico

Parte importante de todo trabajo es el fundamento teórico que lo respalda, buscando desarrollar estudios robustos de los fenómenos abarcados. Debido a que las decisiones asociadas a la elección de cuál tienda visitar y el monto de compra en el que se desea incurrir el cliente, se tiene que estas 2 acciones son actos racionales y están vinculados al beneficio o utilidad que le generará al individuo estas acciones.

Considerando lo anterior, para el correcto desarrollo del estudio planteado es pertinente utilizar como base la Teoría de Utilidad Aleatoria y la Teoría de Elección Discreta. En específico, se utiliza un Logit Mixed para el modelo de incidencia de compra y una Regresión Lineal para el modelo de monto de compra, en donde ambos son abarcados bajo la teoría bayesiana que considera los parámetros de la función de utilidad aleatorios y los datos muestrales constantes.

A continuación se presenta una descripción teórica de los principales conceptos que se utilizan para realizar los modelos de incidencia de compra y monto de compra.

6.1) Teoría de Utilidad Aleatoria

Según Domencich y McFadden (1975) y Williams (1977), la Teoría de Utilidad Aleatoria plantea que las personas son racionales, lo cual significa que al momento de realizar una elección dentro de un conjunto de alternativas relevantes, escogen la opción que maximiza su utilidad neta. Adicionalmente, se asume que los individuos pertenecientes a una población homogénea, actúan racionalmente y poseen información perfecta.

Esta teoría está definida por una población homogénea y un conjunto de alternativas, I y $A = \{A_1, \dots, A_j, \dots, A_J\}$ respectivamente, en donde cada una le produce una utilidad U_{ij} al individuo "i" al elegir la alternativa "j". A su vez, la persona escogerá "j" tal que su utilidad sea la mayor posible. Con el término "Utilidad" se cuantifica el nivel de satisfacción que le produce a la persona cierta alternativa en particular, a través de una función compuesta por una combinación adecuada de atributos y características de esta última. Modelando lo anterior, queda de la siguiente forma:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (1)$$

Se aprecia que la utilidad está conformada por dos componentes principalmente, una parte determinística (V_{ij}) y otra aleatoria (ε_{ij}) que no se observa. Otra interpretación con respecto a la parte aleatorio ε_{ij} , es que representa la diferencia entre la utilidad real y la utilidad observada por el individuo.

Un supuesto que realiza esta teoría es que los errores son independientes, lo cual significa que la porción no observable de la utilidad para una alternativa es irrelevante a la parte aleatoria de la utilidad de otra, entendiéndose que la información del error de una alternativa no proporciona información sobre el error de otra. Además, es aceptable asumir que los errores ε tienen media cero sin pérdida de generalidad, como también que son independientes e idénticamente distribuidos (iid) para cada alternativa.

Tomando en cuenta la parte determinística de la función de utilidad (V_{ij}), el objetivo que tiene el investigador al momento de definir esta función es realizar una formulación suficientemente completa y consistente como para que el error asociado a la misma sólo sea “ruido blanco”, es decir que la formulación sea capaz por si sola de capturar toda la información necesaria para describir en su totalidad al efecto estudiado.

Como se menciona anteriormente, se tiene que el individuo escoge dentro del conjunto de elección la alternativa que le proporciona una mayor utilidad personal, entonces esta situación queda representada en la siguiente probabilidad:

$$P_{ij} = \text{Probabilidad} (U_{ij} > U_{il}) \quad \forall l \in A \quad (2)$$

Reemplazando (1) en la ecuación (2) y despejando el error, se obtiene que la probabilidad es:

$$P_{ij} = \text{Probabilidad} (\varepsilon_{il} \leq V_{ij} - V_{il} + \varepsilon_{ij}) \quad \forall l \in A \quad (3)$$

El desarrollo algebraico anterior, indica que la derivación de un modelo en particular a partir de la Teoría de Utilidad Aleatoria depende de la distribución del error ε , es decir al escoger una distribución específica para los errores se obtiene una formulación concreta del modelo. Específicamente, cuando se ocupa la distribución Logística se esta hablando de un modelo Logit y cuando se usa la distribución Normal se esta haciendo referencia a un modelo Probit. La elección de cuál modelo utilizar dependerá de las características del fenómeno a estudiar.

6.2) Teoría de Elección Discreta

El objetivo de la Teoría de Elección Discreta es describir el comportamiento de un individuo cuando éste debe escoger entre un conjunto de alternativas mutuamente excluyente, es decir se tiene un número finito de opciones y sólo se puede elegir una de ellas. La decisión está basada en la teoría de utilidad aleatoria en donde se elige aquella alternativa que le proporciona una mayor utilidad a la persona.

Los modelos de elección discreta fueron estudiados por Amemiya (1981), McFadden (1984), Maddala (1983) y Dhrymes (1984) en donde existen dos grandes grupos, los cuales son los modelos Binomiales (Logit, Probit y Utilidad Aleatoria) y los Multinomiales (Logit Multinomial, Logit Jerarquico y Logit Mixed), siendo la diferencia entre estos la cantidad de valores que puede tomar la variable dependiente.

6.3) Logit Multinomial

El modelo de elección discreta Logit Multinomial se caracteriza por una variación sistemática de los gustos de los individuos de la muestra, en donde el término de error ε_{ij} se distribuye idéntica e independientemente Gumbel. Considerando la distribución del error, la probabilidad que el individuo “i” escoja la alternativa “j” es la siguiente:

$$P_{ij} = Prob (Y_i = j) = \frac{e^{\mu V_{ij}}}{\sum_{j=1}^J e^{\mu V_{ij}}} \quad (4)$$

En donde el término $\mu > 0$ es el parámetro de Gumbel el cual es inversamente proporcional a la desviación estándar del componente aleatorio:

$$\mu = \frac{\pi}{\sqrt{6\sigma_{ij}}} \quad (5)$$

Como mencionan Hensher, Louviere y Swait (1999), el vector de parámetros que se estima en este modelo es en realidad $\mu\beta$ al reemplazar la ecuación (4) en la probabilidad. A su vez, McFadden (1974) demuestra que la función de probabilidad logarítmica con esas probabilidades de elección son globalmente cóncavas en los parámetros β , lo cual es beneficioso para la maximización numérica.

Siguiendo esta línea, el modelo Logit Multinomial produce una indeterminación en sus parámetros al ser estimado, por consiguiente la forma de solucionar este inconveniente es normalizar el modelo considerando los parámetros de la primera alternativa iguales a cero, quedando las probabilidades de la siguiente forma:

$$P_{ij} = \frac{1}{\sum_{j=1}^J e^{v_{ij}}} \quad \text{para } j = 1 \quad (6)$$

$$P_{ij} = \frac{e^{v_{ij}}}{\sum_{j=1}^J e^{v_{ij}}} \quad \text{para } j = 2, 3, \dots, J \quad (7)$$

Por último, tomando en cuenta las características del estudio de incidencia de compra, el modelo Logit Multinomial tiene desventajas que son perjudiciales para el trabajo deseado, siendo las siguientes:

- No permite heterogeneidad en los gustos de los individuos.
- Asume que los gustos de un individuo varían sistemáticamente, no permitiendo cambios aleatorios.
- La propiedad de Independencia de Alternativas Relevantes, la cual es un impedimento al tener alternativas correlacionadas.
- No permite el manejo de situaciones en donde la parte aleatoria de las utilidades estén correlacionadas.

- Los coeficientes estimados son fijos para toda la población.
- Se asumen independientes las respuestas repetidas de los individuos.

6.4) Logit Mixed

La particularidad del modelo Logit Mixed es que considera más de una componente aleatoria, es decir toma en cuenta el error con una distribución Gumbel iid como lo hace el Logit Multinomial y agrega otra componente que permite modelar correlaciones y heteroscedasticidad⁵ en los términos de error. El error adicional δ_{ij} es función de datos observables de cada alternativa, lo cual permite estimar correlaciones y heteroscedasticidad.

El beneficio que tiene este modelo es el aumento de generalidad para desarrollar el estudio pero trae consigo la pérdida de simpleza en su estimación, ya que se hace imperiosa la necesidad de utilizar herramientas de simulación para estimar los parámetros de la función de utilidad.

Escribiendo lo anterior como función de utilidad, queda de la siguiente forma:

$$U_{ij} = V_{ij} + \delta_{ij} + \varepsilon_{ij} \quad (8)$$

Teniendo como base el modelo Logit, se tiene que el error ε_{ij} sigue una distribución Gumbel iid y el término δ_{ij} sigue cualquier función de distribución que esté definida por una densidad $f(\delta/\theta^*)$ ⁶. Entonces la probabilidad condicional en δ que tiene la persona "i" al seleccionar la alternativa "j" es:

$$P_i(i/\delta) = L_{ij}(\delta) = \frac{e^{V_{ij} + \delta_{ij}}}{\sum_{j=1}^J e^{V_{ij} + \delta_{ij}}} \quad (9)$$

⁵ Distinta varianza.

⁶ θ son parámetros fijos que describen la función de distribución (Ejemplo: media y varianza).

En consecuencia, la probabilidad de elegir la alternativa “j” por el individuo “i” es la integral de la probabilidad condicional sobre todos los valores posibles de δ . Quedando representada por la siguiente formula:

$$P_{ij} = \int L_{ij}(\delta) f(\delta/\theta^*) d\delta \quad (10)$$

Ahora bien, se puede detallar de mejor forma el término aleatorio agregado a través de la incorporación de parámetros y datos observados asociados a la alternativa elegida:

$$\delta_{ij} = \mu_i z_{ij} \quad (11)$$

Siendo μ_i ciertos parámetros y z_{ij} los datos observados relacionados a la alternativa. Lo relevante de esta descripción es que el término μ es propio del individuo y no varía entre las alternativas (Ben Akiva y Bolduc, 1996; Brownstone y Train, 1999).

Por otro lado, al enfocarse en la parte observable de la utilidad, ésta puede adquirir información tanto de la alternativa elegida como del individuo. Siendo X_{ij} el vector de atributos de la alternativa “j” que son planteados por el individuo “i” y S_i el conjunto de características propias de la persona que toma la decisión, la parte determinística de la función de utilidad queda de la siguiente forma:

$$V_{ij} = V(X_{ij}, S_i) \quad (12)$$

Generalmente la forma de esta función es lineal en sus variables explicativas relacionadas a la alternativa, constituyéndose por atributos medibles y observables.

$$V_{ij} = \sum_{k=1}^K \beta_i X_{ikj} \quad (13)$$

Revisando la ecuación (13), los parámetros β que acompañan a las variables explicativas X_{ij} representan el “peso” que tiene cada una sobre la utilidad, teniendo relevancia el signo de cada uno porque esta información indica la dirección de influencia que tiene esa variable con respecto a la variable dependiente.

Es importante mencionar que los únicos parámetros que pueden ser estimados son los que capturan diferencias entre las alternativas, es decir sólo se estiman las diferencias y no los valores absolutos de las opciones. Lo anterior indica que los términos que no varían de una alternativa a otra desaparecen de la probabilidad.

A su vez, a la función de utilidad se le puede agregar una constante específica C_{ij} para cada alternativa, la cual captura el efecto promedio de todos los factores que no están incluidos en la ecuación. Con esta modificación también se altera la parte aleatoria del modelo dejándola con media cero por construcción. La nueva función de utilidad se muestra en la ecuación 14.

$$V_{ij} = \sum_{k=1}^K \beta_i X_{ikj} + C_{ij} \quad (14)$$

Para completar aún más la función de utilidad determinística, se pueden agregar características propias del individuo. Para ingresar este tipo de variables, que son iguales para todas las alternativas, existen dos formas:

1. Tomar en cuenta estas variables personales de forma separada de los atributos de las alternativas:

$$V_{ij} = \sum_{k=1}^K \beta_i X_{ikj} + \sum_{q=1}^Q \theta_{ij} Y_{iq} + C_{ij} \quad (15)$$

En esta opción, el parámetro θ captura el efecto de variación en la característica de la persona sobre la utilidad. Para estimar los parámetros mencionados, al igual que en el caso anterior, se debe normalizar el modelo llevando a cero uno de ellos.

2. Introducir las características del individuo relacionándolas con los atributos de las alternativas, ya sea a través de una división o multiplicación dependiendo de si son directa o indirectamente proporcionales. Suponiendo ambos tipos de relación mencionadas con dos atributos y dos características:

$$V_{ij} = \frac{X_{i1j}}{Y_{i1}} + X_{i2j}Y_{i2} + C_{ij} \quad (16)$$

Para este caso, no es necesario normalizar el modelo pues las variables personales afectan las diferencias en las utilidades a través de su interacción con los atributos de las variables.

Al enfocarse en la estimación de los parámetros y constantes que tiene la función de utilidad del individuo, se tiene la limitación de calcularlas todas porque existe un número infinito de valores que producen la diferencia entre ellas y que resultan en la misma probabilidad de elección. A modo de solución para lo planteado anteriormente, el modelo se debe normalizar para poder estimar correctamente los parámetros, lo cual se lleva a cabo colocando uno de ellos en cero, siendo irrelevante cuál parámetro se fija.

Otra característica muy importante del Logit Mixed para el estudio de aperturas de nuevas tiendas es la propiedad de Independencia de Alternativas Relevantes, lo cual plantea que la inclusión o exclusión de una opción del conjunto de elección no cambia la proporción que existe entre las restantes alternativas, es decir que la probabilidad de elegir una alternativa es independiente de la existencia de las demás.

Por último, Hausman y McFadden (1984) plantean que omitir un conjunto de alternativas verdaderamente irrelevantes no significará cambios sistemáticos en los estimadores de los parámetros, sólo conlleva ineficiencias pero no inconsistencias. Si la división de probabilidades omitidas no son realmente independientes de estas alternativas, los estimadores de los parámetros que se obtienen al eliminar estas alternativas son inconsistentes.

6.5) Método Máxima Verosimilitud

El método de Máxima Verosimilitud consiste en calcular los parámetros de la función de utilidad determinística con el fin de, basados en una muestra de la población, maximizar la probabilidad de que lo predicho por el investigador sea igual a lo observado en los datos. Por otro lado, este método entrega un instrumento de elección de un estimador asintóticamente eficiente de un vector de parámetros.

Al momento de analizar las características de este método, una de las principales cualidades que tiene es que al considerar una utilidad lineal, se tiene un buen comportamiento de éstos ya que se obtiene un máximo único.

Asumiendo la independencia de las observaciones y en donde P_{ij} significa la probabilidad de que el cliente “i” elija la alternativa “j”, la ecuación que representa la verosimilitud de una de ellas es la siguiente:

$$L = \prod_{ij} P_{ij}^{\delta_{ij}} \quad , \text{con } \delta_{ij} = \begin{cases} 1, & \text{si } i \text{ elige } j. \\ 0, & \text{si no.} \end{cases} \quad (17)$$

Luego, considerando la muestra completa y todas las alternativas, la función de verosimilitud señalada en la ecuación 17 queda de la siguiente forma:

$$\text{Max } L(\theta) = \prod_{i=1}^I \prod_{j=1}^J P_{ij}^{\delta_{ij}} \quad (18)$$

Es común maximizar el logaritmo de la función anterior pues tiene mejor maniobrabilidad, disminuye los tiempos computacionales y es monótonamente creciente, quedando:

$$l(\theta) = \ln(L(\theta)) = \sum_{i=1}^I \sum_{j=1}^J \delta_{ij} * \ln(P_{ij}) \quad (19)$$

Otra característica ventajosa de la verosimilitud logarítmica es su característica de concavidad global, lo cual es importante para la maximización, pero una desventaja es que el número de parámetros de este modelo aumenta con el número de alternativas por lo que se pueden generar complicaciones si la cantidad de regresores es elevado.

Al realizar la maximización de la verosimilitud, se obtiene el Estimador Máximo Verosímil, el cual cumple con las siguientes características:

- Es consistente, es decir a medida que aumenta el tamaño de la muestra también se produce un acercamiento al valor del parámetro real.

- Es asintóticamente normal, hace referencia a que la ley aproximada de los estimadores es la normal.
- Es asintóticamente eficiente, lo cual significa que es el estimador insesgado de menor varianza.

Las características mencionadas reafirman el hecho que la maximización tiene solución única.

6.6) Enfoque Bayesiano

El enfoque bayesiano está basado en el Teorema de Bayes, el cual posibilita la inclusión de información muestral para identificar la distribución del parámetro que se quiera estimar. La visión que plantea esta teoría aborda la incertidumbre presente en un modelo representándola a través de una distribución de probabilidad sobre los posibles valores de los parámetros. Comparando lo señalado con el enfoque frecuentista, tradicionalmente utilizado en los estudios anteriores a este trabajo, se tiene una total diferencia porque esta última teoría considera fijos los parámetros y variables los datos muestrales de la investigación.

En otras palabras, el enfoque bayesiano considera constantes los datos poblacionales y variables los parámetros a estimar, es decir se asumen aleatorios los parámetros. La representación matemática de lo planteado por Bayes se aprecia en la ecuación 20.

Adicionalmente, la utilización de modelos bayesianos se usan en trabajos donde se busca obtener parámetros asociados a cada individuo de la muestra poblacional y no sólo un parámetro general para todos los individuos.

$$P(\text{parametros} \setminus \text{datos}) = \frac{P(\text{datos} \setminus \text{parametros}) \text{Prob}(\text{parametros})}{P(\text{datos})} \quad (20)$$

Ahora bien, tomando la ecuación 20 y considerando constantes los datos, la probabilidad condicional que plantea la teoría bayesiana queda de la siguiente forma:

$$P(\text{parametros} \setminus \text{datos}) \propto P(\text{datos} \setminus \text{parametros})P(\text{parametros}) \quad (21)$$

El término $P(\text{datos} \setminus \text{parametros})$ de la ecuación 21 representa la distribución conjunta de las variables aleatorias observadas, dados los parámetros. A su vez, el término $P(\text{parametros})$ de la misma ecuación, identifica las creencias previas del investigador (densidad priori). Por el otro lado, el término $P(\text{parametros} \setminus \text{datos})$ es la densidad posterior de los parámetros dada la estructura actual de los datos o de las creencias sobre la distribución de los parámetros luego de utilizar la información muestral.

La densidad posterior es una mezcla entre la información previa y la información actualizada de los datos, en donde esta densidad está constantemente siendo actualizada por los nuevos datos que se obtengan. Es decir, este principio implica una acumulación continua de conocimientos acerca de los parámetros.

Una de las principales dificultades de este enfoque está en la compleja aplicación de éste debido al cálculo de integrales múltiples que lo hacen intratable analíticamente. Es por ello, que se utilizan métodos numéricos eficientes para calcular estas integrales en varias dimensiones, tales como herramientas de simulación, el método de Aproximación de Laplace y el método de Monte Carlo.

Las dos mayores ventajas que tienen los métodos mencionados en el párrafo anterior, son que para resolver las integrales múltiples no requieren maximizar una función, en donde los estimadores obtenidos por ellos cumplen propiedades deseables bajo condiciones más relajadas.

6.7) Markov Chain Monte Carlo (MCMC)

La metodología de Markov Chain Monte Carlo es una simulación no determinística, basada en conceptos estocásticos discretos de las Cadenas de Markov. En simples palabras, la teoría de cadenas de Markov plantea que la probabilidad de que ocurra un evento depende exclusivamente del acontecimiento producido inmediatamente anterior, es decir no considera la información histórica del evento.

La utilización de este método en el enfoque bayesiano es para encontrar soluciones aproximadas a integrales complejas asociadas a las densidades de probabilidad sin tener que calcularla directamente, en donde el planteamiento central se basa en que si se tiene una muestra aleatoria suficientemente grande, a través de un histograma es posible aproximar la densidad de probabilidad (posterior) deseada. Adicionalmente, uno de los beneficios principales de este tipo de métodos, es que no se necesita maximizar una función, a diferencia del método de máxima verosimilitud, para obtener los resultados.

Para concluir el razonamiento mencionado, Metropolis, et al. (1953) demuestra que se puede construir un proceso estocástico de Cadena de Markov para un parámetro que se desarrolla a través del tiempo de tal manera que:

- 1- Tenga el mismo conjunto de valores posibles.
- 2- Sea fácil de simular.
- 3- La distribución estacionaria que se utiliza para extraer muestras de la densidad de probabilidad después de la Cadena de Markov sea ejecutada por un periodo largo de tiempo.

Dado lo anterior, es posible construir y ejecutar una cadena de Markov para un número elevado de iteraciones, teniendo como fin el desarrollo de un conjunto de datos (muestra) del parámetro desde una distribución posterior y usar estadísticas descriptivas simples para examinar rasgos que fuesen relevantes.

Una generalización de la metodología MCMC la realiza Hasting (1970), el cual sugiere que mediante la obtención de una secuencia de muestras aleatorias de una distribución a priori, al realizar un número elevado de iteraciones la distribución de las muestras generadas corresponde a la distribución real. Este método se denomina Metropolis-Hasting y se utiliza principalmente para probar distribuciones multidimensionales.

En otras palabras, dado un valor inicial β_0 se contruye una cadena mediante el reconocimiento de que cualquier Cadena de Markov que ha encontrado un camino al estado β_t puede ser completamente caracterizada por la distribución de

probabilidad para el tiempo $t + 1$. El algoritmo de este enfoque se basa en una distribución candidata, $f(\beta|\beta_t)$ para el tiempo $t + 1$, dado que tenemos β_t .

Considerando una extracción de la distribución propuesta como un valor candidato β^* y siendo $f(\beta^*|\beta_t)$ una densidad condicional a priori que sea fácil de mostrar, entonces se tiene la siguiente condición de aceptación:

- 1- El valor candidato es aceptado como $\beta_{t+1} = \beta^*$ con probabilidad (α_H):

$$\alpha_H(\beta_t, \beta^*) = \min \left[1, \frac{p(\beta^*)f(\beta_t|\beta^*)}{p(\beta_t)f(\beta^*|\beta_t)} \right] \quad (22)$$

- 2- Si no, $\beta_{t+1} = \beta_t$, quedándose con el actual valor de β .

El ratio de aceptación indica el grado de probabilidad de que la nueva muestra es la distribución deseada, es decir se elige el candidato β^* con probabilidad α_H o se mantiene el valor β_t con probabilidad $1 - \alpha_H$. Además, si el valor candidato se mueve a un punto más probable que el punto existente (un punto en la región de densidad más alta), siempre aceptará el movimiento.

A su vez, esta situación puede ser vista como un evento de Bernoulli en donde con probabilidad α_H se obtiene el evento A y realizar un movimiento para $\beta_{t+1} = \beta^*$ si se obtiene el evento A, sino $\beta_{t+1} = \beta_t$.

Por último, Hastings demuestra que el extraer muestras representa una cadena de Markov con una distribución de equilibrio correcta capaz de producir muestras desde la densidad posterior que interesa para el estudio.

6.8) Regresión Lineal Bayesiano

El presente capítulo quiere entregar los fundamentos teóricos asociados a una regresión lineal tradicional pero utilizando el enfoque bayesiano que se presenta en secciones anteriores. Para esto, se considera un modelo de regresión lineal como se

presenta en la ecuación matricial 23 y se asume que el error tiene una distribución normal con parámetros señalados en la ecuación 24.

$$y = X\beta + \varepsilon \quad (23)$$

$$\varepsilon \sim N(0, \sigma^2 I_n) \quad (24)$$

Se tiene que el vector y representa las variables dependientes (dimensión $n \times 1$), el vector X representa las variables explicativas (dimensión $n \times k$) y ε representa el vector de error (dimensión $n \times 1$), el cual se asumen variables aleatorias normales idénticamente distribuidas independientes.

Dentro de este modelo, lo que se busca estimar son los parámetros que acompañan a las variables explicativas (β) y la desviación estándar (σ). Para ello, se utiliza una densidad a priori de la forma $p(\beta, \sigma) = p_1(\beta)p_2(\sigma)$, en donde se asume la independencia de los parámetros.

Para el caso de β , se considera una densidad a priori normal, lo cual requiere que se especifique una media y una varianza de forma lineal, representando en la ecuación 23, en donde, r es el vector de medias a priori y T es la matriz que contiene las varianzas y covarianza a priori, $Q'Q = T^{-1}$ y $q = Qr$.

$$p_1(\beta) \propto e^{-\frac{1}{2}(Q\beta - q)'(Q\beta - q)} \quad (25)$$

Para el caso de σ , se asume $p_2(\sigma) \propto (1/\sigma)$ para mayor simplicidad pero todos los resultados siguen una densidad previa gamma conjugada para este parámetro.

Uniendo la información señalada para los parámetros β y σ , se obtiene la función de verosimilitud del modelo bayesiano representado en la ecuación 26.

$$L(\beta, \sigma) \propto (1/\sigma^n) e^{-\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}} \quad (26)$$

A su vez, considerando las primeras densidades a priori, se logra desarrollar la densidad a posteriori de la ecuación 27, en donde, $\hat{\beta}(\sigma)$ y $V(\sigma)$ son la media y varianza respectivamente de la densidad a posteriori, ambas dependientes de σ .

$$p(\beta, \sigma) \propto \left(1/\sigma^{n+1}\right) e^{(\beta - \hat{\beta}(\sigma))' V(\sigma)^{-1} (\beta - \hat{\beta}(\sigma))} \quad (27)$$

$$\hat{\beta}(\sigma) = (X'X + \sigma^2 Q'Q)^{-1} (X'y + \sigma^2 Q'q) \quad (28)$$

$$V(\sigma) = \sigma^2 (X'X + \sigma^2 Q'Q)^{-1} \quad (29)$$

Un punto importante a considerar al momento de estimar estos parámetros, es lo planteado por Theil y Goldberger (1961) que proponen que σ^2 es reemplazado por un valor estimado $\hat{\sigma}^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n - k}$, basado en la estimación de mínimos cuadrados de $\hat{\beta}$. La ventaja de esta solución es que el problema es resuelto a través del método de mínimos cuadrados. Debido a esto, la media de la densidad a posteriori se mantiene igual pero la varianza queda de la forma:

$$V(\hat{\beta}_{TG}) = \hat{\sigma}^2 (X'X + \hat{\sigma}^2 Q'Q)^{-1} \quad (30)$$

6.9) Muestreo de Gibbs

Una de las principales dificultades de la aplicación del enfoque bayesiano es la compleja resolución de integrales múltiples relacionadas a una densidad conjunta, como la señalada en la ecuación 31. En este sentido, un avance importante en la aplicación del enfoque bayesiano es el Generador de Muestras de Gibbs, el cual consiste en una herramienta de simulación para obtener distribuciones marginales⁷ desde una densidad conjunta no normalizada (Gelfand and Smith, 1990).

⁷ Teniendo dos variables aleatorias X e Y, la distribución marginal de X es la distribución de probabilidad de X sin considerar la información referente a Y.

$$f(x, y_1, y_2, \dots, y_k) \quad (31)$$

En otras palabras, el muestreo de Gibbs proporciona una forma de muestra de una densidad de probabilidad multivariada basada sólo en las densidades de los subconjuntos de vectores condicionales.

El Generador de Muestras de Gibbs permite generar extracciones aleatorias de la densidad marginal $f(x)$ sin tener que calcularla (Casella and George, 1992). Es por ello que dada una densidad conjunta $f(x, y_1, y_2, \dots, y_k)$, se busca estudiar las características de la distribución marginal $f(x)$. Al igual que en la metodología MCMC, no se requiere maximizar ninguna función para obtener los parámetros deseados.

Al considerar un número suficientemente grande de iteraciones (extracciones), la observación final converge en el límite a la densidad conjunta real de los parámetros, es decir a pesar de utilizar distribuciones condicionales en el esquema de muestreo, una amplia muestra de ellas puede ser usada para producir inferencias posteriores válidas sobre los parámetros.

$$f(x) = \int_{y_k} \dots \int_{y_1} f(x, y_1, y_2, \dots, y_k) dy_1 \dots dy_k. \quad (32)$$

Enfocándose en las extracciones realizadas por el algoritmo, la secuencia de valores no es independiente sino más bien actúa como una Cadena de Markov⁸, en donde la principal característica es que un evento en particular depende exclusivamente del evento inmediatamente anterior.

Para el caso descrito en el capítulo anterior, la densidad posteriori para β condicionada sobre σ , $p(\beta|\sigma)$, es multivariante normal con media y varianza indicadas en las ecuaciones 28 y 29 respectivamente.

A su vez, la densidad posteriori de σ condicionada sobre β , $p(\sigma|\beta)$, es:

⁸ Una Cadena de Markov es un proceso estocástico discreto en el cual la probabilidad de ocurrencia de un evento en particular depende exclusivamente del evento inmediatamente anterior, es decir son proceso que “no tienen memoria”.

$$\frac{(y-X\beta)'(y-X\beta)}{\sigma^2} \setminus \beta \sim \chi^2(n) \quad (33)$$

Tomando en consideración las dos distribuciones condicionales posteriormente señaladas, el algoritmo computacional del Muestreo de Gibbs es:

1. Se inicia la iteración con un valor arbitrario para los parámetros β^0 y σ^0 .
2. Se calcula la media y la varianza de β utilizando las ecuaciones 28 y 29 condicionales sobre el valor de σ^0 .
3. Se utiliza la media y varianza de β para crear un vector aleatorio normal multivariante llamado β^1 .
4. Se usa el valor de β^1 junto con $\chi^2(n)$ al azar para determinar σ^1 a través de la ecuación 33.

Luego de terminar los 4 pasos anteriores (una iteración), se vuelve al paso 1 considerando como valores iniciales β^1 y σ^1 , en donde se produce un nuevo conjunto de valores β^2 y σ^2 , así sucesivamente.

Por último, el objetivo de la extracción de datos es revisar la distribución a priori para agregarle la nueva información que se obtiene de la muestra, teniendo como resultado una nueva distribución a posteriori. El algoritmo permite analizar la distribución marginal posteriori de β iterando un procedimiento de muestreo entre las dos distribuciones mencionadas.

7) Base de Datos

Como se menciona en la parte inicial del documento, uno de los alcances que tiene este trabajo de investigación es utilizar la información transaccional de los clientes para identificar los factores que inciden en la decisión de compra, tanto sobre el lugar donde asistir como por el monto de la compra, y posteriormente usar esta información para tomar decisiones estratégicas de la empresa asociada a la localización de nuevas tiendas. Es por ello que se utiliza una base de datos transaccional de los clientes de Falabella que comprende información desde el año 2000 hasta el 2010 en donde se tienen los ingresos de 5 nuevas tiendas: Plaza Norte, Plaza puente, La Dehesa, Manquehue y Estación Central.

Específicamente, se obtienen 5 bases de datos distintas, una por cada ingreso de una nueva tienda, en donde se seleccionan los clientes activos al momento de cada apertura, con su respectivo historial de compras del periodo señalado anteriormente. Cliente activo se denomina a aquellos consumidores que hayan realizado al menos una compra en cualquier tienda antigua de la misma cadena, un año antes de la apertura de la tienda nueva.

Dentro de la información transaccional de cada cliente, se tiene un identificador del individuo (se utilizaron número correlativos para guardar la identidad de la persona), con su respectiva ubicación cartesiana asociada a su dirección de facturación y comuna. A su vez, se tiene información acerca de la compra en sí con la fecha (día, mes y año), monto y tienda en que se realiza la transacción. Además, se tienen datos relacionados a la tienda como su ubicación cartesiana, el formato (Mall o calle) y fecha de apertura.

Considerando la selección de cliente de forma separada (un archivo por cada tienda nueva), se tiene el sesgo de obtener clientes que habían comprado al menos una vez durante un año antes de alguna apertura, lo cual no incluye clientes nuevos que se hubieran adicionado después de la primera apertura y que luego fuesen clientes regulares de Falabella. Es necesario eliminar este sesgo para que el estudio sea representativo de la realidad, entonces para descartar este efecto se unen las 5 bases de datos obteniendo una base de datos global que cuenta con el único sesgo de tener clientes activos, lo cual favorece el desarrollo de la investigación.

El archivo unificado que se obtiene luego de eliminar los sesgos indeseados contiene un total de 276.514 clientes que representan 15.980.500 transacciones.

Además, tomando en cuenta el total de las compras, se tiene que un 89,6% fueron realizadas en tiendas abiertas con anterioridad al año 2000 (tiendas antiguas).

Considerando que la base de datos es bastante extensa, más de 15 millones de filas, se utiliza un programa especializado en manejo de bases de datos llamado Access, el cual permite realizar consultas SQL para obtener la información que se solicita.

Al desglosar el total de las transacciones por tienda, se tiene una inclinación importante hacia las tiendas con mayor antigüedad, lo cual es lógico ya que tienen mayor tiempo en funcionamiento en el mercado. La distribución de compras es 89.6% para tiendas antiguas, 4.0% para Plaza Norte, 3.1% para Plaza Puente, 1.6% para La Dehesa, 0.9% para Manquehue y 0.8% para Estación Central. Ver Tabla N° 2.

Tabla 2: Distribución de transacciones base de datos original.

Tiendas	Transacciones	Porcentaje
Antiguas	14.314.985	89,60%
Plaza Norte	639.791	4,00%
Plaza Puente	502.743	3,10%
La Dehesa	253.252	1,60%
Manquehue	146.846	0,90%
Estación Central	122.883	0,80%
Total	15.980.500	100%

A continuación se presentan los filtros realizados y la respectiva reducción de clientes que conlleva:

1. El alcance geográfico fue dentro de la Región Metropolitana.
2. No se consideran las tiendas de Melipilla y San Bernardo porque están fuera del área de influencia de las nuevas tiendas.
3. No se consideran los departamentos Administración, servicios y CMR puntos porque no son compras normales.

4. No se consideran montos de compra inferiores a \$100 por ser saldos de devoluciones o cambios.
5. Sólo se tomaron en cuenta transacciones que estén totalmente identificadas (cliente, fecha, monto, entre otras).

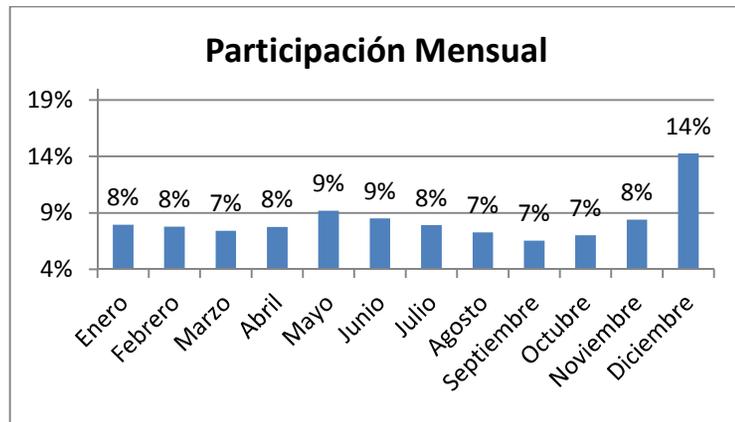
Las consideraciones anteriores reducen los clientes en un 16% y las transacciones en un 14%, obteniendo un total de 233.029 clientes y 13.797.122 transacciones hasta el momento.

6. Se utiliza una temporalidad quincenal, dividiendo cada mes en 2 periodos (periodo 1 desde 1° al 15vo día y periodo 2 desde el 16avo al 31avo día).
7. Se consideran las compras entre los periodos Noviembre 2002 (un año antes de la apertura de Plaza Norte) y Junio 2009 (un año después de la apertura de Estación Central). Se toma en cuenta un año antes de la primera apertura y un año después de la última apertura para no incluir en el estudio efectos que no estén relacionados con los ingresos de nuevas tiendas. Se redujeron los clientes en 0.9% (2.025) y las transacciones en un 32% (4.371.271).
8. Los montos de aquellas transacciones realizadas por un cliente dentro de la misma tienda y en la misma quincena, se suman. Se mantiene el número de clientes pero se reducen las transacciones en un 37% (3.486.208).
9. Los clientes que compran dentro de una misma quincena, mismo mes pero en distintas tiendas, son eliminados de la base de datos. Se reducen los clientes en un 49% (118.369) y las transacciones en un 80% (4.744.579).
10. Se realiza un histograma de transacciones mensuales para visualizar la distribución de las compras dentro del año. Como resultado se obtiene que el 14% de las compras se realizan en el mes de Diciembre, lo cual es lógico debido al impacto en las ventas que tienen las festividades navideñas. En consecuencia y con el fin de no tener elementos distorsionadores del comportamiento habitual de los consumidores, se decide eliminar las transacciones realizadas durante el mes de Diciembre. Este filtro resulta con una disminución del 2% de los clientes (1.727) y el 14% de transacciones (170.494). Ver Ilustración N° 3.

11. El último filtro que se realiza y que selecciona a los clientes finales de la base de datos global, fue el criterio de que al menos deben realizar una compra un año antes del ingreso de Plaza Norte y tener al menos una compra un año después de la apertura de Estación Central, ambos criterios considerando cualquier tienda de Falabella. Con este filtro se asegura que los clientes elegidos son regulares para Falabella. Se reducen los clientes en un 78% (86.275) y las transacciones en un 57% (588.620).

Se debe considerar que teniendo en cuenta la base de datos final obtenida después de los filtros realizados, se tiene una sub-estimación de la participación de mercado de cada una de las tienda de Falabella porque no se permite la inclusión de nuevos clientes que hayan ingresado dentro del horizonte de tiempo, es decir clientes que no compraban antes del año 2000 y que pudieran haberse agregado con posterioridad a esa fecha.

Ilustración 3: Grafico Histograma Transacciones Mensuales Base Datos Final.



La base de datos final que se obtiene después de realizar los filtros mencionados queda conformada por 24.867 clientes y 435.950 transacciones, de las cuales corresponden en un 92,5% a tiendas antiguas, 3,9% a Plaza Norte, 2,1% a plaza Puente, 0,6% a La Dehesa y Estación Central, y 0,4% a Manquehue. Ver Tabla N° 3.

Tabla 3: Distribución de transacciones base de datos final.

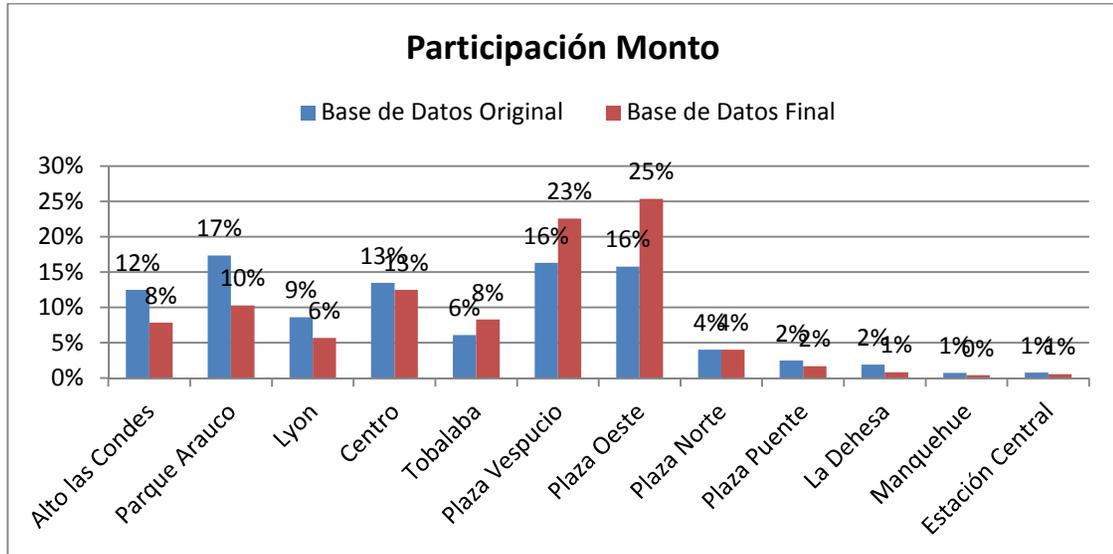
Tiendas	Transacciones	Porcentaje
Antiguas	403.237	92,50%
Plaza Norte	16.966	3,90%
Plaza Puente	8.953	2,10%
La Dehesa	2.515	0,60%
Manquehue	1.863	0,40%
Estación Central	2.416	0,60%
Total	435.950	100%

Al comparar la base de datos original contra la base de datos obtenida luego de la realización de los filtros, se aprecia que se mantiene la participación transaccional para cada una de las tiendas y la tendencia hacia las tiendas con mayor antigüedad. Esta mantención de las características mencionadas es importante ya que manifiesta que la disminución en el número de clientes y transacciones no afecta la representatividad de la muestra.

Ahora bien, considerando la distribución de género dentro de las bases de datos, existe una leve predominación del género femenino, en donde ésta diferencia entre hombre y mujeres no supera el 4% en ambas bases de datos.

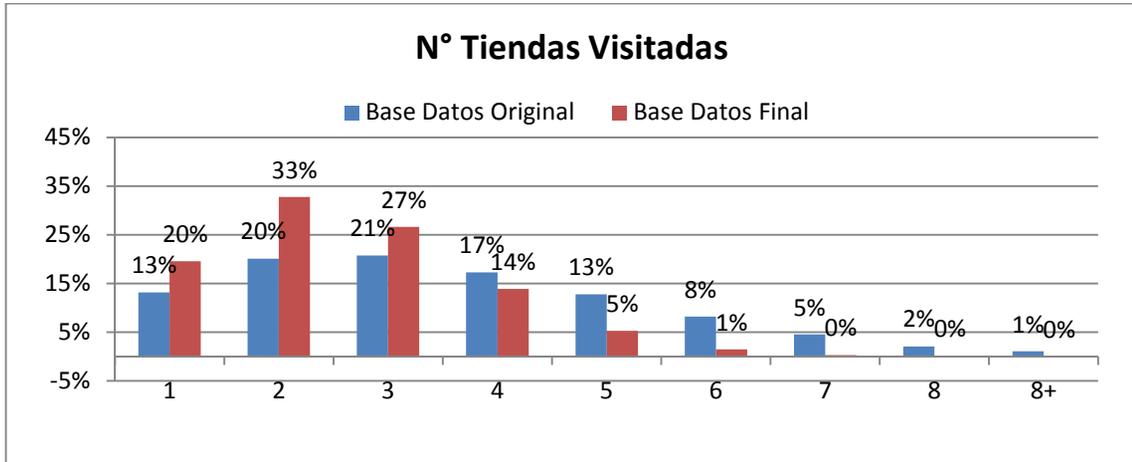
Con relación a la distribución de los montos para las distintas tiendas, al igual que con el número de transacciones, existe una importante concentración de los montos en las tiendas abiertas con posterioridad al año 2000, sobresaliendo las tiendas de Plaza Oeste con un 16% y 25% y Plaza Vespucio con un 16% y 23% respectivamente a las bases de datos original y post filtraciones. Ver Ilustración N°4 para mayor detalle.

Ilustración 4: Grafico Distribución de montos por tienda Falabella.



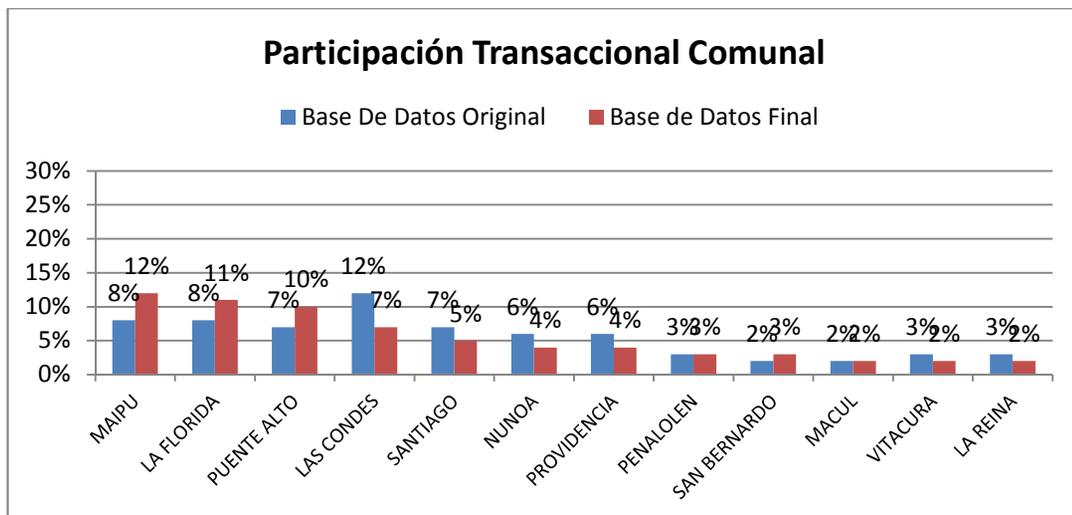
Por otro lado, analizando el número de tiendas que visita un cliente dentro del periodo de estudio, se aprecia que tanto para las bases de datos original como para la final, la mayoría de los clientes tiene un conjunto de tienda entre 2 (20% y 33%) y 3 tiendas (21% y 27%). Además, para ambas bases de datos, considerando todo el conjunto de tiendas, se muestra una distribución similar de las tiendas visitadas. Ver Ilustración N° 5.

Ilustración 5: Grafico número de tiendas por cada consumidor.



Ahora bien revisando la distribución comunal de compras que se observa, tanto antes como después de la aplicación de los filtros se mantiene una distribución homogénea entre las distintas comunas, teniendo mayor participación las comunas de Maipú (8% y 12%), La Florida (8% y 11%) y Las Condes (12% y 7%). Ver Ilustración N° 6.

Ilustración 6: Grafico distribución de compras por comuna (principales 12).



Como conclusión a los distintos comentarios y análisis señalados en este capítulo y teniendo en mente el objetivo de limpiar la base de datos original para obtener sólo la información idónea para realizar el estudio de buena forma, se considera que en términos generales la disminución en el número de clientes y cantidad de transacciones producto de los filtros realizados, no afecta la

representatividad de la población obtenida en la base de datos final, ya que se mantienen las principales características poblacionales que pueden influir en las decisiones de incidencia y monto de compra.

7.1) Muestra

La base de datos final que se desarrolla en el capítulo anterior está conformada por 24.867 clientes y 435.950 transacciones, la cual contiene demasiados datos para los modelos que se quieren trabajar, en especial para el modelo Mixed Logit con teoría bayesiana que se quiere realizar para la incidencia de compra porque los tiempos computacionales son muy elevados y por tanto impracticables para una toma de decisiones real.

Debido a lo anterior y basándose en los estudios previos realizados por Asim Ansari, Carl F. Mela y Scott A. Neslin (2008)⁹ y Gupta (2008)¹⁰, se elige de la base de datos global una muestra aleatoria de 500 clientes, de los cuales se obtiene un total de 8.797 transacciones. Es importante señalar que la muestra utilizada para los modelos tiene una temporalidad quincenal, en donde cada quincena del mes representa una oportunidad de compra para el consumidor, es decir cada cliente dentro de un año tiene 24 oportunidades de compra para cada alternativa del conjunto de elección disponible en ese momento.

Ahora bien, un punto fundamental para la muestra es que ésta sea representativa de la realidad que tienen los clientes de Falabella en cuanto a su comportamiento compra. Para ello, se realiza un análisis descriptivo de la muestra para luego compararla con la base de datos global y poder concluir su representatividad.

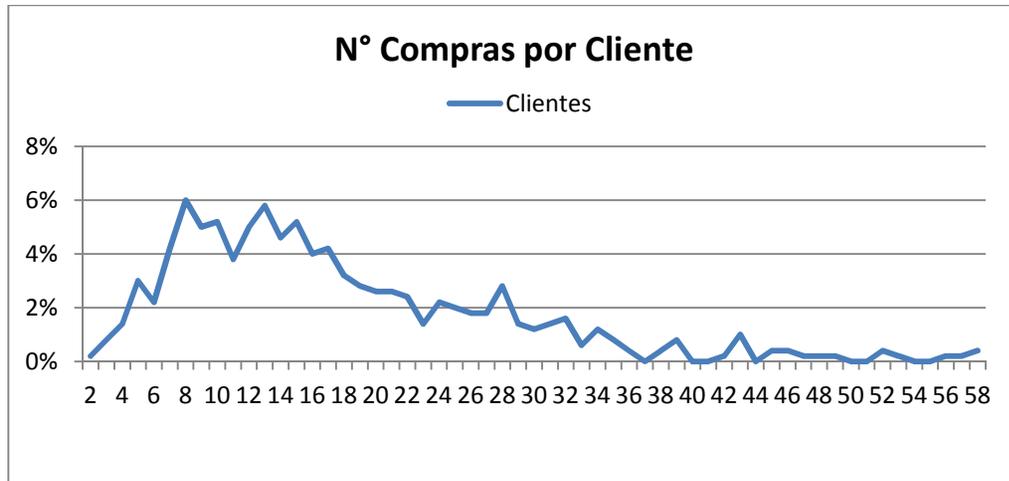
Como se menciona al inicio, la muestra cuenta con 500 clientes y 8.797 transacciones en total, en donde la cantidad de compras promedio de los individuos es 18, con un mínimo de 2 y un máximo de 58. Además, el 50% de los consumidores realiza entre 2 a 15 compras en todo el periodo contemplado para el trabajo, tanto que el 80% realiza entre 2 a 25 compras.

⁹ "Customer Channel Migration", Journal of Marketing Research, Vol. XLV (February 2008), 60-76.

¹⁰ "Impact of sales Promotions on When, What, and Much to Buy", Journal of Marketing Research 25 (4), 342-355.

Al analizar la frecuencia de compra dentro de la muestra considerando el número de compras promedio por individuo, se tiene que los clientes de Falabella realizan una compra cada 8 meses entregando un número anual de 1,5 compras, lo cual es coherente con la información utilizada por Gupta (2008) para realizar su estudio. Ver Ilustración N° 7.

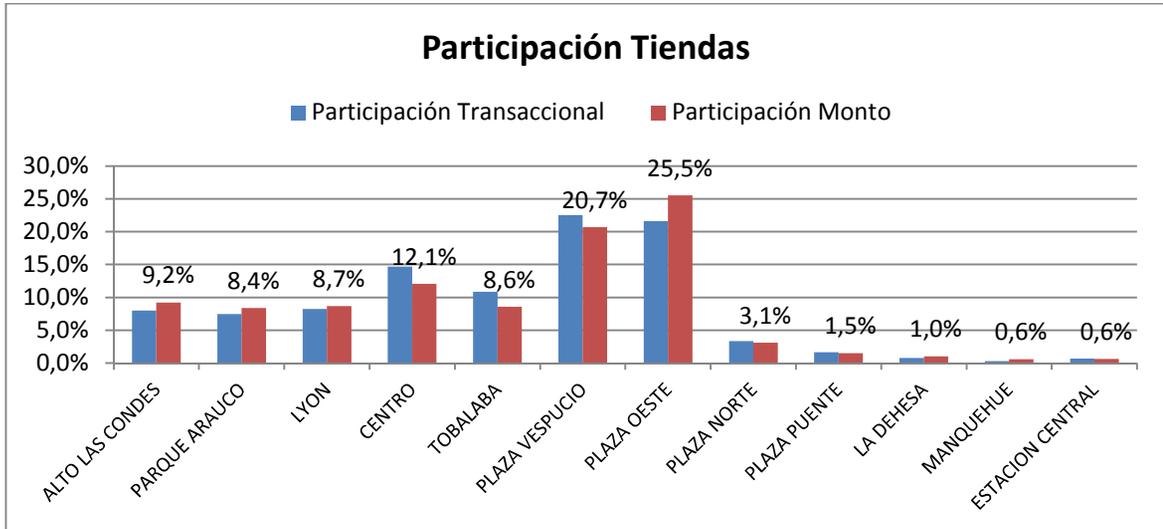
Ilustración 7: Gráfico número compras realizadas por cliente.



Enfocándose en la distribución de los montos entre las distintas tiendas, se aprecia una tendencia importante hacia las tiendas abiertas antes del año 2000, concentrando más del 90% de los ingresos especialmente en Plaza Vespucio y Plaza Oeste, con un 20,7% y 25,5%. Situación similar ocurre al revisar la distribución en la cantidad de transacciones por tienda, en donde se tiene el mismo porcentaje entre tiendas lanzadas antes y después del año 2000. .Ver Ilustración N° 8.

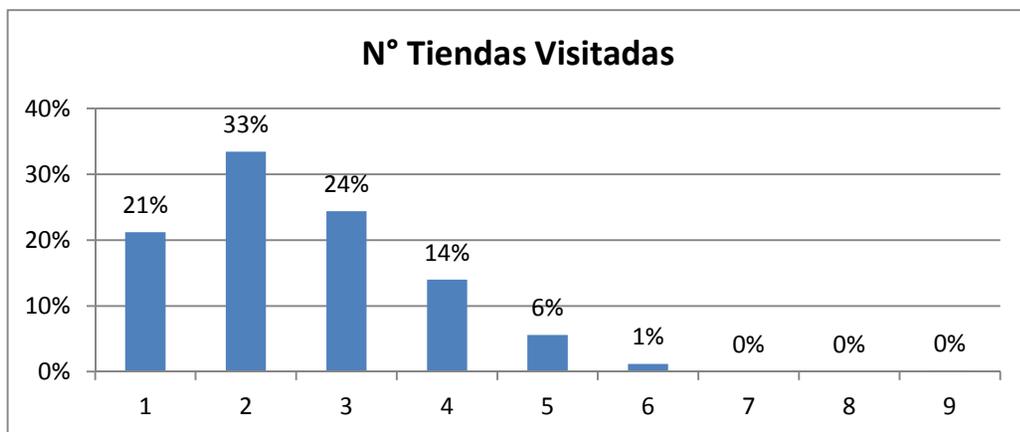
Por otro lado, la participación que tienen hombres y mujeres dentro de la muestra es totalmente equilibrado, teniendo para cada género el mismo número de transacciones.

Ilustración 8: Gráfico participación tiendas.



Continuando con el análisis descriptivo de la muestra, otro factor importante es el conjunto de elección que tienen los consumidores de la muestra, es decir el número de tiendas en las cuales realiza sus compras. Se tiene que el 78% de los clientes visita máximo 3 tiendas en todo su historial, en donde el 33% visita sólo 2 tiendas. Ver Ilustración N° 9.

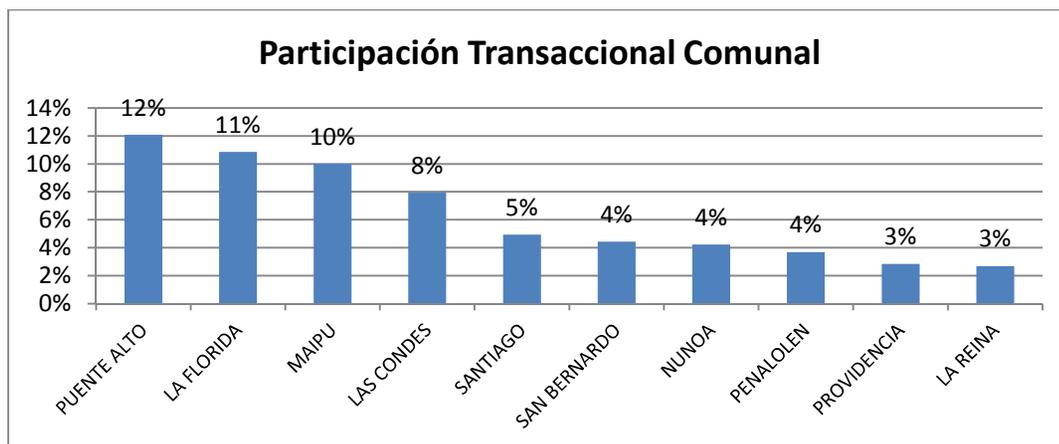
Ilustración 9: Gráfico número de tiendas por cada consumidor.



El alcance que tiene este trabajo es dentro de la región metropolitana, por tanto se debe revisar la distribución de compras a través de las distintas comunas con el fin de ver se está teniendo una muestra heterogénea en este sentido. Se aprecia que la distribución es bastante pareja, lo cual entrega indicios de heterogeneidad de la población. Las comunas con mayor participación dentro de la muestra son

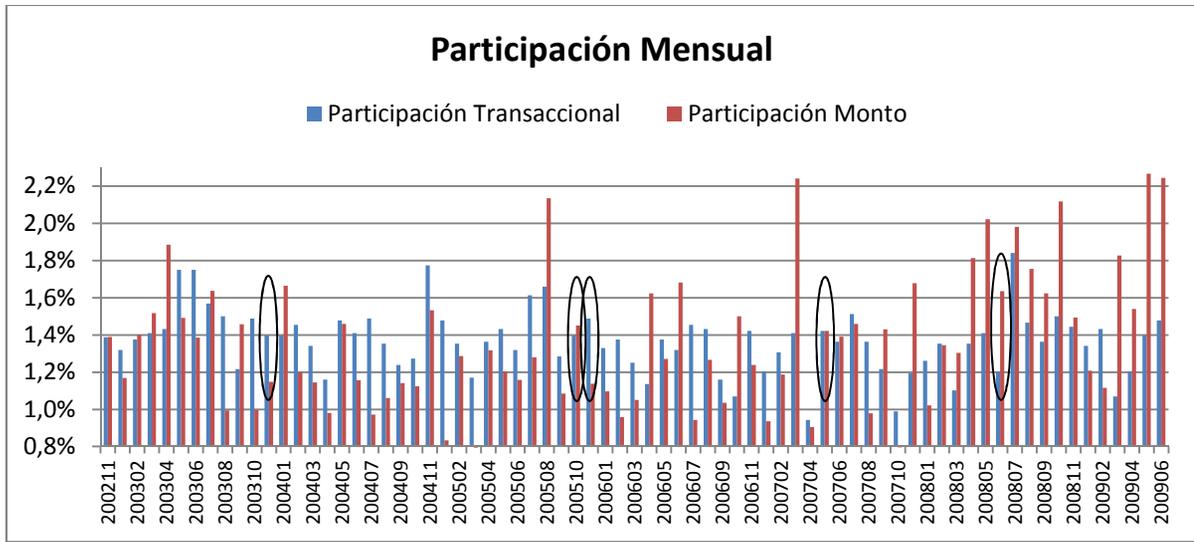
Puente Alto (12%), La Florida (11%), Maipú (10%) y Las Condes (8%). Ver Ilustración N° 10.

Ilustración 10: Gráfico distribución de compras por comuna (principales 10).



Para finalizar el análisis descriptivo de la muestra, es fundamental revisar la distribución de la cantidad de transacciones y respectivos montos durante el horizonte temporal de la investigación, considerando la influencia que tienen los ingresos de las nuevas tiendas sobre los índices señalados. Basándose en el Ilustración N° 11, en donde se señalan con una circunferencia la fecha en que ingresa cada tienda, se tiene que la mayor cantidad de compras y montos se realizan antes y después de los ingresos de las tiendas, lo cual es favorable para el estudio ya que se necesitan clientes activos durante todo el horizonte de tiempo.

Ilustración 11: Gráfico distribución de compras y montos desde Noviembre 2002 hasta Junio 2009.



La conclusión final luego de consolidar todos los análisis realizados sobre la muestra aleatoria seleccionada, es que los 500 clientes seleccionados de la base de datos global es totalmente representativa del comportamiento y características generales que se encuentran en los datos que se tenían originalmente. En resumen, se tiene una participación de monto similar, en donde predominan las tiendas de Plaza Oeste y Plaza Vespucio, también el número de tiendas visitadas es igual predominando el conjunto de elección de 2 a 3 tiendas, y los sectores geográficos con mayor influencia son Maipú, La Florida y Las Condes.

8) Modelo Incidencia de Compra

El objetivo principal de este trabajo es desarrollar un modelo sobre incidencia de compra para analizar la variación en el comportamiento de compra de los clientes actuales de Flabella al ingresar una nueva tienda de la misma cadena. En primer lugar se quiere describir y caracterizar el comportamiento del cliente asociado a la decisión de la tienda donde realiza su compra teniendo un conjunto acotado de alternativas, en donde posteriormente se utiliza esta información para estimar la participación de mercado de cada una de las tiendas de Falabella y luego se cuantifica la canibalización producida al ingresar nuevas tiendas de la misma cadena. El desarrollo del modelo señalado entrega un valor adicional si se compara a la mera utilización de la información agregada de cada tienda porque permite tener una metodología establecida para tomar una decisión de localización de tienda y un conocimiento de los factores influyentes en la decisión de compra de los clientes.

Considerando la variable endógena, ésta toma un solo valor dentro de un conjunto de elección establecido, mutuamente excluyente, finito y con más de dos alternativas porque representa la decisión de dónde compra el cliente, para lo cual se utiliza un modelo Logit Mixed para abarcar este efecto. Adicionalmente a lo anterior, la decisión de utilizar este modelo y no un modelo Logit Multinomial tradicional es que cuenta con los siguientes beneficios con respecto a este último:

1. Permite heterogeneidad en los consumidores.
2. Acepta cambios aleatorios en los gustos de los consumidores, lo cual es fundamental porque en este trabajo el conjunto de elección de cada cliente va variando con el tiempo.
3. Entrega distintos coeficientes para cada consumidor y no una igual para toda la muestra, permitiendo obtener conclusiones individuales o por grupos de personas.
4. Al agregar una nueva tienda, no sustrae participación de forma proporcional a las antiguas tiendas, entregando la proporción real de cada alternativa al introducir una nueva opción al conjunto de elección.

Al tomar en cuenta las características del efecto que se quiere estudiar, queda de manifiesto que los 4 beneficios del Mixed Logit que se plantean resultan

fundamentales para obtener resultados que nos permitan concluir de buena forma los objetivos e hipótesis requeridos.

Complementariamente y para entregar un valor agregado al trabajo de investigación, el modelo Mixed Logit es desarrollado a través de un enfoque bayesiano que afronta la incertidumbre con una distribución de probabilidad sobre los parámetros y asume constantes los datos muestrales.

8.1) Metodología

En el trabajo desarrollado, se elabora una metodología que permita cuantificar la canibalización al ingresar una nueva tienda, en donde una primera etapa es identificar los factores que determinan la decisión de compra de un cliente al tener un conjunto de elección finito y que varía con el tiempo, utilizando información transaccional de la persona e información de la tienda que sea percibida por el individuo en la experiencia de compra.

Una segunda etapa y ya con una función de utilidad elaborada para cada cliente, se procede a obtener la participación de mercado de las distintas tiendas de Falabella considerando que el cliente asiste a aquella que le produzca una mayor utilidad dentro del conjunto de elección disponible en ese momento. Es importante señalar que para cada apertura, sólo se tienen dentro del conjunto de elección las tiendas existentes para esa fecha, incluyendo la opción de “No Comprar”.

La tercera etapa es cuantificar el efecto de canibalización, asociada a la pérdida o ganancia de participación de mercado (participación transaccional) de las tiendas, al comparar los periodos “Antes” y “Después” de las aperturas de las nuevas tiendas (Koen y Neslin, 2006; Kumar, Pancras y Sriram, 2009 y 2012). Esta valorización es una cota superior de la canibalización porque puede haber otros factores externos adicionales que hayan producido variaciones en las participaciones de mercado. Además, se realizan conclusiones acerca de los objetivos e hipótesis planteadas para el trabajo.

Ahora bien, la metodología previa que se realiza para desarrollar el modelo de incidencia de compra se divide en 4 etapas principales: Transformación de la información (base de datos) a un formato específico para la utilización del Mixed Logit, formulación de la función de utilidad del modelo, ejecución del Mixed Logit y análisis de resultados para estimar el valor de la canibalización.

La transformación de la información a un formato específico se refiere a modificar la base de datos muestral para colocar los datos en un formato que sea apto para que el Mixed Logit pueda ejecutarse. A grandes rasgos, este nuevo formato señalada que para cada unidad de tiempo (una quincena), el cliente debe tener todo el conjunto de elección disponible en ese momento, incluyendo la opción de “No Comprar”, en donde para el modelo sólo se activa la opción que ocurre realmente, es decir si el cliente compra en la tienda Plaza Norte, sólo se activa la fila correspondiente a esa tienda y las restantes no. Este punto se revisa en detalle en el capítulo 9.2.

La formulación de la función de utilidad hace referencia a la explicación de las variables explicativas utilizadas, la temporalidad de cada una y la definición de la ecuación matemática que la representará.

Para la ejecución del modelo de incidencia de compra se debe primeramente desarrollar el código del modelo Mixed Logit (ver Anexo N° 1) para luego utilizar un software matemático llamado Matlab, en donde se introduce la modelación (código) y la base de datos modificada con el horizonte temporal de un año antes de la apertura y un año después de la misma para obtener los resultados. Como se menciona en el capítulo teórico de este trabajo, al utilizar un enfoque bayesiano se dificulta la ejecución práctica del modelo debido al cálculo de integrales múltiples, es por ello que se hace imprescindible la utilización del método de simulación Markov Chain Monte Carlo (Boatwright, Borle, Kadane, Nunes y Shmueli, 2005), en donde se realiza un procedimiento de muestreo para obtener los parámetros, es decir no se maximiza ninguna función.

8.2) Formato Base de Datos Utilizados

La base de datos original cuenta con información transaccional de los consumidores y con sus respectivos datos individuales, en donde cada una de las filas representaba una transacción realizada por el consumidor. Este formato no es correcto para el modelo que se quiere realizar porque en este caso, el conjunto de elección y la temporalidad de cada cliente se resume sólo a la tienda y fecha donde se realiza la compra.

El modelo Mixed Logit, en especial el algoritmo MCMC, necesita que la base de datos muestral contenga, para cada transacción, todo el espectro temporal y todo el conjunto de elección, en donde otra columna indica la fecha y tienda donde se

realiza la compra. Considerando lo anterior, por cada transacción se tienen 146 periodos temporales (desde Noviembre del 2002 hasta Junio del 2009) y el número de alternativas disponibles, incluyendo la opción de No Comprar.

Por otro lado, para crear los interceptos que capturan la atracción que tiene cada individuo hacia las tiendas antiguas, se agrega una matriz identidad que activa estos parámetros (se van agregando a medida que la tienda nueva exista) Ahora bien, para incluir los interceptos de las tiendas nuevas, se crea una matriz identidad por cada tienda, en donde éstas son cero si para un cierto periodo temporal la tienda aún no está inaugurada.

Las siguientes columnas son las variables explicativas que se quieren utilizar: Distancia, distancia al cuadrado, mall, densidad población, distancia entre tiendas y antigüedad. Cabe destacar que cada valor de estas variables hace referencia a la tienda respectiva a esa fila. Es importante mencionar que todas las filas que hacen referencia a la alternativa "No Comprar" son completamente rellenas con ceros, ya que con esto se busca normalizar el modelo.

En el Anexo N° 2 se muestra un fragmento de la base de datos para entregar un ejemplo del formato utilizado y detallado anteriormente. En este extracto se considera el ingreso de Plaza Norte (PN), para un cliente y en la fecha Noviembre 2002.

Luego de tener definido el formato que se necesita para el modelo Mixed Logit, para poder armar la base de datos a utilizar en el programa Matlab se crean variadas Macros en el software Excel. Específicamente, se tiene una macro general llamada "Armar Base de Datos" (ver Anexo N° 3), la cual a su vez está compuesta por 14 macros adicionales:

- Contar: Cuenta la cantidad de compras que tiene cada cliente (Anexo N° 4).
- Cambio clientes: Se cambia el número identificador de cada cliente por números correlativos desde el 1 hasta el 500 (Anexo N° 5).

- Meses Formato Antigüedad: Agrega los periodos temporales para cada transacción y completa las variables explicativas de Mall y Antigüedad (Anexo N° 6).

- Densidad Población: Calcúla la variable explicativa relacionada a la participación poblacional que existe en las comunas en donde se ubican las distintas alternativas (Anexo N° 7).

- Distancia: Calcúla la distancia euclidiana entre cada cliente y las distintas tiendas. Además, calcula la distancia al cuadrado (Anexo N° 8).

- Distancia Nueva Tienda: Calcúla la distancia euclidiana entre las tiendas ya inauguradas y la nueva tienda entrante (Anexo N° 9).

- Elección: Activa la tienda en la cual se realiza la compra y si no hubiera transacción en el periodo, se activa la alternativa “No Comprar”. Además, coloca el identificador y monto de compra en la opción que se activa (Anexo N° 10).

- Código: Coloca los códigos representativos de cada alternativa (Anexo N° 11).

- Identidad: Crea la matriz identidad que activa los interceptos de las tiendas antiguas (Anexo N° 12).

- Identidad Entrada PN, PP, DE, MA y EC: Crea la matriz identidad que activa los interceptos de Plaza Norte (Anexo N° 13), Plaza Puente (Anexo N° 14), La Dehesa (Anexo N° 15), Manquehue (Anexo N° 16) y Estación Central (Anexo N° 17) respectivamente.

Además, para calcular las distancias euclidianas relacionadas a las tiendas de Falabella, se utilizan las siguientes coordenadas cartesianas:

1. Alto las Condes: Antigua, mall y coordenadas (-70,5456; -33,3891).

2. Parque Arauco: Antigua, mall y coordenadas (-70,5788; -33,4014).

3. Lyon: Antigüedad, calle y coordenadas (-70,6104; -33,4213).
4. Centro: Antigüedad, calle y coordenadas (-70,6503; -33,4432).
5. Tobalaba: Antigüedad, mall y coordenadas (-70,5544; -33,5699).
6. Plaza Vespuccio: Antigüedad, mall y coordenadas (-70,6003; -33,5187).
7. Plaza Oeste: Antigüedad, mall y coordenadas (-70,7177; -33,5147).
8. Plaza Norte: Nueva, mall y coordenadas (-70,6745; -33,3687).
9. Plaza Puente: Nueva, mall y coordenadas (-70,6508; -33,4353).
10. La Dehesa: Nueva, mall y coordenadas (-70,5157; -33,3585).
11. Manquehue: Nueva, calle y coordenadas (-70,5663; -33,4111).
12. Estación Central: Nueva, mall y coordenadas (-70,6825; -33,4519).

Por último, una de las variables descriptivas considera la variación en la distribución poblacional entre las distintas comunas de la región metropolitana en donde están ubicadas las tiendas, con un total de habitantes de 6.318.299 en el año 2002, 6.391.827 en el 2003, 6.465.348 en el 2004, 6.538.896 en el 2005, 6.607.805 en el 2006, 6.676.745 en el 2007, 6.745.651 en el 2008 y 6.814.630 en el año 2009. Ver Tabla N° 4.

Tabla 4: Participación poblacional desde 2002 hasta 2009.

Tienda	Comuna	2002	2003	2004	2005	2006	2007	2008	2009
Alto Las Condes	Las Condes	4,10%	4,11%	4,13%	4,14%	4,15%	4,15%	4,15%	4,16%
Parque Arauco	Las Condes	4,10%	4,11%	4,13%	4,14%	4,15%	4,15%	4,15%	4,16%
Lyon	Providencia	1,93%	1,92%	1,91%	1,91%	1,90%	1,88%	1,87%	1,86%
Centro	Santiago	3,09%	3,01%	2,93%	2,85%	2,76%	2,68%	2,60%	2,52%
Tobalaba	Puente Alto	8,64%	8,85%	9,06%	9,26%	9,49%	9,72%	9,94%	10,15%
Plaza Vespucio	Santiago	3,09%	3,01%	2,93%	2,85%	2,76%	2,68%	2,60%	2,52%
Plaza Oeste	Maipú	8,52%	8,89%	9,24%	9,59%	10,04%	10,47%	10,89%	11,30%
Plaza Norte	Santiago	3,09%	3,01%	2,93%	2,85%	2,76%	2,68%	2,60%	2,52%
Plaza Puente	Puente Alto	8,64%	8,85%	9,06%	9,26%	9,49%	9,72%	9,94%	10,15%
La Dehesa	Lo Barnechea	1,30%	1,33%	1,36%	1,39%	1,42%	1,46%	1,49%	1,52%
Manquehue	Las Condes	4,10%	4,11%	4,13%	4,14%	4,15%	4,15%	4,15%	4,16%
Estación Central	Estación Central	2,03%	1,98%	1,93%	1,88%	1,84%	1,79%	1,74%	1,70%

8.3) Formulación Modelo Logit Mixed

Antes de la ejecución del modelo Logit Mixed para la incidencia de compra, se deben establecer los conjuntos de datos, la función de utilidad y las respectivas variables explicativas que describen el fenómeno.

- Consumidores $i = \{1, \dots, 500\}$.
- Tiendas $j = \{\text{Alto las Condes, Parque Arauco, Lyon, Centro, Tobalaba, Plaza Vespucio, Plaza Oeste, Plaza Norte, Plaza Puente, La Dehesa, Manquehue, Estación Central y No Comprar}\}$. Ver Tabla N° 5.
- Periodos Temporales $t = \{2002111 \text{ (Noviembre 2002, 1° Quincena)}, \dots, 2009062 \text{ (Junio 2009, 2° Quincena)}\}$.

Es importante señalar que para cada apertura de una nueva tienda de Falabella, sólo se consideran las tiendas de la cadena abiertas hasta esa fecha, es decir las tiendas van ingresando al conjunto de elección de los clientes a medida que van siendo abiertas, entonces una tienda no existente no está dentro del conjunto de elección de los clientes.

La función de utilidad para un consumidor “i”, que visita la tienda “j” y en el periodo “t” es:

$$U_{ijt} = \sum_j \beta_{ij} + \beta_i^1 \text{Distancia}_{ij} + \beta_i^2 \text{Distancia}_{ij}^2 + \beta_i^3 \text{Mall}_j + \beta_i^4 \text{Participación Población}_{jt} + \beta_i^5 \text{Distancia Tienda Nueva}_j + \beta_i^6 \text{Antigüedad Tienda}_j + \varepsilon_{ijt}. \quad (34)$$

Al revisar la función de utilidad y los índices incluidos en ella, se aprecia que no hay problemas en los grados de libertad al tener la información suficiente para realizar los cálculos de los parámetros.

Tabla 5: Codificación alternativas (tiendas Falabella).

Código	Alternativas
100021	Alto las Condes
100022	Parque Arauco
100023	Lyon
100025	Centro
100045	Tobalaba
100026	Plaza Vespucio
100027	Plaza Oeste
100064	Plaza Norte
100067	Plaza Puente
100065	La Dehesa
100071	Manquehue
100073	Estación Central
100000	No Comprar

Para definir la función de utilidad anterior, se revisa la información disponible y la literatura confeccionada hasta la fecha, en donde se incluyen variables explicativas relacionadas a características de los clientes de Falabella y también características propias de las tiendas pero que sean percibidas por el cliente en la experiencia de compra (visita a la tienda).

- β_{ij} = Representa el “atractivo” que tiene la tienda “j” para el cliente “i” con respecto a la alternativa No Comprar.

Esta variable está relacionada al “agrado” o “atractivo” que tiene el cliente para asistir a una respectiva tienda, tomando como base la alternativa de no asistir si no la agrada la tienda.

- $Distancia_{ij}$ = Distancia euclidiana entre el cliente “i” y la tienda “j”.
- $Distancia_{ij}^2$ = Distancia euclidiana entre el consumidor “i” y la tienda “j” al cuadrado.

Para las dos últimas variables, se utiliza la distancia euclidiana a través de las coordenadas cartesianas de los clientes y las respectivas tiendas. Esta información fue entregada por la misma cadena Falabella.

- $Mall_j$ = Toma valor 1 si la tienda “j” está ubicada dentro de un mall y 0 si es formato calle.
- $Participación Población_{jt}$ = Participación poblacional de la comuna en que se encuentra ubicada la tienda “j” en el periodo “t”.

La variable relacionada a la participación poblacional se agrega para verificar si la densidad poblacional de las comunas está correlacionada con la demanda.

- $Distancia Tienda Nueva_j$ = Distancia entre la tienda “j” y las restantes tiendas del conjunto de elección (misma cadena).

La inclusión de la variable anterior busca ratificar si la nueva estrategia de localización de tienda (agregada) es beneficiosa o no para Falabella al entregarle mayor utilidad a los clientes, es decir verificar si el cliente considera mejor tener una tienda nueva de Falabella cerca de las restantes tiendas de la misma cadena ya establecidas.

- *Antigüedad Tienda_j*= Toma el valor 1 si la tienda “j” fue abierta antes del año 2000 y 0 si fue inaugurada posterior a esa fecha.

Por otro lado, cada intercepto β_{ijt} está compuesto de dos partes:

$$\beta_{ij} = \beta_{ij}^0 + \beta_{ij}^{TN} X_{TN} \quad (35)$$

$$X_{TN} = \begin{cases} 1, & \text{si TN (Tienda Nueva) dentro del conjunto de elección.} \\ 0, & \text{si no existe TN dentro del conjunto de elección.} \end{cases}$$

La razón de separar el intercepto β_{ij} en dos partes es identificar la influencia que tiene la nueva tienda sobre el “atractivo” del cliente hacia una tienda antigua (j), es decir con el primer término se obtiene el atractivo original que tiene el cliente sobre la tienda y con el segundo término se incorpora al intercepto general el atractivo que le produce al individuo la nueva tienda, al activarse el intercepto β_{ij}^{TN} que captura este fenómeno.

8.4) Resultados y Análisis

Al tener los detalles del trabajo de investigación ya abordados, se procede a la ejecución del modelo Mixed Logit para estudiar el efecto de canibalización a través de un modelo de incidencia de compra, para ello se utilizó el programa Matlab en donde en términos generales, la simulación MCMC se realiza con 200.000 iteraciones.

Para detallar de forma correcta los resultados, se analiza este capítulo en 4 puntos principales: Influencia de las variables descriptivas (parámetros) sobre la incidencia de compra, preferencia de los consumidores con respecto a las distintas tiendas de Falabella, la canibalización producida con los ingresos de nuevas tiendas y los valores de la herramienta Hit Rate para ver el grado de ajuste del modelo.

8.4.1) Variables Descriptivas y Parámetros

Los parámetros que describen la función de utilidad son calculados para los 500 consumidores seleccionados, es decir cada cliente tiene un conjunto de parámetros que permite identificar la utilidad asociada a cada una de las tiendas individualmente, obteniendo 63 valores. Previamente a analizar los parámetros, es importante señalar que se utiliza el promedio de cada una de ellas para los análisis gruesos y los valores individuales para los análisis específicos. Además, el intervalo de confianza utilizado es un 95%.

La principal interpretación de los parámetros asociados a las variables descriptivas de todo modelo es la identificación de influencia (dirección) que tienen éstas sobre la utilidad percibida por el consumidor, es decir cuando el valor del parámetro es negativo significa que la variable es indirectamente proporcional a la utilidad e interpretación contraria cuando el parámetro tiene valor positivo.

Las variables descriptivas tienen los siguientes valores promedio: Distancia (-2.90), Distancia al Cuadrado (-0.42), Mall (-4.46), Participación Poblacional (-0.40), Distancia Tienda Nueva (-0.99) y Antigüedad (-1.55).

Tabla 6: Datos estadísticos variables explicativas Mixed Logit.

Variables Descriptivas	Parámetros Promedio	N° Iteraciones Modelo	Estadísticamente Distinto Cero	Convergencia	INTERVALO CONFIANZA		
					0.025	0.5	0.095
Distancia	-2,9	200.000	Si	No	-3,28	-2,9	-2,59
Distancia al Cuadrado	-0,42		Si	Si	-0,54	-	0,42
Mall	-4,46		Si	Si	-4,6	-	4,46
Participación Poblacional	-0,4		Si	Si	-0,53	-0,4	-0,27
Distancia Tienda Nueva	-1		Si	Si	-1,1	-	0,99
Antigüedad	-1,55		Si	Si	-1,66	-	1,55

Considerando esta forma de analizar los signos de los parámetros y a su vez al revisar los resultados obtenidos por el modelo en la Tabla N°6, las variables de Distancia, Distancia Cuadrado, Mall, Participación Poblacional, Distancia Nueva Tienda y Antigüedad adquieren la siguiente interpretación:

- Distancia (-2.90) y Distancia Cuadrado (-0.42): Considerando la distancia que tienen los consumidores con respecto a las tiendas disponibles, se tiene que a menor distancia entre ellos, aumenta la probabilidad de incidencia.

- Mall (-4.46): El formato de la tienda mall afecta negativamente a la incidencia de compra, lo cual es coherente si se piensa que una mayor competencia (tiendas de diferentes cadenas de retail) afecta negativamente a la demanda. Además, en tiendas que están geográficamente cerca aumenta la probabilidad que los individuos tengan comportamiento de compra y patrones de demanda similares, por ende las cadenas compiten por los mismos consumidores. Este resultado está validado por diversos estudios anteriores (Reilly 1931; Huff 1964; Huff y Batsell, 1977; Bronnenberg y Sismeiro, 2002; Deleersnyder, 2002; Allenby y Yang, 2003; Bronnenberg y Mela, 2004; Bell y Song, 2007; Kumar, Pancras y Sriram, 2009).

- Participación Poblacional (-0.40): Tomando en cuenta los datos poblacionales, el modelo señala que a mayor participación poblacional de la comuna en que se encuentra la tienda, menor probabilidad de incidencia. A su vez, esta variables tiene un valor muy cercano a cero dentro de la función y por tanto, no toma alta relevancia en la utilidad final y por tanto en la decisión de compra del cliente.

- Distancia Nueva Tienda (-1.00): Al revisar el efecto que produce la distancia entre las tiendas de la misma cadena, se tiene que a menor distancia entre la tienda nueva y la antigua, aumenta la incidencia de compra. Este resultado va en total concordancia con la nueva estrategia que quiere seguir la industria del retail, agrupando sus tiendas en radios cercanos de influencia para potenciar la incidencia de compra de sus clientes.

- Antigüedad (-1.55): Considerando la antigüedad de las tiendas de Falabella, se tiene que las tiendas ingresadas antes del año 2000, manifiestan una menor incidencia de compra. Lo anterior es explicable debido al atractivo que tienen nuevas tiendas a través de publicidad u otros elementos que cumplen el objetivo de posicionar a la tienda rápidamente.

Por otro lado, analizando los intervalos de confianza que tiene cada una de las variables explicativas, se aprecia que el cero no esta dentro del intervalo para ninguna de las variables, lo cual significa que son estadísticamente distintas de cero, propiedad relevante para el modelo (ver Anexo N° 19). Además, cada variables converge despues de las 200.000 iteraciones del algoritmo salvo la

variables “Distancia” que no logra estabilizarse dentro del periodo. Ver el detalle en el Anexo N° 18 (gráficos).

Ahora bien, al revisar los valores promedio de los parámetros y sus respectivos signos, en terminos generales los consumidores de Falabella tienen una mayor incidencia de compra en aquellas tiendas que tienen formato calle, ingresadas después del año 2000, y que se encuentran ubicadas a una menor distancia del consumidor y cercanas a otras tiendas de la misma cadena. Lo anterior tiene sentido al considerar que los consumidores de hoy en día le entregan un elevado valor a la comodidad del servicio y a la satisfacción que produzca la experiencia de compra en la tienda (Kumar, Sriram y Pancras, 2009 y 2012),

Como se señala en la sección teórica, el modelo Mixed Logit y el enfoque bayesiano permite obtener parámetros diferentes para cada uno de los individuos, es por ello que esta información permite la realización de análisis más detallados. Es por ello que se realiza una interpretación de las variables explicativas diferenciando por zonas geográficas de la región metropolitana y por género del consumidor, en donde los resultados se pueden apreciar en la Tabla N° 7 y Tabla N°8 respectivamente.

Tabla 7: Parámetros variables explicativas por zonas RM Mixed Logit.

	Norte	Poniente	Sur	Centro	Oriente	Sub Urbana
Distancia	-3,03	-2,82	-3,13	-2,66	-2,81	-2,95
Distancia al Cuadrado	-0,30	-0,45	-0,45	-0,4	-0,41	-0,39
Mall	-3,98	-4,42	-4,57	-4,68	-4,4	-4,33
Participación Poblacional	-0,19	-0,22	-0,26	-0,44	-0,76	-0,09
Distancia Tienda Nueva	-1,06	-1,05	-1	-0,97	-0,91	-1,07
Antigüedad	-2,1	-1,65	-1,58	-1,59	-1,27	-1,66
Total Clientes	36	84	159	68	131	22

Se tiene que para todas las zonas de la región metropolitana se mantienen los mismos signos e interpretaciones para cada parámetro pero existen variaciones en las magnitudes que tiene cada una sobre la utilidad del consumidor, llamando la atención la zona norte. Para el caso de la zona norte, se genera un aumento en la importancia de la distancia entre el consumidores y la tienda pero una disminución en la relevancia del formato, lo cual puede deberse a dificultades de acceso o tránsito de las personas y por tanto, una mayor preferencia a tiendas ubicadas en Mall, las cuales generalmente tiene mejor conectividad.

Tabla 8: Comunas por zona RM.

Zonas RM	Comunas
Norte	Independencia, Recoleta, Huechuraba, Colina, Conchalí, Quilicura, Tiltil
Oriente	Las Condes, La Reina, Lo Barnechea, Macul, Ñuñoa, Peñalolén, Providencia, Vitacura
Centro	Santiago, Estación Central, Pedro Aguirre Cerda, San Joaquín, San Miguel
Poniente	Cerro Navia, Lo Prado, Maipú, Cerrillos, Pudahuel, Quinta Normal, Renca
Sur	La Florida, Puente Alto, San Bernardo, La Pintana, El Bosque, La Cisterna, La Granja, Lo Espejo, San Ramón
Zona Sub Urbana	Curacaví, Lampa, Pirque, San Jose de Maipo, Alhue, El Monte, Isla de Maipo, Maria Pinto, Melipilla, Peñaflo, Padre Hurtado, San pedro, Talagante, Buin, Calera de Tango, Paine.

Considerando el análisis por genero de los individuos, no se aprecia una diferencia relevantes entre cada uno, lo cual indica que la incidencia de compra entre hombres y mujeres se rige por efectos similares.

Tabla 9: Parámetros variables descriptivas por genero Mixed Logit.

	Masculino	Femenino
Distancia	-2,97	-2,86
Distancia al Cuadrado	-0,42	-0,42
Mall	-4,44	-4,48
Participación Poblacional	-0,36	-0,44
Distancia Tienda Nueva	-1	-0,98
Antigüedad	-1,6	-1,51
Total Clientes	259	241

8.4.2) Preferencia de los Consumidores

Un tema relevante de analizar es la preferencia que tienen los consumidores con respecto a las distintas tiendas de Falabella que están disponibles en su conjunto de elección. Para identificar esto, se utiliza la información obtenida en los interceptos de la función de utilidad, en donde cabe recordar que está ecuación está separada en dos partes (ver ecuación 35).

El intercepto representa la preferencia o agrado del consumidor sobre una tienda con respecto a la opción de “No Comprar”, ya que ésta es la alternativa que se fijó en cero para poder ejecutar el modelo. En otras palabras y debido a que la opción de “No Comprar” está en negación, los interceptos que tengan valores positivos indican una preferencia menor hacia la tienda y viceversa para el caso de interceptos con valores negativos.

Revisando los resultados entregados por el modelo, las tiendas antiguas de Falabella que tienen preferencia positiva por parte de los consumidores son Lyon con -3.51 y Centro con -2.33, escenario contrario obtienen las tiendas Alto Las Condes con 0.06, Parque Arauco con 0.19, Plaza Vespucio con 1.77 y Plaza Oeste con 1.29. Esta situación puede deberse a que las últimas tiendas mencionadas tienen las principales participaciones en monto entonces tienen una mayor probabilidad de perder ventas por el tamaño que representan.

A su vez, se aprecia que el agrado que manifiestan los consumidores hacia las tiendas de Falabella aumenta con el tiempo, es decir la imagen de la cadena mejora con el ingreso de nuevas tiendas, fortaleciendo la preferencia de los clientes sobre Lyon y Centro y mejorando la percepción de agrado de los clientes sobre las tiendas Plaza Vespucio y Plaza Oeste, en donde éstas nunca llegan a ser percibidas por los consumidores como gratas en todo el horizonte de tiempo:

- Alto Las Condes: 0.06, -0.10, -0.79, -1.26, -1.63 y -1.76.
- Parque Arauco: 0.19, 0.05, -0.54, -0.78, -0.85 y -0.63.
- Lyon: -3.51, -4.25, -4.83, -5.23, -5.58 y -5.37.
- Centro: -2.33, -2.34, -3.33, -3.76, -3.95 y -3.91.
- Tobalaba: 0.31, 0.28, -0.67, -1.36, -1.83 y -1.87.
- Plaza Vespucio: 1.77, 2.32, 1.37, 0.81, 0.60 y 0.48.
- Plaza Oeste: 1.29, 1.73, 0.54, 0.10, 0.06 y 0.09.
- Plaza Norte: -1.09, -1.03, -1.04, -0,37 y 0,10.

- Plaza Puente: -0.82, -1.32, -0.20 y -0.17.
- La Dehesa: -2.43, -1.08 y -0.35.
- Manquehue: -2.10 y -0.77.

Tabla 10: Interceptos tiendas Falabella.

Interceptos Tiendas Falabella.												
	Alto Las Condes	Parque Arauco	Lyon	Centro	Tobalaba	Plaza Vespucio	Plaza Oeste	Plaza Norte	Plaza Puente	La Dehesa	Manquehue	Estación Central
Intercepto Cero	0,06	0,19	-3,51	-2,33	0,31	1,77	1,29					
Ingreso Plaza Norte	-0,16	-0,14	-0,74	-0,01	-0,03	0,55	0,43	-1,09				
Total Después Ingreso	-0,1	0,05	-4,25	-2,34	0,28	2,32	1,73	-1,09				
Ingreso Plaza Puente	-0,69	-0,59	-0,58	-0,99	-0,95	-0,96	-1,18	-1,03	-0,82			
Total Después Ingreso	-0,79	-0,54	-4,83	-3,33	-0,67	1,37	0,54	-2,12	-0,82			
Ingreso La Dehesa	-0,47	-0,24	-0,4	-0,43	-0,69	-0,56	-0,45	-1,04	-1,32	-2,43		
Total Después Ingreso	-1,26	-0,78	-5,23	-3,76	-1,36	0,81	0,1	-3,16	-2,14	-2,43		
Ingreso Manquehue	-0,37	-0,07	-0,35	-0,2	-0,48	-0,21	-0,04	-0,37	-0,2	-1,08	-2,1	
Total Después Ingreso	-1,63	-0,85	-5,58	-3,95	-1,83	0,6	0,06	-3,53	-2,34	-3,51	-2,1	
Ingreso Estación Central	-0,13	0,22	0,21	0,04	-0,03	-0,12	0,03	0,1	-0,17	-0,35	-0,77	-1,2
Total Después Ingreso	-1,76	-0,63	-5,37	-3,91	-1,87	0,48	0,09	-3,43	-2,51	-3,86	-2,87	-1,2

Por otro lado, todas las tiendas nuevas, es decir tiendas con apertura posterior al año 2000, obtuvieron preferencias positivas por los consumidores: -1.09 Plaza Norte, -0.82 Plaza Puente, -2.43 La Dehesa, -2.10 Manquehue y -1.20 Estación Central. Estos valores son coherentes a lo esperado porque todas las tiendas nuevas realizan esfuerzos y gastos importantes en publicidad y ofertas en las inauguraciones que permitan tener la atención de los clientes (Kumar, Pancras y Sruram, 2009 y 2012; Gupta, 2008), siendo validado lo anterior al ver que los mayores valores son en aquellas tiendas ubicadas en zonas de mayores recursos económicos.

Como resumen de los resultados asociados a las preferencias de los consumidores hacia las tiendas de Falabella, se tiene que siempre hay una preferencia positiva hacia las tiendas nuevas y que estas últimas también ayudan a mejorar las preferencias de los clientes hacia las tiendas ya establecidas con

anterioridad. Además, se aprecia que las tiendas anteriores al año 2000 con mayor participación tienen una mayor pérdida de su popularidad al ingresar nuevas tiendas, es decir les impacta con mayor fuerza las nuevas aperturas.

Aprovechando la ventaja de tener parámetros para cada uno de los clientes, se realiza una separación por zonas dentro de la región metropolitana para los análisis: Zona norte, poniente, sur, centro, oriente y sub-urbana).

Tabla 11: Interceptos tiendas Falabella zona norte.

INTERCEPTOS ZONA NORTE												
	Alto Las Condes	Parque Arauco	Lyon	Centro	Tobalaba	Plaza Vespucio	Plaza Oeste	Plaza Norte	Plaza Puente	La Dehesa	Manquehue	Estación Central
Intercepto Cero	0,08	0,77	-3,97	-0,96	-0,18	0,74	0,91	-----	-----	-----	-----	-----
Ingreso Plaza Norte	-0,42	-0,17	-1,23	0,36	-0,62	0,3	0,27	1,1	-----	-----	-----	-----
Total Después Ingreso	-0,34	0,61	-5,2	-0,61	-0,8	1,04	1,17	1,1				
Ingreso Plaza Puente	-0,62	-0,78	-0,62	-1,17	-0,97	-1,11	-1,42	-0,74	-0,31	-----	-----	-----
Total Después Ingreso	-0,96	-0,17	-5,82	-1,78	-1,77	-0,07	-0,25	0,36	-0,31			
Ingreso La Dehesa	-0,58	-0,41	-0,05	-0,21	-0,9	-0,44	-0,48	-0,4	-0,66	-2,01	-----	-----
Total Después Ingreso	-1,54	-0,58	-5,87	-1,99	-2,66	-0,51	-0,73	-0,04	-0,98	-2,01		
Ingreso Manquehue	-0,45	0,16	-0,2	-0,7	-0,44	-0,23	-0,22	-0,22	-0,18	-0,94	-2,33	-----
Total Después Ingreso	-1,99	-0,42	-6,07	-2,69	-3,1	-0,74	-0,95	-0,25	-1,16	-2,95	-2,33	
Ingreso Estación Central	-0,07	0,25	0,12	-0,01	0,17	-0,07	0	0,08	-0,05	-0,46	-0,87	-1,08
Total Después Ingreso	-2,06	-0,17	-5,95	-2,7	-2,93	-0,81	-0,95	-0,18	-1,21	-3,41	-3,19	-1,08

Para el caso de la zona norte, todas las tiendas ingresadas antes del año 2000 tienen preferencia negativa sólo Lyon (-3.97), Centro (-0.96) y Tobalaba (-0.18) que obtienen preferencias positivas, en donde las 2 últimas tienen valores muy cercanos a cero lo que indicaría cierta indiferencia. Revisando las nuevas tiendas de Falabella, éstas obtuvieron preferencia positiva en sus inauguraciones salvo Plaza Norte (1.10).

Tabla 12: Interceptos tiendas Falabella zona poniente.

	INTERCEPTOS ZONA PONIENTE											
	Alto Las Condes	Parque Arauco	Lyon	Centro	Tobalaba	Plaza Vespucio	Plaza Oeste	Plaza Norte	Plaza Puente	La Dehesa	Manquehue	Estación Central
Intercepto Cero	-0,37	-0,06	-3,5	-1,72	-0,39	1,16	2,74					
Ingreso Plaza Norte	-0,36	-0,42	-0,82	0,46	-0,41	0,38	1	-0,76				
Total Después Ingreso	-0,73	-0,49	-4,32	-1,25	-0,8	1,54	3,75	-0,76				
Ingreso Plaza Puente	-0,73	-0,44	-0,43	-1,03	-1,01	-1,03	-1,21	-1,07	-0,55			
Total Después Ingreso	-1,46	-0,92	-4,75	-2,28	-1,81	0,51	2,53	-1,82	-0,55			
Ingreso La Dehesa	-0,48	-0,33	-0,27	-0,28	-0,81	-0,47	-0,24	-0,84	-0,93	-2,48		
Total Después Ingreso	-1,94	-1,26	-5,02	-2,56	-2,62	0,04	2,3	-2,66	-1,48	-2,48		
Ingreso Manquehue	-0,38	-0,15	-0,28	-0,4	-0,48	-0,36	0,06	-0,4	-0,25	-1,14	-2,21	
Total Después Ingreso	-2,32	-1,41	-5,3	-2,96	-3,1	-0,32	2,36	-3,06	-1,73	-3,62	-2,21	
Ingreso Estación Central	-0,05	0,28	0	0,03	0,01	-0,29	-0,02	0,1	-0,26	-0,63	-0,74	-0,74
Total Después Ingreso	-2,37	-1,13	-5,3	-2,93	-3,09	-0,61	2,34	-2,97	-2	-4,25	-2,95	-0,74

Considerando la zona poniente, todas las tiendas ingresadas antes del año 2000 tienen preferencia positiva salvo Plaza Vespucio (1.16) y Plaza Oeste (2.74). Además, todas las nuevas tiendas de Falabella obtuvieron preferencia positiva en sus inauguraciones, sobre saliendo La Dehesa (-2.48) y Manquehue (-2.21).

Tabla 13: Interceptos tiendas Falabella zona sur.

	INTERCEPTOS ZONA SUR											
	Alto Las Condes	Parque Arauco	Lyon	Centro	Tobalaba	Plaza Vespucio	Plaza Oeste	Plaza Norte	Plaza Puente	La Dehesa	Manquehue	Estación Central
Intercepto Cero	-0,45	-0,72	-3,94	-2,69	1,48	2,94	1,23	-----	-----	-----	-----	-----
Ingreso Plaza Norte	-0,46	-0,45	-1,03	-0,31	0,58	0,98	0,42	-1,67	-----	-----	-----	-----
Total Después Ingreso	-0,91	-1,17	-4,98	-3	2,06	3,93	1,65	-1,67				
Ingreso Plaza Puente	-0,78	-0,54	-0,56	-1	-0,81	-1	-1	-1,16	-0,95	-----	-----	-----
Total Después Ingreso	-1,69	-1,71	-5,54	-4	1,25	2,93	0,65	-2,83	-0,95			
Ingreso La Dehesa	-0,49	-0,06	-0,7	-0,64	-0,56	-0,41	-0,57	-1,14	-1,68	-2,64	-----	-----
Total Después Ingreso	-2,18	-1,77	-6,24	-4,64	0,69	2,52	0,08	-3,97	-2,63	-2,64		
Ingreso Manquehue	-0,36	-0,04	-0,5	-0,09	-0,54	-0,16	-0,16	-0,25	-0,15	-1,17	-2,21	-----
Total Después Ingreso	-2,54	-1,81	-6,74	-4,73	0,15	2,35	-0,08	-4,22	-2,78	-3,81	-2,21	
Ingreso Estación Central	-0,21	0,3	0,38	0,21	-0,09	-0,05	0,16	0,12	-0,23	-0,3	-1,04	-1,49
Total Después Ingreso	-2,75	-1,51	-6,36	-4,52	0,05	2,31	0,09	-4,1	-3,01	-4,11	-3,25	-1,49

Para la zona sur, las tiendas Alto Las Condes (-0.45), Parque Arauco (-0.72), Lyon (-3.94) y Centro (-2.69) tienen preferencias positivas de los consumidores. A su vez, todas las tiendas inauguradas con posterioridad al 2000 tienen preferencias positivas. La zona centro tiene el mismo comportamiento que la zona sur salvo que la tienda Parque Arauco (0.03) tiene preferencia negativa de los consumidores.

Tabla 14: Interceptos tiendas Falabella zona centro.

	INTERCEPTOS ZONA CENTRO											
	Alto Las Condes	Parque Arauco	Lyon	Centro	Tobalaba	Plaza Vespucio	Plaza Oeste	Plaza Norte	Plaza Puente	La Dehesa	Manquehue	Estación Central
Intercepto Cero	-0,33	0,03	-3,32	-1,63	-0,18	2,1	1,3	-----	-----	-----	-----	-----
Ingreso Plaza Norte	-0,33	-0,31	-0,57	0,37	-0,21	0,58	0,48	-1,09	-----	-----	-----	-----
Total Después Ingreso	-0,67	-0,27	-3,89	-1,26	-0,39	2,69	1,78	-1,09				
Ingreso Plaza Puente	-0,62	-0,43	-0,59	-1,03	-0,95	-1,03	-1,26	-1,02	-0,58	-----	-----	-----
Total Después Ingreso	-1,28	-0,71	-4,49	-2,28	-1,34	1,66	0,53	-2,11	-0,58			
Ingreso La Dehesa	-0,34	-0,31	-0,37	-0,35	-0,72	-0,44	-0,55	-0,97	-0,97	-2,54	-----	-----
Total Después Ingreso	-1,63	-1,01	-4,86	-2,63	-2,06	1,22	-0,02	-3,08	-1,55	-2,54		
Ingreso Manquehue	-0,46	0,04	-0,35	-0,17	-0,6	-0,25	-0,05	-0,41	-0,17	-1,19	-2,11	-----
Total Después Ingreso	-2,09	-0,98	-5,21	-2,8	-2,66	0,97	-0,07	-3,49	-1,72	-3,72	-2,11	
Ingreso Estación Central	-0,07	0,19	0,08	-0,04	-0,03	-0,14	0,07	-0,03	-0,16	-0,44	-0,75	-0,83
Total Después Ingreso	-2,15	-0,78	-5,13	-2,84	-2,69	0,84	0	-3,52	-1,88	-4,16	-2,86	-0,83

Considerando la zona oriente, las únicas tiendas antiguas que tienen preferencias positivas son Lyon (-2.93), Centro (-3.11) y Tobalaba (-0.20), en donde predomina la tienda Centro. Las tiendas nuevas obtuvieron buenas preferencias de los consumidores.

Tabla 15: Interceptos tiendas Falabella zona oriente.

	INTERCEPTOS ZONA ORIENTE											
	Alto Las Condes	Parque Arauco	Lyon	Centro	Tobalaba	Plaza Vespucio	Plaza Oeste	Plaza Norte	Plaza Puente	La Dehesa	Manquehue	Estación Central
Intercepto Cero	1,26	1,44	-2,93	-3,11	-0,2	1	0,17	-----	-----	-----	-----	-----
Ingreso Plaza Norte	0,55	0,58	-0,22	-0,33	-0,19	0,22	-0,04	-1,29	-----	-----	-----	-----
Total Después Ingreso	1,8	2,02	-3,16	-3,44	-0,39	1,21	0,13	-1,29				
Ingreso Plaza Puente	-0,58	-0,84	-0,72	-0,86	-1,06	-0,78	-1,27	-0,92	-1,12	-----	-----	-----
Total Después Ingreso	1,22	1,18	-3,88	-4,3	-1,45	0,44	-1,14	-2,21	-1,12			
Ingreso La Dehesa	-0,45	-0,32	-0,24	-0,38	-0,68	-0,93	-0,4	-1,34	-1,54	-2,19	-----	-----
Total Después Ingreso	0,77	0,86	-4,12	-4,68	-2,13	-0,49	-1,54	-3,55	-2,66	-2,19		
Ingreso Manquehue	-0,32	-0,16	-0,26	0	-0,34	-0,13	0,07	-0,51	-0,21	-0,9	-1,8	-----
Total Después Ingreso	0,45	0,71	-4,38	-4,67	-2,47	-0,62	-1,47	-4,06	-2,87	-3,09	-1,8	
Ingreso Estación Central	-0,14	0,07	0,25	-0,12	-0,04	-0,08	-0,11	0,14	-0,06	-0,08	-0,43	-1,44
Total Después Ingreso	0,31	0,77	-4,13	-4,79	-2,51	-0,71	-1,58	-3,92	-2,93	-3,17	-2,24	-1,44

Por último, se tomó en cuenta las comunas ubicadas en la zona sub-urbana, en donde todas las tiendas antiguas tienen preferencias positivas salvo Plaza Vespucio (0.96) y Plaza Oeste (3.42). Llama la atención que la tienda Plaza oeste tiene una elevada desconformidad en este grupo de consumidores. Al igual que en las otras zonas, las tiendas inauguradas tienen preferencias positivas.

Tabla 16: Interceptos tiendas Falabella zona sub urbana.

	INTERCEPTOS ZONA SUB-URBANA											
	Alto Las Condes	Parque Arauco	Lyon	Centro	Tobalaba	Plaza Vespucio	Plaza Oeste	Plaza Norte	Plaza Puente	La Dehesa	Manquehue	Estación Central
Intercepto Cero	-0,54	-0,18	-3,76	-1,8	-0,22	0,96	3,42	-----	-----	-----	-----	-----
Ingreso Plaza Norte	-0,5	-0,49	-1,09	0,51	-0,41	0,41	1,28	-0,49	-----	-----	-----	-----
Total Después Ingreso	-1,05	-0,67	-4,85	-1,28	-0,63	1,37	4,71	-0,49				
Ingreso Plaza Puente	-0,78	-0,39	-0,37	-1,1	-1,02	-1	-1,23	-1,11	-0,65	-----	-----	-----
Total Después Ingreso	-1,82	-1,06	-5,22	-2,38	-1,65	0,37	3,48	-1,6	-0,65			
Ingreso La Dehesa	-0,57	-0,2	-0,27	-0,28	-0,84	-0,32	-0,23	-0,67	-1,08	-2,48	-----	-----
Total Después Ingreso	-2,39	-1,26	-5,49	-2,67	-2,49	0,05	3,25	-2,27	-1,73	-2,48		
Ingreso Manquehue	-0,32	-0,17	-0,33	-0,63	-0,44	-0,3	0,04	-0,28	-0,36	-1,13	-2,31	-----
Total Después Ingreso	-2,71	-1,43	-5,82	-3,29	-2,93	-0,25	3,29	-2,55	-2,1	-3,61	-2,31	
Ingreso Estación Central	0	0,37	0,01	0,05	-0,02	-0,28	-0,03	0,19	-0,37	-0,72	-0,85	-0,89
Total Después Ingreso	-2,7	-1,06	-5,81	-3,24	-2,95	-0,53	3,26	-2,36	-2,46	-4,33	-3,15	-0,89

8.4.3) Canibalización

Uno de los objetivos principales de este trabajo de investigación es analizar la canibalización al momento de ingresar una nueva tienda de la misma cadena al conjunto de elección de los consumidores, lo cual se identifica revisando la variación en la participación de mercado que tiene cada tienda de Falabella.

Para determinar la canibalización, se calcula la participación de mercado para cada tienda en los periodos “Antes” y “Después” del ingreso de las nuevas alternativas (tiendas). Específicamente, en ambos periodos se calcula para cada uno de los clientes la utilidad que le proporcionan las alternativas disponibles en ese momento, en donde se obtiene como resultado de asistencia la tienda que le produce una mayor utilidad.

Es importante señalar que considerando el diseño del modelo Mixed Logit, la única parte del modelo que varía entre ambos periodos son los interceptos que reflejan el atractivo que tiene el cliente hacia cada una de las tiendas. Esta variación se produce al activarse la segunda parte del intercepto, capturando la

influencia que tiene la nueva tienda sobre el intercepto general de las tiendas ya establecidas (ecuación 35).

Como se menciona anteriormente, en este trabajo no se agregan nuevos clientes a lo largo del tiempo debido al filtro número 11 realizado a la base de datos, entonces debido a esto cada una de las columnas de participación de las tiendas en los periodos “Antes” y “Después” siempre suman el 100%. Además, el foco principal del trabajo es analizar el comportamiento de clientes antiguos de Falabella y no clientes nuevos.

Al analizar la canibalización, la comparación entre los dos periodos siempre es utilizando el 100% contra el 100% de la muestra. De esta manera, el efecto de variación del mercado total afecta de igual forma a todas las tiendas, es decir este efecto externo no ensucia los resultados obtenidos.

Tabla 17: Canibalización tiendas Falabella.

Canibalización Tiendas Falabella															
Tiendas	Antes	Después	Caniba.												
Alto las Condes	7,80%	8,40%	0,60%	8,40%	10,00%	1,60%	10,00%	9,00%	-1,00%	9,00%	8,20%	-0,80%	8,20%	7,60%	-0,60%
Parque Arauco	7,20%	7,20%	0,00%	7,20%	6,00%	-1,20%	6,00%	6,20%	0,20%	6,20%	6,60%	0,40%	6,60%	7,20%	0,60%
Lyon	8,00%	7,40%	-0,60%	7,40%	7,20%	-0,20%	7,20%	7,60%	0,40%	7,60%	7,80%	0,20%	7,80%	7,40%	-0,40%
Centro	25,80%	20,80%	-5,00%	20,80%	19,80%	-1,00%	19,80%	19,80%	0,00%	19,80%	17,60%	-2,20%	17,60%	15,80%	-1,80%
Tobalaba	8,80%	8,40%	-0,40%	8,40%	9,20%	0,80%	9,20%	9,00%	-0,20%	9,00%	8,80%	-0,20%	8,80%	8,60%	-0,20%
Plaza Vespucio	21,60%	22,80%	1,20%	22,80%	22,60%	-0,20%	22,60%	22,20%	-0,40%	22,20%	22,60%	0,40%	22,60%	21,60%	-1,00%
Plaza Oeste	20,80%	21,20%	0,40%	21,20%	20,20%	-1,00%	20,20%	19,80%	-0,40%	19,80%	20,40%	0,60%	20,40%	21,00%	0,60%
Plaza Norte		3,80%	3,80%	3,80%	4,20%	0,40%	4,20%	3,80%	-0,40%	3,80%	4,20%	0,40%	4,20%	4,60%	0,40%
Plaza Puente					0,80%	0,80%	0,80%	1,60%	0,80%	1,60%	2,80%	1,20%	2,80%	3,20%	0,40%
Dehesa								1,00%	1,00%	1,00%	1,00%	0,00%	1,00%	1,00%	0,00%
Manquehue											0,00%	0,00%	0,00%	0,60%	0,60%
Estación Central														1,40%	1,40%

Al momento de ingresar la tienda Plaza Norte, ésta obtiene una participación de 3.8%, en donde la tienda antigua más afectada es Centro que disminuyó 5.0% su participación. Por otro lado, la entrada de Plaza Puente produce una participación de 0.8%, en donde se identifica una disminución en la participación para las tiendas de Parque Arauco (-1.2%), Centro (-1.0%) y Plaza Oeste (-1.0%) pero la tienda Alto Las Condes tiene un aumento en su participación de 1.6%.

La entrada de la tienda La Dehesa obtiene una participación de mercado de 1,0%, en donde Alto Las Condes tiene el mismo porcentaje de disminución. A su vez, la tienda Manquehue tiene un 0% de participación, lo cual se explica porque esta tienda tuvo un periodo de “marcha blanca” desde la fecha de inauguración en donde no se produce un elevado número de ventas.

Por último, la apertura de Estación Central genera un 1.4% de participación de mercado, siendo las principales tiendas afectadas con un decrecimiento en su participación son Centro (-1.8%) y Plaza Vespucio (-1.0%). Ver la totalidad de los resultados en la Tabla N° 17.

Otro análisis importante es que la tienda Centro, una de las tiendas con mayor participación inicial, está constantemente perdiendo relevancia a medida que ingresaban nuevas tiendas, partiendo con 25.8%, y terminando con 15.8%. Situación diferente ocurrió para Plaza Puente que se inaugura con un 0.8% y finaliza el periodo en 3.2% de participación, es decir tiene un crecimiento sostenido durante el tiempo.

Además, al revisar el historial de canibalización de las tiendas, se aprecia que los valores de este fenómeno para cada una de ellas varía con el tiempo, lo cual tiene sentido porque este fenómeno va respondiendo a los factores de la situación actual.

Tabla 48: Distancia Euclidiana entre tiendas Falabella.

Tiendas	Alto Las Condes	Parque Arauco	Lyon	Santiago Centro	Plaza Tobalaba	Plaza Vespucio	Plaza Oeste	Plaza Norte	Plaza Puente	La Dehesa	Manquehue	Estación Central
Alto las Condes	0	0,04	0,07	0,12	0,18	0,14	0,21	0,13	0,11	0,04	0,03	0,15
Parque Arauco	----	0	0,04	0,08	0,17	0,12	0,18	0,1	0,08	0,08	0,02	0,12
Lyon	----	----	0	0,05	0,16	0,1	0,14	0,08	0,04	0,11	0,05	0,08
Centro	----	----	----	0	0,16	0,09	0,1	0,08	0,01	0,16	0,09	0,03
Tobalaba	----	----	----	----	0	0,07	0,17	0,23	0,17	0,21	0,16	0,17
Plaza Vespucio	----	----	----	----	----	0	0,12	0,17	0,1	0,18	0,11	0,11
Plaza Oeste	----	----	----	----	----	----	0	0,15	0,1	0,26	0,18	0,07
Plaza Norte	----	----	----	----	----	----	----	0	0,07	0,16	0,12	0,08
Plaza Puente	----	----	----	----	----	----	----	----	0	0,16	0,09	0,04
Dehesa	----	----	----	----	----	----	----	----	----	0	0,07	0,19
Manquehue	----	----	----	----	----	----	----	----	----	----	0	0,12
Estación Central	----	----	----	----	----	----	----	----	----	----	----	0

Por último, tomando en cuenta los resultados de canibalización señalados al inicio, se identifica que las tiendas que fueron más afectadas con cada ingreso tienen la menor distancia euclidiana entre ellas, lo cual ratifica que la distancia es un factor importante para la canibalización (Bell, Ho y Tang, 1998; Dekimpe, Delesnyder, Geyskens y Gielens, 2002); Kumar, Pancras y Sriram, 2009 y 2012)

8.4.4) Hit Rate

Luego de desarrollar cualquier modelo, es muy relevante analizar el nivel de predicción que tiene el modelo con respecto a la realidad (medida de ajuste), característica que le entrega validez al trabajo. Es por ello que se elabora una herramienta llamada Hit Rate, la cual calcula el grado de acierto que indica el modelo de incidencia de compra, es decir, calcula el lugar donde irá cada consumidor considerando que el individuo asiste a la alternativa que le proporciona una mayor utilidad, luego compara este resultado con el número de veces que realmente la persona compra en esa tienda.

Otra forma de interpretar el Hit Rate es como la probabilidad con que el modelo de incidencia de compra predice la tienda que visita un consumidor, en dónde por ejemplo si un cliente ha realizado 8 de sus 10 compras en Alto las Condes y el modelo selecciona que el cliente asiste a la tienda mencionada, el Hit Rate de ese cliente es 80%.

Tabla 19: Hit Rate.

Ingreso Plaza Norte			Ingreso Plaza Puento			Ingreso La Dehesa			Ingreso Manquehue			Ingreso Estación Central		
Índice de Acierto	Cantidad Clientes	Porcen.	Índice de Acierto	Cantidad Clientes	Porcen.	Índice de Acierto	Cantidad Clientes	Porcen.	Índice de Acierto	Cantidad Clientes	Porcen.	Índice de Acierto	Cantidad Clientes	Porcen.
0	0	0%	0	0	0%	0	0	0%	0	0	0%	0	0	0%
0,1	0	0%	0,1	0	0%	0,1	0	0%	0,1	0	0%	0,1	0	0%
0,2	3	1%	0,2	2	0%	0,2	4	1%	0,2	6	1%	0,2	11	2%
0,3	13	3%	0,3	9	2%	0,3	9	2%	0,3	14	3%	0,3	28	6%
0,4	36	7%	0,4	28	6%	0,4	31	6%	0,4	37	7%	0,4	34	7%
0,5	46	9%	0,5	50	10%	0,5	43	9%	0,5	38	8%	0,5	34	7%
0,6	41	8%	0,6	45	9%	0,6	47	9%	0,6	42	8%	0,6	38	8%
0,7	49	10%	0,7	54	11%	0,7	56	11%	0,7	55	11%	0,7	52	10%
0,8	78	16%	0,8	78	16%	0,8	76	15%	0,8	74	15%	0,8	70	14%
0,9	63	13%	0,9	63	13%	0,9	63	13%	0,9	63	13%	0,9	62	12%
1	171	34%	1	171	34%	1	171	34%	1	171	34%	1	171	34%
500 100%			500 100%			500 100%			500 100%			500 100%		
Hit Rate Promedio: 78%			Hit Rate Promedio: 79%			Hit Rate Promedio: 79%			Hit Rate Promedio: 78%			Hit Rate Promedio: 77%		

Los resultados obtenidos por la herramienta Hit Rate para los ingresos de cada una de las tiendas se presentan en la Tabla N° 19 y son los siguientes:

- Plaza Norte: 78%.
- Plaza Puente: 79%.
- La Dehesa: 79%.
- Manquehue: 78%.
- Estación Central: 77%.

Se tiene que en promedio, el modelo tiene un Hit Rate de 78.2%, lo cual es muy bueno porque significa que predice el 78.2% de las visitas de los consumidores.

9) Modelo Monto de Compra

El segundo efecto que se quiere estudiar con el ingreso de una nueva tienda, es la potencial variación en los montos de compra de los consumidores a través de identificar los factores influyentes de este fenómeno, para analizar la variación existente. Para esto, se desarrolla un modelo de regresión lineal utilizando un enfoque bayesiano, el cual se procede a explicar a continuación.

9.1) Metodología

Para desarrollar el modelo de monto señalado, se utiliza una regresión lineal con un enfoque bayesiano, lo cual corresponde a estimar los parámetros de las variables explicativas de la función a través de una densidad de probabilidad. El inconveniente de este enfoque es la dificultad para resolver integrales múltiples, es por ello que se procede a estimar los parámetros con una herramienta de simulación llamada Muestreo de Gibbs (Kai, 1998; Ho-Chuan River Huang, 2001).

La metodología para este modelo es estimar los parámetros de las variables utilizando la totalidad de las transacciones en la muestra de clientes, es decir no se realiza una estimación individual para cada consumidor. De esta forma se obtienen parámetros agregados. Al igual como ocurre para el modelo de incidencia de compra, el enfoque bayesiano tiene graves dificultades en su ejecución debido a integrales múltiples. Es por ello que se procede a utilizar el método Muestreo de Gibbs para trabajar las probabilidades y obtener los resultados (Ho-Chuan y Huang, 2010), en donde no es necesaria la maximización de ninguna función. Ver Anexo N° 20.

Por otro lado, se busca realizar un análisis individual para 5 nuevas tiendas de Falabella ingresadas desde el año 2002 hasta el 2009 (Plaza Norte, Plaza Puente, La Dehesa, Manquehue y Estación Central). Para ellos, se obtienen parámetros específicos para cada una de las aperturas (periodos).

Teniendo los parámetros de las variables para cada una de las entradas, se procede a analizar la variación del monto promedio del cliente en los periodos “Antes” y “Después” de la entrada de una nueva tienda. Al señalar un monto promedio, se refiere a estimar el monto en cada una de sus incidencias de compra considerando la información transaccional hasta ese instante. Luego, con esta

información se calcula un monto promedio antes y un monto promedio después de la apertura de la nueva tienda, de esta forma se obtiene la variación.

9.2) Formulación Modelo de Regresión Lineal Bayesiano

Como se menciona anteriormente, se realiza un modelo de regresión lineal, por tanto la función de utilidad debe tener la forma señalada en la ecuación 36.

$$Monto_t = \beta_1 Monto \acute{U}ltima Compra_{t-1} + \beta_2 Monto Promedio_{t-1} + \beta_3 Mall + \beta_4 N^\circ Periodos \acute{U}ltima Compra_t + \varepsilon_t \quad (36)$$

En donde las variables explicativas señaladas en la función de utilidad significan lo siguiente:

- *Monto Última Compra_{t-1}*: Como lo dice el nombre, es el monto de la última transacción realizada por el consumidor.
- *Monto Promedio_t*: Esta variable corresponde al monto promedio hasta el periodo “t” considerando el historial de cada consumidor.
- *Mall*: Esta es una variable binaria en donde el valor 1 corresponde a que la transacción fue realizada en una tienda que está dentro de un mall (formato de tienda mall) y el valor 0 significa que la tienda tiene formato calle.
- *N° Periodos Última Compra_t*: Diferencia temporal entre dos compras consecutivas, es decir el número de unidades temporales entre compra en el periodo “t” y la realizada inmediatamente anterior.

Tomando en cuenta las variables seleccionadas y la bibliografía revisada anteriormente, se tiene la hipótesis de que los signos de los parámetros asociados a las variables descriptivas del modelo de monto son:

- Monto última compra: Positivo.

- Monto promedio: Positivo.
- Mall: Negativo.
- N° periodos última compra: Positivo.

Tabla 20: Hipótesis signos variables explicativas modelo monto.

Variable	Signo
Monto Última Compra	+
Monto Promedio	+
Mall	-
N° Periodos Última Compra	+

9.3) Formato Base de Datos Utilizados

Para ejecutar la regresión lineal con enfoque bayesiano para el modelo de monto, se utiliza el formato original de la muestra que se utiliza para el modelo inicial de incidencia de compra, ya que en esta oportunidad los parámetros que se estiman son a nivel agregado, es decir uno para todos los clientes de la muestra.

El enfoque anterior entrega una mayor facilidad en el manejo de datos porque cada una de las filas de la muestra poblacional corresponde a una compra realizada y sólo se indica la tienda en donde se realiza y no todo el conjunto de elección de cada consumidor. En la Tabla N°21 se muestra un ejemplo del formato utilizado para la ejecución del modelo de monto, en donde se presentan 5 compras realizadas en Enero del año 2003.

Tabla 21: Abstracto base de datos modelo Monto.

Monto Predicho (Y)	Monto Ult. Compra (X1)	Monto Promedio (X2, t-1)	Mall (X3)	Meses Sin Comprar (X4)
\$ 34.900	0	34900	1	0
\$ 5.980	39364	22672	1	4
\$ 324.140	110892	151674	1	3
\$ 9.990	0	9990	1	0
\$ 32.981	73960	53471	1	3

9.4) Resultados y Análisis

Para estimar los parámetros de la función de utilidad se utiliza el programa Matlab y se ocupa la información incluida en la boletas de cada uno de los clientes. A través de la información transaccional se obtiene una valorización cuantitativa para los parámetros e interpretación cualitativa para los signo de los parámetros de las variables explicativas del monto de compra.

Tabla 22: Parámetros variables explicativas modelo Monto.

Variables	Plaza Norte		Plaza Puente		La Dehesa		Manquehue		Estación Central	
	Antes	Después	Antes	Después	Antes	Después	Antes	Después	Antes	Después
Monto Última Compra	-0,362	-0,272	-0,163	-0,197	-0,172	-0,173	-0,175	-0,096	-0,096	-0,101
Monto Promedio	1,157	1,253	1,161	1,267	1,176	1,259	1,255	1,24	1,24	1,366
Mall	0,003	-0,001	0,001	-0,004	-0,001	0,001	-0,005	0,001	0,004	0
Meses Sin Compras	0,014	0,003	-0,006	0,005	-0,006	0,002	0,006	0,005	0,003	0,002
N° Observaciones	1425	1475	1492	1389	1506	1391	1340	1367	1367	1473

En la Tabla N° 22 se aprecian los valores luego de desarrollar el modelo de monto, en donde la variable que tiene una mayor importancia sobre el monto es la variable “Monto Promedio” (Gupta, 2008), seguida por la variable “Monto Última Compra”. Por otro lado, se tiene que las variables “Mall” y “Meses Sin Compras” no tienen un peso predominante porque son muy cercanos a cero, significando que el formato de la tienda tiene poco impacto sobre el monto (Pancras, Sriram y Kumar, 2009) y que el tener un tiempo mayor sin compras no es relevante sobre la variable estudiada.

A su vez, al analizar el estadístico “t-probability” asociado a la significancia de las variables utilizadas, coincide el resultado con el valor cuantitativo de los parámetros, ya que las variables representativas son “Monto Promedio” y “Monto Última Compra”. Ver Tabla N° 23.

Tabla 23: Estadístico t-probability modelo Monto.

Variables	Plaza Norte		Plaza Puente		La Dehesa		Manquehue		Estación Central	
	Antes	Después	Antes	Después	Antes	Después	Antes	Después	Antes	Después
Monto Última Compra	0	0	0	0	0	0	0	0	0	0
Monto Promedio	0	0	0	0	0	0	0	0	0	0
Mall	0,998	0,999	0,998	0,997	1	0,999	0,996	0,999	0,997	1
Meses Sin Compras	0,989	0,998	0,997	0,996	0,995	0,998	0,996	0,006	0,998	0,998

Otro punto relevante a la hora de ver los resultados, son los signos asociados a los parámetros, los cuales indican la dirección de influencia que tiene la variables sobre el monto de compra. Las variables principales tienen signos apostados sobre el monto para todas las tiendas nuevas, tanto en los periodos “Antes” y “Después” de las aperturas, siendo “Monto Última Compra” indirectamente proporcional y “Monto Promedio” directamente proporcional al monto. Lo anterior indica que los consumidores no varían demasiado los montos de sus compras con el transcurso del tiempo pero que les afecta negativamente el valor de la última compra.

Tabla 24: Signos variables explicativas modelo Monto.

Variables	Plaza Norte		Plaza Puente		La Dehesa		Manquehue		Estación Central	
	Antes	Después	Antes	Después	Antes	Después	Antes	Después	Antes	Después
Monto Última Compra	-	-	-	-	-	-	-	-	-	-
Monto Promedio	+	+	+	+	+	+	+	+	+	+
Mall	+	-	+	-	-	+	-	+	+	+
Meses Sin Compras	+	+	-	+	-	+	+	+	+	+

Considerando los signos obtenidos para la variable “Meses Sin Compras”, se aprecia que en todas las tiendas tiene un efecto positivo sobre la utilidad del consumidor salvo antes de la entrada Plaza Puente y La Dehesa, en donde a pesar que aumentarían su frecuencia de compra no disminuyeron su monto. Los signos positivos de esta variable tiene mucho sentido, ya que al tener menor separación de tiempo entre compras y asumiendo que el presupuesto es limitado, el monto disminuye en la compra inmediatamente posteriores (Boatwright, Borle, Kadane, Nunes y Shmueli, 2005).

Al revisar la variables “Mall” y tomando en cuenta el periodo posterior a la inauguración de las nuevas tiendas, se tiene que todas ellas cambiaron su signo salvo Estación Central que para ambas etapas fue positivo, es decir para los ingresos de Plaza Norte y Plaza Puente el formato mall fue percibido por los consumidores como un efecto negativo, situación contraria ocurre para las aperturas de La Dehesa y Manquehue.

Además, para confirmar el poder predictivo que tiene el modelo de monto desarrollado, se confeccionan comparativos entre los resultados obtenidos y la muestra poblacional real, en los ámbitos de “Monto Promedio” y “Monto Total” para cada una de las tiendas (ver Tablas N°25 y Tabla N° 26 respectivamente).

Para la calificación de “Monto Promedio”, se tiene que en promedio el modelo sobreestima los montos en un 15%, teniendo como valor mínimo -10% antes de la apertura de Plaza Norte y valor máximo 39% posteriormente a la apertura de Estación Central. Profundizando el análisis, se tiene que para todos los casos, salvo Plaza Norte, el monto promedio de compra aumenta con el ingreso de una nueva tienda.

Tabla 25: Comparativo resultados v/s base de datos (monto promedio).

Montos Promedio	Plaza Norte		Plaza Puente		La Dehesa		Manquehue		Estación Central	
	Antes	Después	Antes	Después	Antes	Después	Antes	Después	Antes	Después
Base de Datos	49.782	45.731	46.196	50.042	46.720	49.258	54.664	56.409	56.409	65.377
Modelo	44.927	42.658	54.236	58.530	55.048	60.191	60.152	70.057	68.900	90.845
Diferencia (%)	-10%	-7%	17%	17%	18%	22%	10%	24%	22%	39%

Ahora bien, considerando el “Monto Total” de cada una de las tiendas, en promedio el modelo sobreestima en un 12% los montos con respecto a la muestra real, teniendo como valor mínimo 2% antes del ingreso de la tienda Manquehue y valor máximo después de la apertura de Estación Central con un 39%. Además, se tiene que para las tiendas Plaza Norte, Plaza Puente y La Dehesa hay una disminución de los montos totales después de su lanzamiento (Neslin y Pauwels, 2006), situación adversa ocurrió para Manquehue y estación Central.

Tabla 26: Comparativo resultados v/s base de datos (monto total).

Montos Total	Plaza Norte		Plaza Puente		La Dehesa		Manquehue		Estación Central	
	Antes	Después	Antes	Después	Antes	Después	Antes	Después	Antes	Después
Base de Datos	70.939.887	67.497.596	68.925.071	68.448.128	70.360.883	69.488.128	73.249.494	77.111.407	77.111.407	96.294.256
Modelo	64.021.489	62.920.939	75.008.976	81.297.880	76.736.226	83.725.782	74.587.865	95.767.676	85.574.146	133.815.110
Diferencia (%)	-10%	-7%	9%	19%	9%	20%	2%	24%	11%	39%

10) Conclusiones Generales

Luego de haber finalizado el desarrollo, ejecución y análisis de resultados de los modelos Incidencia de Compra (Mixed Logit) y Monto de Compra (Regresión Lineal) considerando una muestra de 500 clientes activos de Falabella, se obtienen las siguientes conclusiones generales:

1. La cuantificación del efecto de canibalización para las tiendas inauguradas son: 3.8% Plaza Norte, 0.8% Plaza Puente, 1.0% La Dehesa, 0% Manquehue y 1.4% Estación Central. Los resultados confirman la hipótesis que los valores de canibalización producidos al ingresar una nueva tienda de la misma cadena tienen valores pequeños, en este caso no superan el 4%. A su vez, se identifica que las tiendas que fueron más afectadas con cada ingreso tenían la menor distancia euclidiana entre ellas.

2. Se tiene que la canibalización varía con el tiempo, obteniendo mayores o menores valores para este fenómeno dependiendo del momento específico de la evaluación (conjunto de elección disponible).

3. Considerando los factores influyentes en la incidencia de compra de un consumidor, se tiene que tanto el formato de la tienda (mall o calle) como la distancia entre el individuo y la tienda son las principales características que afectan la utilidad que le produce cierta tienda al cliente. Interpretando este resultado, el consumidor prefiere la tienda que esté más cerca de su ubicación y que no tenga un formato mall, lo cual se explica por el valor agregado que entrega una experiencia de compra agradable donde no se incurra en gastos de traslado y no haya aglomeraciones de gente (validado también por la variable Participación Poblacional). En resumen, los clientes tienen una mayor probabilidad de asistir a una tienda que tenga formato calle, ingresada después del año 2000, y que se encuentre ubicada a una menor distancia del consumidor.

4. Un resultado relevante para esta investigación está asociada a que los consumidores aumentan su probabilidad de asistencia a una tienda cuando está se encuentra cerca de otras de la misma cadena, lo cual confirma que la nueva estrategia de localización de tiendas de forma agregada beneficia a la cadena. A su vez, ocurre una situación contraria en tiendas ubicadas en lugares con mucha competencia debido a la cercanía de cadenas competidoras (mall).

5. Con respecto a las preferencias que tiene los consumidores, se tiene que los clientes demuestran mayor atractivo hacia las tiendas nuevas de Falabella, lo cual tiene sentido por las diversas iniciativas que estas tiendas realizan para posicionarse. A su vez, el ingreso de nuevas tiendas ayuda a mejorar el nivel de agrado de las tiendas ya establecidas con anterioridad a la inauguración.

6. Tomando en cuenta la variación en los montos para las entradas de las nuevas tiendas de Falabella, se tiene que en todos los casos, salvo Plaza Norte, aumenta el promedio del monto de compra de los consumidores. Esto demuestra que la apertura de nuevos locales beneficia los ingresos de la cadena completa.

7. La principal variable representativa del modelo de monto es “Monto Promedio”, lo cual significa que los clientes no realizan grandes variaciones en los montos de compra a lo largo del tiempo.

8. El modelo de incidencia de compra (Mixed Logit) y modelo de monto de compra son ejecutados con éxito en la obtención de los parámetros al aplicarlos con un enfoque bayesiano, obteniendo resultados individuales para cada cliente que permitan realizar análisis y conclusiones detallados.

9. Con la información transaccional incluida en las boletas entregadas a los clientes se pueden tomar decisiones estratégicas para la cadena, en este caso la localización de una nueva tienda.

11) Simulación tienda Costanera Center v/s Parque Padre Hurtado

Una de las expectativas que se tiene para este trabajo de investigación es conformar una metodología que se pueda aplicar a la vida real al momento de tener dos o más alternativas de locaciones para una nueva tienda. Para demostrar la aplicabilidad de los resultados se realiza una simulación de dos alternativas de posibles lugares para una nueva tienda de Falabella:

- Mall Costanera Center, ubicado en la comuna de Providencia con coordenadas X= -70.608 e Y= -33.417. Opción denominada como "CC".
- Parque Padre Hurtado, ubicado en la esquina Jorge de Alessandri con José Arrieta de la comuna de La Reina, en donde sus coordenadas son X= -70.547 e Y= -33.463. Se asume que al colocarse una tienda ahí es en formato mall. Alternativa abreviada como "LR".

Para simplificar el escenario, se realizan los cálculos considerando la decisión de localización en Enero del año 2009, con las 12 tiendas de Falabella ya inauguradas y para un solo consumidor denominado "A", en donde su ubicación cartesiana es X= -70.78 e Y= -33.56. Ahora bien, utilizando la ecuación 34 y considerando los parámetros promedio obtenidos para el modelo de incidencia de compra, la utilidad que le produce al cliente "A" cada alternativa está definida de la siguiente forma:

$$\begin{aligned}
 \text{Uti. Inc. Compra}_{A,CC,2009/1} = & (-1.76 - 0.63 - 5.37 - 3.91 - 1.87 - 0.48 - 0.09 - 3.43 - \\
 & 2.51 - 3.86 - 2.87 - 1.20) - 2.9 * \text{Distancia}_{A,CC} - 0.42 * \text{Distancia}_{A,CC}^2 - 4.46 * \text{Mall}_{CC} - \\
 & 0.4 * \text{Participación Población}_{CC,2009/1} - 1.00 * \text{Distancia Tienda Nueva}_{CC} - 1.55 * \\
 & \text{Antigüedad Tienda}_{CC} \quad (37)
 \end{aligned}$$

$$\begin{aligned}
 \text{Uti. Inc. Compra}_{A,LR,2009/1} = & (-1.76 - 0.63 - 5.37 - 3.91 - 1.87 - 0.48 - 0.09 - 3.43 - \\
 & 2.51 - 3.86 - 2.87 - 1.20) - 2.9 * \text{Distancia}_{A,LR} - 0.42 * \text{Distancia}_{A,LR}^2 - 4.46 * \text{Mall}_{LR} - \\
 & 0.4 * \text{Participación Población}_{LR,2009/1} - 1.00 * \text{Distancia Tienda Nueva}_{LR} - \\
 & 1.55 * \text{Antigüedad Tienda}_{LR} \quad (38)
 \end{aligned}$$

Para continuar con el ejercicio, en las ecuaciones 37 y 38 se coloca la información relacionada con las variables descriptivas para cada alternativa (ver

tabla N° 27), obteniendo la utilidad que le proporciona cada una de ellas al consumidor.

Tabla 27: Datos variables para simulación Costanera Center v/s La Reina.

Variable	Costanera Center	La Reina
Distancia	0,224	0,252
Distancia2	0,05	0,064
Mall	1	1
Participación	0,019	0,014
Dist. Nuev. Tien.	Ver Tabla	Ver Tabla
Tienda Antigua	0	0

Distancia Nueva Tienda	Costanera Center	La Reina
Alto Las Condes	0,068	0,074
Parque Arauco	0,033	0,069
Lyon	0,005	0,076
Santiago Centro	0,05	0,105
Plaza Tobalaba	0,162	0,107
Plaza Vespucio	0,102	0,077
Plaza Oeste	0,147	0,178
Plaza Norte	0,082	0,159
Plaza Puente	0,047	0,107
La Dehesa	0,109	0,109
Manquehue	0,042	0,055
Estación Central	0,082	0,136

$$Uti. Inc. Compra_{A,CC,2009/1} = -26.840 - 0.650 - 0.02 - 4.460 - 0.008 - 0.929 - 0 = -32.91. \quad (39)$$

$$Uti. Inc. Compra_{A,LR,2009/1} = -26.840 - 0.731 - 0.03 - 4.460 - 0.006 - 1.253 - 0 = -33.32. \quad (40)$$

Los resultados obtenidos después de resolver las ecuaciones de utilidad son: -32.91 para la ubicación Costanera Center y -33.32 para la ubicación Parque Padre Hurtado. Lo anterior significa que la ubicación Costanera Center le produce mayor utilidad al cliente y por ende, es a la tienda que asiste.

Teniendo elegido el lugar en donde va el consumidor "A", se debe estimar el monto de compra para la visita a la tienda. Para ello, se desarrolla la ecuación 36 con los parámetros promedio obtenidos para el modelo de monto:

$$\begin{aligned}
 & \mathbf{Monto}_{A,CC,2009/1} = \\
 & -0.18 * \mathbf{Monto\ Ultima\ Compra}_{t-1} + 1.24 * \mathbf{Monto\ Promedio}_{t-1} - 0.0001 * \\
 & \mathbf{Mall} + 0.003 * \mathbf{N^\circ\ Periodos\ Ultima\ Compra}_t \qquad \qquad \qquad (41)
 \end{aligned}$$

Tabla 28: Datos variables modelo monto para Costanera Center.

Variable	Valor
Monto Última Compra	\$ 60.490
Monto Promedio	\$ 52.562
Mall	1
N° Periodos Última Compra	9

Ahora que se tiene definida la función de monto en la ecuación 41, en donde se procede a completar los datos de acuerdo a la información transaccional del cliente señalado en la Tabla N°28, quedando de la siguiente forma:

$$\mathbf{Monto}_{A,CC,2009/1} = -10888 + 65177 - 0.0001 + 0.027 = \$54.289. \qquad \qquad (42)$$

En resumen, en resultado de la simulación considerando como fecha Enero 2009, el consumidor “A” asiste a la tienda ubicada en el Costanera Center (providencia) y realiza una compra por un monto de \$54.289.

12) Bibliografía

- Asim Ansari, Carl F. Mela, and Csott A. Neslin (2008), "Customer Channel Migration", *Journal of Marketing Research* XLV, 60-76.
- Avery, Jill, Thomas J. Steenburgh, John Deighton, and Mary Caravella (2008), "The Effects of Store Openings on Sales Through Direct Channels".
- Bell, Davis R., Teck-Hua Ho, y Christopher S. Tang (1998), "Determining Where to Shop: Fixed and Variable Costs of Shopping." *Journal of Marketing Research* 35 (3), 352-69.
- Bell y Sangyoung Song (2007), "Neighborhood Effects and Trial in the Internet: Evidence from Online Grocery Retailing." *Quantitative Marketing and Economics* 5 (4), 361-400.
- Biyalogorsky, E y P. Neik. 2003. Clicks and mortar: The effect of on-line activities on offline sales. *Marketing Letters*. 14 (1), 21-32.
- Bronnenberg, Bart, y Catarina Sismeiro (2002), "Using Multimarket Data to Predict Brand Performance in Market for Which No or Poor Data Exist." *Journal of Marketing Research* 39 (1), 1-17.
- Bronnenberg y Carl Mela (2004), "Market Roll-Out and Retailer Adoption for New Brands." *Marketing Science* 23 (4), 500-18.
- Boatwright, Peter, Shara Borle, Joseph B. Kadane, Joseph C. Nunes, and Galit Shmueli (2005), "The Effect of Product Assortment Changes on Customer Retention", *Marketing Science* 24 (4): 616-622.
- Datta, Sumon, K. Sudhir, y Debu Talukdar (2008), "A Structural Model of Entry and Location Choice: The Differentiation-Agglomeration Tradeoff." New Haven, Conn.: Yale School of Management, Working Paper.
- Deleersnyder, B., I. Geyskens, K. Gielens y M.G. Dekimpe. 2002. How Cannibalistic is the Internet Channel: A study of the newspaper industry in the

United Kingdom and the Netherlands. *International Journal of research in Marketing*. 19, 337-348.

- Fotheringham AS (1991) Statistical modeling of spatial choice: An overview. *Res. Marketing* 5 (Supplement 5): 95-117.

- Gupta, Sunil 2008, "Impact of Sales Promotions on When, What, and How Much to Buy", *Journal of Marketing Research* 25 (4), 342-355.

- Hanssens, D.M., L.J. Parsons y R.L. Schults. 2001. *Market response models: Econometric and time series analysis (2nd Edition ed.)*. Norwell, MA: Kluwer Academic Publishers.

- Huang, Ho-Chuan (2010), "Bayesian analysis of the dividend behaviour", *Applied Financial Economics* 11, 333-339.

- Huff D (1964) Defining and estimating a Trading Area. *J. Marketing* 28 (July): 34-38.

- Huff DL, Batsell RR (1977) Delimiting the areal extent of a market area. *J Marketing Res.* 14 (4): 581-585.

- Kumar, V. y R. Venkatesan. 2005. Who are the multichannel shoppers and how do they perform?: Correlates of multichannel shopping behavior. *Journal of Interactive Marketing*. 19 (2), 44-62.

- Li, Kai (1998), "Bayesian inference in a simultaneous equation model with limited dependent variables", *Journal of Econometrics* 85, 387-400.

- Neslin, Scott A., and Pauwels Koen (2006), "Building with Bricks and Mortar: The Revenue Impact of Opening Physical Stores in a Multichannel Environment".

- Neslin, Scott A., Caroline Herderson, y John Quelch (1985), "Consumer Promotions and the Acceleration of Product Purchases." *Marketing Science*, 4 (Spring), 147-65.

- Pancras, Joseph, S.Sriram, and V. Kumar (2009), "Investigating the Effect of New Store Opening in a Dynamic Environment", *Marketing Science Institute MSI Reports*, No 09-003, 77-100.
- Pancras, Joseph, S.Sriram, and V. Kumar (2012), "Empirical Investigation of Retail Expansion and Cannibalization in a Dynamic Environment.", *Management Science*, 25-1909, 1-18.
- Reilly W (1931) *The Law of retail Gravitation* (Putnam, New York).
- Rust RT, Donthu N (1995) Capturing geographically localized misspecification error in retail choice models. *J. Marketing Res.* 32 (1): 103-110.
- William H. Greene, 2002, *Análisis Económico*, Tercera Edición, New York University.
- Yang, Sha, y Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences." *Journal of Marketing Research* 40 (3), 282-94.

13) Anexos

Anexo N° 1: Código Matlab mixed logit y MCMC (incidencia de compra).

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%% MCMC Mixed Logit (ML) Model
%% Full heterogeneity
%% Estimation
%% 11/20/08
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%clear;
close all;
warning off;

rand('seed',0);
randn('seed',0);
time0=clock;
disp('*****');
ni=300;          % number of people          (n° clientes)
nj=64;          % number of attributes (or attribute levels) (n° variables // betas)
na=13;          % number of alternatives      (n° alternativas)
nt=160;         % number of choice tasks     (n° quincenas)
%nlevels=[4 4 4 4];
%np=sum(nlevels)-nj; %numCber of part-worths
np=np-nj;
fprintf('Mixed Logit model\n');
data=load('Datos.txt','-ascii'); %debe tener "puntos" y no "comas" / No poner
titulos.
%change individuals id
%data=transform_data(datos,ni,na,nt,nlevels,np);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%% METROPOLIS -HASTINGS ALGORITHM          %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

niter=150000;
nout=niter/2; %burn-in period
nchains=1;
```

```

acc=zeros(nchains,ni);
ntheta = np; % intercepts -1 can be estimated
%MasterPar_theta=double(zeros(ntheta,ni,(niter-nout))); %to allocate space
for r=1:nchains
LLN=-Inf;
kk=1;
l=0;
k=1;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Parameters initialization
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Initialize theta and mean theta
theta_old = .05*randn(ntheta,ni); %heterogeneity in the conditional probabilities
mu_theta = .5*randn(ntheta,1);
%priors for theta only
f0 = (ntheta)+5; %f0 = (ntheta)+1; before
G0 = f0*eye(ntheta);
un0=zeros(ntheta,1);
Vn0=10^2*eye(ntheta); %Vn0=10^(3)*eye(ntheta); before
sigmabar = randn(ntheta,ntheta);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Adaptation process parameters
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

gamma_n_theta= repmat(eye(ntheta),ni,1);
tau_bar=0.3;
epsilon1=10^(-7);
epsilon2=10^(-6);
A1=10^7;
sigma_n_theta=0.3;
mu_n_theta=.5*randn(ntheta,ni);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%covariance matrices
MLE_matrix_theta=eye(ntheta);
%compute LLN only the first time
LLNi=likeML(theta_old,data); %old loglikelihood
LLN=sum(LLNi);
LL0i=LLNi;

```

```

%%%%%%%%%%
%%%%%%%%%%
% START ITERATIONS
%%%%%%%%%%
%%%%%%%%%%

cc=-1;
acc_theta=zeros(ni,1);
Vn0inv=inv(Vn0);
while k<=niter
    accept_theta(r)=0;
    g_n=10/k; %for the adaptation process

%%%%%%%%%%
%%%%%%%%%%
    %UPDATE THETA FROM FULL CONDITIONALS

%%%%%%%%%%
%%%%%%%%%%

    SIGMA=(theta_old-mu_theta*ones(1,ni))*(theta_old-mu_theta*ones(1,ni))';
    invtest = inv(SIGMA + G0);
    invtest=triu(invtest)+triu(invtest)'-diag(diag(invtest));
    sigmainv=wishrnd(invtest,ni+f0); %sigma_theta = inv(sigmainv);

%%%%%%%%%%
%%%%%%%%%%

    %update mu
    Vn=inv(Vn0inv+ ni*sigmainv);
    un=Vn*(Vn0inv*un0 + sigmainv*(ni*mean(theta_old,2)));
    % full conditional mu_theta
    tmp=chol(Vn);
    mu_theta=un+tmp*randn(ntheta,1);
    lambda_n=gamma_n_theta+epsilon2*repmat(eye(ntheta),ni,1);
    if k<=5000
        lambda_new_theta=MLE_matrix_theta;
    else
        lambda_new_theta=(reshape(lambda_n',ntheta,ntheta,ni));
    end
    %heterogeneity in M
    theta_candidate=theta_old +
(sigma_n_theta*mvnrnd(zeros(ni,ntheta),lambda_new_theta))'; %using estimated
sigma
    LL0=LLN; %assign LL0 to the last accepted value of loglikelihood

```

```

LL0i=LLNi;
LL1i=likeML(theta_candidate,data); %candidate vector
LL1=sum(LL1i); %candidate loglikelihood
a1=-(1/2)*(theta_candidate-mu_theta*ones(1,ni))*sigmainv*(theta_candidate-
mu_theta*ones(1,ni));
a1=diag(a1)';
a0=-(1/2)*(theta_old-mu_theta*ones(1,ni))*sigmainv*(theta_old-
mu_theta*ones(1,ni));
a0=diag(a0)';
%compute pr(acceptance)
pr_theta=exp(a1-a0 + LL1i-LL0i);
pr2_theta=isnan(pr_theta);
pr3_theta=isinf(pr_theta);
%generate u ~U[0,1]
u=rand(ni,1);
pu_theta=(u<pr_theta)';
accept_theta(r)=sum(pu_theta(:));
pu_theta=diag(pu_theta);
if (isnan(LL1)==1)||isinf(LL1)==1
    theta_new=theta_old;
    LLNi=LL0i;
    pr_theta=0;
else
    theta_new=theta_candidate*pu_theta + theta_old*(eye(ni)-pu_theta);
    LLNi=LL1i*pu_theta +LL0i*(eye(ni)-pu_theta);
    acc_theta=acc_theta+diag(pu_theta);
end
LLN=sum(LLNi);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ADAPTIVE METHOD - Updating scale matrix
%gamma_n_2=gamma_n;

if k>3000
    %gamma_n
    tmp_gamma_n_theta=(zeros(ntheta*ni,ntheta));
    jp=0;
    for i=1:ni
        tmp_gamma_n_theta(jp+1:i*ntheta,:)=tmp_gamma_n_theta(jp+1:i*ntheta,:)
+ (theta_new(:,i)-mu_n_theta(:,i))*(theta_new(:,i)-mu_n_theta(:,i))';
        jp=i*ntheta;
    end
    gamma_n_theta=gamma_n_theta+g_n*(tmp_gamma_n_theta-

```

```

gamma_n_theta);

    dd_x= norm(gamma_n_theta,'fro');
    if dd_x>A1
        gamma_n_theta=(A1/dd_x)*gamma_n_theta;
    end
end
%mu_n
mu_n_theta=mu_n_theta+g_n*(theta_new-mu_n_theta);
for i=1:ni
    dd_x(i,1)= norm(mu_n_theta(:,i),'fro');
    if dd_x(i,1)>A1
        mu_n_theta(:,i)=(A1/dd_x(i,1))*mu_n_theta(:,i);
    end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Updating step
%sigma_n_phi and sigma_n_theta
if k>100
    sigma_n_theta = abs(sigma_n_theta + g_n*diag(min(pr_theta,1)-tau_bar));
    sigma_n_theta=diag(diag(sigma_n_theta));
end
%save values of phi and theta
if (k>nout)&&(mod(k,20)==0)%(before I was saving one every 20th)
    l=l+1;
    MasterPar_theta(:,l)=theta_new; %all the parameters
    param_mu_theta(r,l,:)=mu_theta'; %only the mean
    s=1;
    for k1=1:ntheta
        for k2=1:ntheta
            param_sigma_theta(r,l,s)=sigmainv(k1,k2);
            s=s+1;
        end
    end
    paramN(r,l,1)=(LLN);
    %paramN(r,l,2)=single(mean(acc(r,:))/k);
    LNi=exp(LLNi);
    lsi(r,l)=sum(log(1./LNi));
end

%for next iteration
if mod(k,1000)==0 %(it was 100)
    fprintf('-----\n' ,r,k,LLN);

```

```

    fprintf('Chain %2.0f Iteration:%2.0f LogLikelihood:%2.2f\n' ,r,k,LLN);
    fprintf('theta:[%2.2f %2.2f %2.2f]\n' ,mu_theta);
    fprintf('sigma_n_theta : %2.3f\n',trace(sigma_n_theta)/ni);
    fprintf('number of NAN : %2.0f INF: %2.0f mpr:
%2.2f\n',sum(pr2_theta),sum(pr3_theta),mean(min(pr_theta,1)));
    fprintf('cumulative time:%2.0f min\n',etime(clock,time0)/60);
end
k=k+1;
theta_old=theta_new;
end %end while
end %end for nchains
time=etime(clock,time0);
%MLL - marginal log-likelihood
nn1=1;
nn=size(paramN,2);
log_S=-paramN(nn1:nn);
cmax=max(log_S);
LML=-(log(mean(exp(log_S-cmax)))+cmax);
MP_theta=mean(MasterPar_theta(:, :, nn1:nn),3);
%PARAMETERS
%95%CI
for j=1:ntheta
    CI_MROL_theta(j,1)=quantile(param_mu_theta(1,nn1:nn,j),0.025);
    CI_MROL_theta(j,2)=quantile(param_mu_theta(1,nn1:nn,j),0.5);
    CI_MROL_theta(j,3)=quantile(param_mu_theta(1,nn1:nn,j),0.975);
end
save resultados

```

Anexo N° 2: Ejemplo base de datos para modelo de incidencia de compra.

Cliente	Mes Compra	Tienda	Compra	Eleccion	Amount	1	2	3	4	5	6	7	PH/Año
1	200401	10001	13	0	0	1	0	0	0	0	0	0	1
1	200401	10002	13	0	0	0	1	0	0	0	0	0	0
1	200401	10003	13	0	0	0	0	1	0	0	0	0	0
1	200401	10005	13	0	0	0	0	0	1	0	0	0	0
1	200401	10006	13	0	0	0	0	0	0	1	0	0	0
1	200401	10007	13	0	0	0	0	0	0	0	1	0	0
1	200401	10004	13	0	0	0	0	0	0	0	0	1	0
1	200401	10007	13	0	0	0	0	0	0	0	0	0	0
1	200401	10005	13	0	0	0	0	0	0	0	0	0	0
1	200401	10005	13	0	0	0	0	0	0	0	0	0	0
1	200401	10007	13	0	0	0	0	0	0	0	0	0	0
1	200401	10007	13	0	0	0	0	0	0	0	0	0	0
1	200401	10000	13	1	0	0	0	0	0	0	0	0	0
1	200402	10001	13	0	0	1	0	0	0	0	0	0	1
1	200402	10002	13	0	0	0	1	0	0	0	0	0	0
1	200402	10003	13	0	0	0	0	1	0	0	0	0	0
1	200402	10005	13	0	0	0	0	0	1	0	0	0	0
1	200402	10006	13	0	0	0	0	0	0	1	0	0	0
1	200402	10008	13	0	0	0	0	0	0	0	1	0	0
1	200402	10006	13	0	0	0	0	0	0	0	0	1	0
1	200402	10004	13	0	0	0	0	0	0	0	0	0	0
1	200402	10007	13	0	0	0	0	0	0	0	0	0	0
1	200402	10005	13	0	0	0	0	0	0	0	0	0	0
1	200402	10006	13	0	0	0	0	0	0	0	0	0	0
1	200402	10007	13	0	0	0	0	0	0	0	0	0	0
1	200402	10000	13	1	0	0	0	0	0	0	0	0	0

Fuente: Elaboración propia.

Anexo N° 3: Código macro “Armar Base de Datos”.

```
Sub Tabla_Final_Todas_Tiendas()
'DATO A INGRESAR
filas_totales = 308
tiendas = 13                '7 tiendas antiguas + "NO COMPRAR" + "PN".
meses = 146
'FIN DATO A INGRESAR
contar (filas_totales)      'Cuenta clientes y n° de transacciones.
Cambio_clientes (filas_totales) 'Enumera correlativamente a los clientes.
ni = Hoja2.Cells(2, 74)
Meses_Formato_Antiguedad (meses) ' Completa las columnas de meses, tiendas,
formato de tienda, antiguedad y densidad depoblación.
Densidad_Poblacion (meses)
Distancia (meses)          'Calcula la distancia de los clientes a cada alternativa.
Distancia_Nueva_Tienda (meses)
Eleccion (meses)          'Identifica la tienda que el individuo decidió comprar.
Codigo (meses)
Identidad (meses)         'Crea la matriz de identidad para tener interceptos
para cada tienda.
Identidad_Entrada_PN (meses)      'Intercepto que captura el efecto de la entrada
de Plaza Norte.
Identidad_Entrada_PP (meses)      'Intercepto que captura el efecto de la entrada
de Plaza Puente.
Identidad_Entrada_DE (meses)      'Intercepto que captura el efecto de la entrada
de La Dehesa.
Identidad_Entrada_MA (meses)      'Intercepto que captura el efecto de la entrada
de Manquehue.
Identidad_Entrada_EC (meses)      'Intercepto que captura el efecto de la entrada
de Estación Central.
End Sub
```

Anexo N° 4: Código macro “Contar”.

```
Sub contar(filas_totales) ' Cuenta las observaciones por cliente.
fila = 2 'contador de filas tabla original.
cliente_inicial = Hoja1.Cells(2, 1) 'primer cliente tabla original.
Hoja2.Cells(2, 75) = cliente_inicial
CLIENTES = 1
observaciones = 0
t = 1
While fila < filas_totales + 1
  If Hoja1.Cells(fila, 1) = cliente_inicial Then
    If filas_totales = fila Then
      Hoja2.Cells(t + 1, 76) = observaciones + 1
    End If
    fila = fila + 1
    observaciones = observaciones + 1
  Else
    Hoja2.Cells(t + 2, 75) = Hoja1.Cells(fila, 1)
    Hoja2.Cells(t + 1, 76) = observaciones
    observaciones = 1
    t = t + 1
    cliente_inicial = Hoja1.Cells(fila, 1)
    fila = fila + 1
    CLIENTES = CLIENTES + 1
    Hoja2.Cells(2, 74) = CLIENTES
    If filas_totales = fila - 1 Then
      Hoja2.Cells(t + 1, 76) = 1
    End If
  End If
End If
Wend
End Sub
```

Anexo N° 5: Código macro “Cambio_clientes”.

```
Sub Cambio_clientes(filas_totales)
CLIENTES = Hoja2.Cells(2, 74)
fila = 2
numero = 1
While fila < filas_totales + 1
  For nu_clientes = 2 To CLIENTES + 1
    observaciones = Hoja2.Cells(nu_clientes, 76)
  Next nu_clientes
  fila = fila + 1
End While
```

```

For i = 1 To observaciones
    Hoja1.Cells(fila, 1) = numero
    fila = fila + 1
Next i
numero = numero + 1
Next nu_clientes
Wend
End Sub

```

Anexo N° 6: Código macro “Meses_Formato_Antiguedad”.

```

Sub Meses_Formato_Antiguedad(meses)
tiendas = 13
ni = Hoja2.Cells(2, 74)
i = 2
tr = 2
r = 2          'fila de cada observación por cliente.
filas_nuevas = 2      'recorre las filas de la tabla nueva.
For CLIENTES = 1 To ni
    For k = 2 To meses + 1
        Hoja2.Cells(i, 3) = "ALTO LAS CONDES"
        Hoja2.Cells(i, 1) = CLIENTES          ' ID_CORREL
        Hoja2.Cells(i, 2) = Hoja1.Cells(k, 22)      ' Fecha
        Hoja2.Cells(i, 5) = 0
        Hoja2.Cells(i, 4) = tiendas
        Hoja2.Cells(i, 6) = 0
        Hoja2.Cells(i, 24) = 1                ' Formato tienda (1=Mall y 0=No)
        Hoja2.Cells(i, 28) = 1                ' Antiguedad tienda (1=Antes 2000 y
0=Despues 2000)
        Hoja2.Cells(i + 1, 3) = "PARQUE ARAUCO"
        Hoja2.Cells(i + 1, 1) = CLIENTES
        Hoja2.Cells(i + 1, 2) = Hoja1.Cells(k, 22)
        Hoja2.Cells(i + 1, 5) = 0
        Hoja2.Cells(i + 1, 4) = tiendas
        Hoja2.Cells(i + 1, 6) = 0
        Hoja2.Cells(i + 1, 24) = 1
        Hoja2.Cells(i + 1, 28) = 1            ' Antiguedad tienda (1=Antes
2000 y 0=Despues 2000)
        Hoja2.Cells(i + 2, 3) = "LYON"
        Hoja2.Cells(i + 2, 1) = CLIENTES
        Hoja2.Cells(i + 2, 2) = Hoja1.Cells(k, 22)
        Hoja2.Cells(i + 2, 5) = 0
        Hoja2.Cells(i + 2, 4) = tiendas

```

```

    Hoja2.Cells(i + 2, 6) = 0
    Hoja2.Cells(i + 2, 24) = 0
    Hoja2.Cells(i + 2, 28) = 1          ' Antigüedad tienda (1=Antes
2000 y 0=Despues 2000)
    Hoja2.Cells(i + 3, 3) = "CENTRO"
    Hoja2.Cells(i + 3, 1) = CLIENTES
    Hoja2.Cells(i + 3, 2) = Hoja1.Cells(k, 22)
    Hoja2.Cells(i + 3, 5) = 0
    Hoja2.Cells(i + 3, 4) = tiendas
    Hoja2.Cells(i + 3, 6) = 0
    Hoja2.Cells(i + 3, 24) = 0
    Hoja2.Cells(i + 3, 28) = 1          ' Antigüedad tienda (1=Antes
2000 y 0=Despues 2000)
    Hoja2.Cells(i + 4, 3) = "TOBALABA"
    Hoja2.Cells(i + 4, 1) = CLIENTES
    Hoja2.Cells(i + 4, 2) = Hoja1.Cells(k, 22)
    Hoja2.Cells(i + 4, 5) = 0
    Hoja2.Cells(i + 4, 4) = tiendas
    Hoja2.Cells(i + 4, 6) = 0
    Hoja2.Cells(i + 4, 24) = 1
    Hoja2.Cells(i + 4, 28) = 1          ' Antigüedad tienda (1=Antes
2000 y 0=Despues 2000)
    Hoja2.Cells(i + 5, 3) = "PLAZA VESPUCIO"
    Hoja2.Cells(i + 5, 1) = CLIENTES
    Hoja2.Cells(i + 5, 2) = Hoja1.Cells(k, 22)
    Hoja2.Cells(i + 5, 5) = 0
    Hoja2.Cells(i + 5, 4) = tiendas
    Hoja2.Cells(i + 5, 6) = 0
    Hoja2.Cells(i + 5, 24) = 1
    Hoja2.Cells(i + 5, 28) = 1          ' Antigüedad tienda (1=Antes
2000 y 0=Despues 2000)
    Hoja2.Cells(i + 6, 3) = "PLAZA OESTE"
    Hoja2.Cells(i + 6, 1) = CLIENTES
    Hoja2.Cells(i + 6, 2) = Hoja1.Cells(k, 22)
    Hoja2.Cells(i + 6, 5) = 0
    Hoja2.Cells(i + 6, 4) = tiendas
    Hoja2.Cells(i + 6, 6) = 0
    Hoja2.Cells(i + 6, 24) = 1
    Hoja2.Cells(i + 6, 28) = 1          ' Antigüedad tienda (1=Antes
2000 y 0=Despues 2000)
    If Hoja2.Cells(i, 2) < 2003111 Then
        Hoja2.Cells(i + 7, 3) = "PLAZA NORTE"
        Hoja2.Cells(i + 7, 1) = CLIENTES
        Hoja2.Cells(i + 7, 2) = Hoja1.Cells(k, 22)
        Hoja2.Cells(i + 7, 5) = 0
        Hoja2.Cells(i + 7, 4) = tiendas
        Hoja2.Cells(i + 7, 6) = 0
        Hoja2.Cells(i + 7, 24) = 0

```

```

        Hoja2.Cells(i + 7, 28) = 0           ' Antiguedad tienda (1=Antes
2000 y 0=Despues 2000)
    Else
        Hoja2.Cells(i + 7, 3) = "PLAZA NORTE"
        Hoja2.Cells(i + 7, 1) = CLIENTES
        Hoja2.Cells(i + 7, 2) = Hoja1.Cells(k, 22)
        Hoja2.Cells(i + 7, 5) = 0
        Hoja2.Cells(i + 7, 4) = tiendas
        Hoja2.Cells(i + 7, 6) = 0
        Hoja2.Cells(i + 7, 24) = 1         ' Mall
        Hoja2.Cells(i + 7, 28) = 0         ' Antiguedad tienda (1=Antes
2000 y 0=Despues 2000)
    End If
    If Hoja2.Cells(i, 2) < 2005101 Then
        Hoja2.Cells(i + 8, 3) = "PLAZA PUENTE"
        Hoja2.Cells(i + 8, 1) = CLIENTES
        Hoja2.Cells(i + 8, 2) = Hoja1.Cells(k, 22)
        Hoja2.Cells(i + 8, 5) = 0
        Hoja2.Cells(i + 8, 4) = tiendas
        Hoja2.Cells(i + 8, 6) = 0
        Hoja2.Cells(i + 8, 24) = 0
        Hoja2.Cells(i + 8, 28) = 0         ' Antiguedad tienda (1=Antes
2000 y 0=Despues 2000)
    Else
        Hoja2.Cells(i + 8, 3) = "PLAZA PUENTE"
        Hoja2.Cells(i + 8, 1) = CLIENTES
        Hoja2.Cells(i + 8, 2) = Hoja1.Cells(k, 22)
        Hoja2.Cells(i + 8, 5) = 0
        Hoja2.Cells(i + 8, 4) = tiendas
        Hoja2.Cells(i + 8, 6) = 0
        Hoja2.Cells(i + 8, 24) = 1         ' Mall
        Hoja2.Cells(i + 8, 28) = 0         ' Antiguedad tienda (1=Antes
2000 y 0=Despues 2000)
    End If
    If Hoja2.Cells(i, 2) < 2005111 Then
        Hoja2.Cells(i + 9, 3) = "LA DEHESA"
        Hoja2.Cells(i + 9, 1) = CLIENTES
        Hoja2.Cells(i + 9, 2) = Hoja1.Cells(k, 22)
        Hoja2.Cells(i + 9, 5) = 0
        Hoja2.Cells(i + 9, 4) = tiendas
        Hoja2.Cells(i + 9, 6) = 0
        Hoja2.Cells(i + 9, 24) = 0
        Hoja2.Cells(i + 9, 28) = 0         ' Antiguedad tienda (1=Antes
2000 y 0=Despues 2000)
    Else
        Hoja2.Cells(i + 9, 3) = "LA DEHESA"
        Hoja2.Cells(i + 9, 1) = CLIENTES
        Hoja2.Cells(i + 9, 2) = Hoja1.Cells(k, 22)

```

```

        Hoja2.Cells(i + 9, 5) = 0
        Hoja2.Cells(i + 9, 4) = tiendas
        Hoja2.Cells(i + 9, 6) = 0
        Hoja2.Cells(i + 9, 24) = 1           ' Mall
        Hoja2.Cells(i + 9, 28) = 0         ' Antigüedad tienda (1=Antes
2000 y 0=Despues 2000)
    End If
    If Hoja2.Cells(i, 2) < 2007051 Then
        Hoja2.Cells(i + 10, 3) = "MANQUEHUE"
        Hoja2.Cells(i + 10, 1) = CLIENTES
        Hoja2.Cells(i + 10, 2) = Hoja1.Cells(k, 22)
        Hoja2.Cells(i + 10, 5) = 0
        Hoja2.Cells(i + 10, 4) = tiendas
        Hoja2.Cells(i + 10, 6) = 0
        Hoja2.Cells(i + 10, 24) = 0
        Hoja2.Cells(i + 10, 28) = 0         ' Antigüedad tienda (1=Antes
2000 y 0=Despues 2000)
    Else
        Hoja2.Cells(i + 10, 3) = "MANQUEHUE"
        Hoja2.Cells(i + 10, 1) = CLIENTES
        Hoja2.Cells(i + 10, 2) = Hoja1.Cells(k, 22)
        Hoja2.Cells(i + 10, 5) = 0
        Hoja2.Cells(i + 10, 4) = tiendas
        Hoja2.Cells(i + 10, 6) = 0
        Hoja2.Cells(i + 10, 24) = 1         ' Mall
        Hoja2.Cells(i + 10, 28) = 0         ' Antigüedad tienda (1=Antes
2000 y 0=Despues 2000)
    End If
    If Hoja2.Cells(i, 2) < 2007051 Then
        Hoja2.Cells(i + 11, 3) = "ESTACION CENTRAL"
        Hoja2.Cells(i + 11, 1) = CLIENTES
        Hoja2.Cells(i + 11, 2) = Hoja1.Cells(k, 22)
        Hoja2.Cells(i + 11, 5) = 0
        Hoja2.Cells(i + 11, 4) = tiendas
        Hoja2.Cells(i + 11, 6) = 0
        Hoja2.Cells(i + 11, 24) = 0
        Hoja2.Cells(i + 11, 28) = 0         ' Antigüedad tienda (1=Antes
2000 y 0=Despues 2000)
    Else
        Hoja2.Cells(i + 11, 3) = "ESTACION CENTRAL"
        Hoja2.Cells(i + 11, 1) = CLIENTES
        Hoja2.Cells(i + 11, 2) = Hoja1.Cells(k, 22)
        Hoja2.Cells(i + 11, 5) = 0
        Hoja2.Cells(i + 11, 4) = tiendas
        Hoja2.Cells(i + 11, 6) = 0
        Hoja2.Cells(i + 11, 24) = 1         ' Mall
        Hoja2.Cells(i + 11, 28) = 0         ' Antigüedad tienda (1=Antes
2000 y 0=Despues 2000)

```

```

End If
Hoja2.Cells(i + 12, 3) = "NO COMPRO"
Hoja2.Cells(i + 12, 1) = CLIENTES
Hoja2.Cells(i + 12, 2) = Hoja1.Cells(k, 22)
Hoja2.Cells(i + 12, 5) = 1
Hoja2.Cells(i + 12, 4) = tiendas
Hoja2.Cells(i + 12, 6) = 0
Hoja2.Cells(i + 12, 24) = 0
Hoja2.Cells(i + 12, 28) = 0           ' Antigüedad tienda (1=Antes
2000 y 0=Despues 2000)
    i = i + tiendas
Next k
Next CLIENTES
End Sub

```

Anexo N° 7: Código macro “Densidad_ Población”.

```

Sub Densidad_Poblacion(meses)
tiendas = 13
ni = Hoja2.Cells(2, 74)
fila = 2
While fila < tiendas * ni * meses + 2
    If Hoja2.Cells(fila, 3) = "ALTO LAS CONDES" Then
        If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2002122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(2, 28)
        ElseIf 2003011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2003122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(2, 29)
        ElseIf 2004011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2004122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(2, 30)
        ElseIf 2005011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2005122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(2, 31)
        ElseIf 2006011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2006122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(2, 32)
        ElseIf 2007011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2007122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(2, 33)
        ElseIf 2008011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then

```

```

        Hoja2.Cells(fila, 25) = Hoja1.Cells(2, 34)
    Elself 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009062 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(2, 35)
    End If
    Elself Hoja2.Cells(fila, 3) = "PARQUE ARAUCO" Then
        If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2002122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(3, 28)
            Elself 2003011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2003122 Then
                Hoja2.Cells(fila, 25) = Hoja1.Cells(3, 29)
                Elself 2004011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2004122 Then
                    Hoja2.Cells(fila, 25) = Hoja1.Cells(3, 30)
                    Elself 2005011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2005122 Then
                        Hoja2.Cells(fila, 25) = Hoja1.Cells(3, 31)
                        Elself 2006011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2006122 Then
                            Hoja2.Cells(fila, 25) = Hoja1.Cells(3, 32)
                            Elself 2007011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2007122 Then
                                Hoja2.Cells(fila, 25) = Hoja1.Cells(3, 33)
                                Elself 2008011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then
                                    Hoja2.Cells(fila, 25) = Hoja1.Cells(3, 34)
                                    Elself 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009122 Then
                                        Hoja2.Cells(fila, 25) = Hoja1.Cells(3, 35)
                                    End If
                                Elself Hoja2.Cells(fila, 3) = "LYON" Then
                                    If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2002122 Then
                                        Hoja2.Cells(fila, 25) = Hoja1.Cells(4, 28)
                                        Elself 2003011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2003122 Then
                                            Hoja2.Cells(fila, 25) = Hoja1.Cells(4, 29)
                                            Elself 2004011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2004122 Then
                                                Hoja2.Cells(fila, 25) = Hoja1.Cells(4, 30)
                                                Elself 2005011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2005122 Then
                                                    Hoja2.Cells(fila, 25) = Hoja1.Cells(4, 31)
                                                    Elself 2006011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2006122 Then
                                                        Hoja2.Cells(fila, 25) = Hoja1.Cells(4, 32)
                                                        Elself 2007011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=

```

```

2007122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(4, 33)
    Elseif 2008011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(4, 34)
    Elseif 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(4, 35)
    End If
    Elseif Hoja2.Cells(fila, 3) = "CENTRO" Then
        If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2002122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(5, 28)
            Elseif 2003011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2003122 Then
                Hoja2.Cells(fila, 25) = Hoja1.Cells(5, 29)
                Elseif 2004011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2004122 Then
                    Hoja2.Cells(fila, 25) = Hoja1.Cells(5, 30)
                    Elseif 2005011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2005122 Then
                        Hoja2.Cells(fila, 25) = Hoja1.Cells(5, 31)
                        Elseif 2006011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2006122 Then
                            Hoja2.Cells(fila, 25) = Hoja1.Cells(5, 32)
                            Elseif 2007011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2007122 Then
                                Hoja2.Cells(fila, 25) = Hoja1.Cells(5, 33)
                                Elseif 2008011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then
                                    Hoja2.Cells(fila, 25) = Hoja1.Cells(5, 34)
                                    Elseif 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009122 Then
                                        Hoja2.Cells(fila, 25) = Hoja1.Cells(5, 35)
                                        End If
                                        Elseif Hoja2.Cells(fila, 3) = "TOBALABA" Then
                                            If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2002122 Then
                                                Hoja2.Cells(fila, 25) = Hoja1.Cells(6, 28)
                                                Elseif 2003011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2003122 Then
                                                    Hoja2.Cells(fila, 25) = Hoja1.Cells(6, 29)
                                                    Elseif 2004011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2004122 Then
                                                        Hoja2.Cells(fila, 25) = Hoja1.Cells(6, 30)
                                                        Elseif 2005011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2005122 Then
                                                            Hoja2.Cells(fila, 25) = Hoja1.Cells(6, 31)

```

```

Elseif 2006011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2006122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(6, 32)
Elseif 2007011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2007122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(6, 33)
Elseif 2008011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(6, 34)
Elseif 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(6, 35)
End If
Elseif Hoja2.Cells(fila, 3) = "PLAZA VESPUCIO" Then
    If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2002122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(7, 28)
Elseif 2003011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2003122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(7, 29)
Elseif 2004011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2004122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(7, 30)
Elseif 2005011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2005122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(7, 31)
Elseif 2006011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2006122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(7, 32)
Elseif 2007011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2007122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(7, 33)
Elseif 2008011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(7, 34)
Elseif 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(7, 35)
End If
Elseif Hoja2.Cells(fila, 3) = "PLAZA OESTE" Then
    If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2002122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(8, 28)
Elseif 2003011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2003122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(8, 29)
Elseif 2004011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2004122 Then

```

```

        Hoja2.Cells(fila, 25) = Hoja1.Cells(8, 30)
    ElseIf 2005011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2005122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(8, 31)
    ElseIf 2006011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2006122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(8, 32)
    ElseIf 2007011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2007122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(8, 33)
    ElseIf 2008011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(8, 34)
    ElseIf 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(8, 35)
    End If
    ElseIf Hoja2.Cells(fila, 3) = "PLAZA NORTE" Then
        If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2003102 Then
            Hoja2.Cells(fila, 25) = 0
        ElseIf 2003111 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2003122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(9, 29)
        ElseIf 2004011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2004122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(9, 30)
        ElseIf 2005011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2005122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(9, 31)
        ElseIf 2006011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2006122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(9, 32)
        ElseIf 2007011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2007122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(9, 33)
        ElseIf 2008011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(9, 34)
        ElseIf 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009122 Then
            Hoja2.Cells(fila, 25) = Hoja1.Cells(9, 35)
    End If
    ElseIf Hoja2.Cells(fila, 3) = "PLAZA PUENTE" Then
        If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2005092 Then
            Hoja2.Cells(fila, 25) = 0
        ElseIf 2005101 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=

```

```

2005122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(10, 31)
    Elseif 2006011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2006122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(10, 32)
    Elseif 2007011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2007122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(10, 33)
    Elseif 2008011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(10, 34)
    Elseif 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(10, 35)
    End If
    Elseif Hoja2.Cells(fila, 3) = "LA DEHESA" Then
        If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2005102 Then
    Hoja2.Cells(fila, 25) = 0
    Elseif 2005111 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2005122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(11, 31)
    Elseif 2006011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2006122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(11, 32)
    Elseif 2007011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2007122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(11, 33)
    Elseif 2008011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(11, 34)
    Elseif 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(11, 35)
    End If
    Elseif Hoja2.Cells(fila, 3) = "MANQUEHUE" Then
        If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2007042 Then
    Hoja2.Cells(fila, 25) = 0
    Elseif 2007051 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2007122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(12, 33)
    Elseif 2008011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(12, 34)
    Elseif 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009122 Then
    Hoja2.Cells(fila, 25) = Hoja1.Cells(12, 35)

```

```

End If
Elseif Hoja2.Cells(fila, 3) = "ESTACION CENTRAL" Then
  If 2002011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008052 Then
    Hoja2.Cells(fila, 25) = 0
    Elseif 2008061 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2008122 Then
      Hoja2.Cells(fila, 25) = Hoja1.Cells(13, 34)
      Elseif 2009011 <= Hoja2.Cells(fila, 2) And Hoja2.Cells(fila, 2) <=
2009122 Then
        Hoja2.Cells(fila, 25) = Hoja1.Cells(13, 35)
      End If
    Else
      Hoja2.Cells(fila, 25) = 0
    End If
  End If
  fila = fila + 1
Wend
End Sub

```

Anexo N° 8: Código macro "Distancia".

```

Sub Distancia(meses)
ni = Hoja2.Cells(2, 74) ' número de clientes en la base de datos
tiendas = 13          ' 8 tiendas más la alternativa de no comprar.
i = 2
tr = 2                ' coordenadas X e Y de cada cliente.
r = 2                 ' recorre tabla de numero de observaciones por cliente
For CLIENTES = 1 To ni
no = Hoja2.Cells(r, 76) ' numero de observaciones por cliente
  For k = 2 To meses + 1
    fecha = Hoja2.Cells(i, 2)
    'Hoja2.Cells(i, 3) = "ALTO LAS CONDES"
    Hoja2.Cells(i, 22) = Sqr((Hoja1.Cells(2, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(2, 16) - Hoja1.Cells(tr, 3)) ^ 2) ' Distancia euclidiana
    Hoja2.Cells(i, 23) = Hoja2.Cells(i, 22) ^ 2
    'Hoja2.Cells(i + 1, 3) = "PARQUE ARAUCO"
    Hoja2.Cells(i + 1, 22) = Sqr((Hoja1.Cells(3, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(3, 16) - Hoja1.Cells(tr, 3)) ^ 2)
    Hoja2.Cells(i + 1, 23) = Hoja2.Cells(i + 1, 22) ^ 2
    'Hoja2.Cells(i + 2, 3) = "LYON"
  
```

```

    Hoja2.Cells(i + 2, 22) = Sqr((Hoja1.Cells(4, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(4, 16) - Hoja1.Cells(tr, 3)) ^ 2)
    Hoja2.Cells(i + 2, 23) = Hoja2.Cells(i + 2, 22) ^ 2
    'Hoja2.Cells(i + 3, 3) = "CENTRO"
    Hoja2.Cells(i + 3, 22) = Sqr((Hoja1.Cells(5, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(5, 16) - Hoja1.Cells(tr, 3)) ^ 2)
    Hoja2.Cells(i + 3, 23) = Hoja2.Cells(i + 3, 22) ^ 2
    'Hoja2.Cells(i + 4, 3) = "TOBALABA"
    Hoja2.Cells(i + 4, 22) = Sqr((Hoja1.Cells(6, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(6, 16) - Hoja1.Cells(tr, 3)) ^ 2)
    Hoja2.Cells(i + 4, 23) = Hoja2.Cells(i + 4, 22) ^ 2
    'Hoja2.Cells(i + 5, 3) = "PLAZA VESPUCIO"
    Hoja2.Cells(i + 5, 22) = Sqr((Hoja1.Cells(7, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(7, 16) - Hoja1.Cells(tr, 3)) ^ 2)
    Hoja2.Cells(i + 5, 23) = Hoja2.Cells(i + 5, 22) ^ 2
    'Hoja2.Cells(i + 6, 3) = "PLAZA OESTE"
    Hoja2.Cells(i + 6, 22) = Sqr((Hoja1.Cells(8, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(8, 16) - Hoja1.Cells(tr, 3)) ^ 2)
    Hoja2.Cells(i + 6, 23) = Hoja2.Cells(i + 6, 22) ^ 2
    If fecha < 2003111 Then
        "PLAZA NORTE"
        Hoja2.Cells(i + 7, 22) = 0
        Hoja2.Cells(i + 7, 23) = 0
    Else
        "PLAZA NORTE"
        Hoja2.Cells(i + 7, 22) = Sqr((Hoja1.Cells(9, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(9, 16) - Hoja1.Cells(tr, 3)) ^ 2)
        Hoja2.Cells(i + 7, 23) = Hoja2.Cells(i + 7, 22) ^ 2
    End If
    If fecha < 2005101 Then
        "PLAZA PUENTE"
        Hoja2.Cells(i + 8, 22) = 0
        Hoja2.Cells(i + 8, 23) = 0
    Else
        "PLAZA PUENTE"
        Hoja2.Cells(i + 8, 22) = Sqr((Hoja1.Cells(10, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(10, 16) - Hoja1.Cells(tr, 3)) ^ 2)
        Hoja2.Cells(i + 8, 23) = Hoja2.Cells(i + 8, 22) ^ 2
    End If
    If fecha < 2005111 Then
        "LA DEHESA"
        Hoja2.Cells(i + 9, 22) = 0
        Hoja2.Cells(i + 9, 23) = 0
    Else

```

```

        "LA DEHESA"
        Hoja2.Cells(i + 9, 22) = Sqr((Hoja1.Cells(11, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(11, 16) - Hoja1.Cells(tr, 3)) ^ 2)
        Hoja2.Cells(i + 9, 23) = Hoja2.Cells(i + 9, 22) ^ 2
    End If
    If fecha < 2007051 Then
        "MANQUEHUE"
        Hoja2.Cells(i + 10, 22) = 0
        Hoja2.Cells(i + 10, 23) = 0
    Else "MANQUEHUE"
        Hoja2.Cells(i + 10, 22) = Sqr((Hoja1.Cells(12, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(12, 16) - Hoja1.Cells(tr, 3)) ^ 2)
        Hoja2.Cells(i + 10, 23) = Hoja2.Cells(i + 10, 22) ^ 2
    End If
    If fecha < 2008061 Then
        "ESTACION CENTRAL"
        Hoja2.Cells(i + 11, 22) = 0
        Hoja2.Cells(i + 11, 23) = 0
    Else
        "ESTACION CENTRAL"
        Hoja2.Cells(i + 11, 22) = Sqr((Hoja1.Cells(13, 15) - Hoja1.Cells(tr, 2)) ^ 2 +
(Hoja1.Cells(13, 16) - Hoja1.Cells(tr, 3)) ^ 2)
        Hoja2.Cells(i + 11, 23) = Hoja2.Cells(i + 11, 22) ^ 2
    End If
    "NO COMPRO"
    Hoja2.Cells(i + 12, 22) = 0
    Hoja2.Cells(i + 12, 23) = 0
    i = i + tiendas
Next k
tr = tr + no
r = r + 1
Next CLIENTES
End Sub

```

Anexo N° 9: Código macro “Distancia_Nueva_Tienda”.

```
Sub Distancia_Nueva_Tienda(meses)
ni = Hoja2.Cells(2, 74) ' número de clientes en la base de datos
tiendas = 13          ' 8 tiendas más la alternativa de no comprar.
i = 2
While i < ni * meses * tiendas + 1
  fecha = Hoja2.Cells(i, 2)
  If fecha < 2003111 Then
    Hoja2.Cells(i, 27) = 0
    Hoja2.Cells(i + 1, 27) = 0
    Hoja2.Cells(i + 2, 27) = 0
    Hoja2.Cells(i + 3, 27) = 0
    Hoja2.Cells(i + 4, 27) = 0
    Hoja2.Cells(i + 5, 27) = 0
    Hoja2.Cells(i + 6, 27) = 0
    Hoja2.Cells(i + 7, 27) = 0
    Hoja2.Cells(i + 8, 27) = 0
    Hoja2.Cells(i + 9, 27) = 0
    Hoja2.Cells(i + 10, 27) = 0
    Hoja2.Cells(i + 11, 27) = 0
    Hoja2.Cells(i + 12, 27) = 0
  ElseIf fecha < 2005101 Then 'PLAZA NORTE
    Hoja2.Cells(i, 27) = Sqr((Hoja1.Cells(2, 15) - Hoja1.Cells(9, 15)) ^ 2 +
(Hoja1.Cells(2, 16) - Hoja1.Cells(9, 16)) ^ 2)
    Hoja2.Cells(i + 1, 27) = Sqr((Hoja1.Cells(3, 15) - Hoja1.Cells(9, 15)) ^ 2
+ (Hoja1.Cells(3, 16) - Hoja1.Cells(9, 16)) ^ 2)
    Hoja2.Cells(i + 2, 27) = Sqr((Hoja1.Cells(4, 15) - Hoja1.Cells(9, 15)) ^ 2
+ (Hoja1.Cells(4, 16) - Hoja1.Cells(9, 16)) ^ 2)
    Hoja2.Cells(i + 3, 27) = Sqr((Hoja1.Cells(5, 15) - Hoja1.Cells(9, 15)) ^ 2
+ (Hoja1.Cells(5, 16) - Hoja1.Cells(9, 16)) ^ 2)
    Hoja2.Cells(i + 4, 27) = Sqr((Hoja1.Cells(6, 15) - Hoja1.Cells(9, 15)) ^ 2
+ (Hoja1.Cells(6, 16) - Hoja1.Cells(9, 16)) ^ 2)
    Hoja2.Cells(i + 5, 27) = Sqr((Hoja1.Cells(7, 15) - Hoja1.Cells(9, 15)) ^ 2
+ (Hoja1.Cells(7, 16) - Hoja1.Cells(9, 16)) ^ 2)
    Hoja2.Cells(i + 6, 27) = Sqr((Hoja1.Cells(8, 15) - Hoja1.Cells(9, 15)) ^ 2
+ (Hoja1.Cells(8, 16) - Hoja1.Cells(9, 16)) ^ 2)
    Hoja2.Cells(i + 7, 27) = 0          'PLAZA NORTE
    Hoja2.Cells(i + 8, 27) = 0
    Hoja2.Cells(i + 9, 27) = 0
    Hoja2.Cells(i + 10, 27) = 0
    Hoja2.Cells(i + 11, 27) = 0
    Hoja2.Cells(i + 12, 27) = 0
  ElseIf fecha < 2005111 Then 'PLAZA PUENTE
    Hoja2.Cells(i, 27) = Sqr((Hoja1.Cells(2, 15) - Hoja1.Cells(10, 15)) ^ 2 +
```

```

(Hoja1.Cells(2, 16) - Hoja1.Cells(10, 16)) ^ 2)
    Hoja2.Cells(i + 1, 27) = Sqr((Hoja1.Cells(3, 15) - Hoja1.Cells(10, 15)) ^ 2
+ (Hoja1.Cells(3, 16) - Hoja1.Cells(10, 16)) ^ 2)
    Hoja2.Cells(i + 2, 27) = Sqr((Hoja1.Cells(4, 15) - Hoja1.Cells(10, 15)) ^ 2
+ (Hoja1.Cells(4, 16) - Hoja1.Cells(10, 16)) ^ 2)
    Hoja2.Cells(i + 3, 27) = Sqr((Hoja1.Cells(5, 15) - Hoja1.Cells(10, 15)) ^ 2
+ (Hoja1.Cells(5, 16) - Hoja1.Cells(10, 16)) ^ 2)
    Hoja2.Cells(i + 4, 27) = Sqr((Hoja1.Cells(6, 15) - Hoja1.Cells(10, 15)) ^ 2
+ (Hoja1.Cells(6, 16) - Hoja1.Cells(10, 16)) ^ 2)
    Hoja2.Cells(i + 5, 27) = Sqr((Hoja1.Cells(7, 15) - Hoja1.Cells(10, 15)) ^ 2
+ (Hoja1.Cells(7, 16) - Hoja1.Cells(10, 16)) ^ 2)
    Hoja2.Cells(i + 6, 27) = Sqr((Hoja1.Cells(8, 15) - Hoja1.Cells(10, 15)) ^ 2
+ (Hoja1.Cells(8, 16) - Hoja1.Cells(10, 16)) ^ 2)
    Hoja2.Cells(i + 7, 27) = Sqr((Hoja1.Cells(9, 15) - Hoja1.Cells(10, 15)) ^ 2
+ (Hoja1.Cells(9, 16) - Hoja1.Cells(10, 16)) ^ 2)
    Hoja2.Cells(i + 8, 27) = 0 'PLAZA PUENTE
    Hoja2.Cells(i + 9, 27) = 0
    Hoja2.Cells(i + 10, 27) = 0
    Hoja2.Cells(i + 11, 27) = 0
    Hoja2.Cells(i + 12, 27) = 0
    Elself fecha < 2007051 Then 'LA DEHESA
    Hoja2.Cells(i, 27) = Sqr((Hoja1.Cells(2, 15) - Hoja1.Cells(11, 15)) ^ 2 +
(Hoja1.Cells(2, 16) - Hoja1.Cells(11, 16)) ^ 2)
    Hoja2.Cells(i + 1, 27) = Sqr((Hoja1.Cells(3, 15) - Hoja1.Cells(11, 15)) ^ 2
+ (Hoja1.Cells(3, 16) - Hoja1.Cells(11, 16)) ^ 2)
    Hoja2.Cells(i + 2, 27) = Sqr((Hoja1.Cells(4, 15) - Hoja1.Cells(11, 15)) ^ 2
+ (Hoja1.Cells(4, 16) - Hoja1.Cells(11, 16)) ^ 2)
    Hoja2.Cells(i + 3, 27) = Sqr((Hoja1.Cells(5, 15) - Hoja1.Cells(11, 15)) ^ 2
+ (Hoja1.Cells(5, 16) - Hoja1.Cells(11, 16)) ^ 2)
    Hoja2.Cells(i + 4, 27) = Sqr((Hoja1.Cells(6, 15) - Hoja1.Cells(11, 15)) ^ 2
+ (Hoja1.Cells(6, 16) - Hoja1.Cells(11, 16)) ^ 2)
    Hoja2.Cells(i + 5, 27) = Sqr((Hoja1.Cells(7, 15) - Hoja1.Cells(11, 15)) ^ 2
+ (Hoja1.Cells(7, 16) - Hoja1.Cells(11, 16)) ^ 2)
    Hoja2.Cells(i + 6, 27) = Sqr((Hoja1.Cells(8, 15) - Hoja1.Cells(11, 15)) ^ 2
+ (Hoja1.Cells(8, 16) - Hoja1.Cells(11, 16)) ^ 2)
    Hoja2.Cells(i + 7, 27) = Sqr((Hoja1.Cells(9, 15) - Hoja1.Cells(11, 15)) ^ 2
+ (Hoja1.Cells(9, 16) - Hoja1.Cells(11, 16)) ^ 2)
    Hoja2.Cells(i + 8, 27) = Sqr((Hoja1.Cells(10, 15) - Hoja1.Cells(11, 15)) ^
2 + (Hoja1.Cells(10, 16) - Hoja1.Cells(11, 16)) ^ 2)
    Hoja2.Cells(i + 9, 27) = 0 'LA DEHESA
    Hoja2.Cells(i + 10, 27) = 0
    Hoja2.Cells(i + 11, 27) = 0
    Hoja2.Cells(i + 12, 27) = 0
    Elself fecha < 2008061 Then 'MANQUEHUE
    Hoja2.Cells(i, 27) = Sqr((Hoja1.Cells(2, 15) - Hoja1.Cells(12, 15)) ^ 2 +
(Hoja1.Cells(2, 16) - Hoja1.Cells(12, 16)) ^ 2)
    Hoja2.Cells(i + 1, 27) = Sqr((Hoja1.Cells(3, 15) - Hoja1.Cells(12, 15)) ^ 2
+ (Hoja1.Cells(3, 16) - Hoja1.Cells(12, 16)) ^ 2)

```

```

        Hoja2.Cells(i + 2, 27) = Sqr((Hoja1.Cells(4, 15) - Hoja1.Cells(12, 15)) ^ 2
+ (Hoja1.Cells(4, 16) - Hoja1.Cells(12, 16)) ^ 2)
        Hoja2.Cells(i + 3, 27) = Sqr((Hoja1.Cells(5, 15) - Hoja1.Cells(12, 15)) ^ 2
+ (Hoja1.Cells(5, 16) - Hoja1.Cells(12, 16)) ^ 2)
        Hoja2.Cells(i + 4, 27) = Sqr((Hoja1.Cells(6, 15) - Hoja1.Cells(12, 15)) ^ 2
+ (Hoja1.Cells(6, 16) - Hoja1.Cells(12, 16)) ^ 2)
        Hoja2.Cells(i + 5, 27) = Sqr((Hoja1.Cells(7, 15) - Hoja1.Cells(12, 15)) ^ 2
+ (Hoja1.Cells(7, 16) - Hoja1.Cells(12, 16)) ^ 2)
        Hoja2.Cells(i + 6, 27) = Sqr((Hoja1.Cells(8, 15) - Hoja1.Cells(12, 15)) ^ 2
+ (Hoja1.Cells(8, 16) - Hoja1.Cells(12, 16)) ^ 2)
        Hoja2.Cells(i + 7, 27) = Sqr((Hoja1.Cells(9, 15) - Hoja1.Cells(12, 15)) ^ 2
+ (Hoja1.Cells(9, 16) - Hoja1.Cells(12, 16)) ^ 2)
        Hoja2.Cells(i + 8, 27) = Sqr((Hoja1.Cells(10, 15) - Hoja1.Cells(12, 15)) ^
2 + (Hoja1.Cells(10, 16) - Hoja1.Cells(12, 16)) ^ 2)
        Hoja2.Cells(i + 9, 27) = Sqr((Hoja1.Cells(11, 15) - Hoja1.Cells(12, 15)) ^
2 + (Hoja1.Cells(11, 16) - Hoja1.Cells(12, 16)) ^ 2)
        Hoja2.Cells(i + 10, 27) = 0 'MANQUEHUE
        Hoja2.Cells(i + 11, 27) = 0
        Hoja2.Cells(i + 12, 27) = 0
    Else
        'ESTACION CENTRAL
        Hoja2.Cells(i, 27) = Sqr((Hoja1.Cells(2, 15) - Hoja1.Cells(13, 15)) ^ 2 +
(Hoja1.Cells(2, 16) - Hoja1.Cells(13, 16)) ^ 2)
        Hoja2.Cells(i + 1, 27) = Sqr((Hoja1.Cells(3, 15) - Hoja1.Cells(13, 15)) ^ 2
+ (Hoja1.Cells(3, 16) - Hoja1.Cells(13, 16)) ^ 2)
        Hoja2.Cells(i + 2, 27) = Sqr((Hoja1.Cells(4, 15) - Hoja1.Cells(13, 15)) ^ 2
+ (Hoja1.Cells(4, 16) - Hoja1.Cells(13, 16)) ^ 2)
        Hoja2.Cells(i + 3, 27) = Sqr((Hoja1.Cells(5, 15) - Hoja1.Cells(13, 15)) ^ 2
+ (Hoja1.Cells(5, 16) - Hoja1.Cells(13, 16)) ^ 2)
        Hoja2.Cells(i + 4, 27) = Sqr((Hoja1.Cells(6, 15) - Hoja1.Cells(13, 15)) ^ 2
+ (Hoja1.Cells(6, 16) - Hoja1.Cells(13, 16)) ^ 2)
        Hoja2.Cells(i + 5, 27) = Sqr((Hoja1.Cells(7, 15) - Hoja1.Cells(13, 15)) ^ 2
+ (Hoja1.Cells(7, 16) - Hoja1.Cells(13, 16)) ^ 2)
        Hoja2.Cells(i + 6, 27) = Sqr((Hoja1.Cells(8, 15) - Hoja1.Cells(13, 15)) ^ 2
+ (Hoja1.Cells(8, 16) - Hoja1.Cells(13, 16)) ^ 2)
        Hoja2.Cells(i + 7, 27) = Sqr((Hoja1.Cells(9, 15) - Hoja1.Cells(13, 15)) ^ 2
+ (Hoja1.Cells(9, 16) - Hoja1.Cells(13, 16)) ^ 2)
        Hoja2.Cells(i + 8, 27) = Sqr((Hoja1.Cells(10, 15) - Hoja1.Cells(13, 15)) ^
2 + (Hoja1.Cells(10, 16) - Hoja1.Cells(13, 16)) ^ 2)
        Hoja2.Cells(i + 9, 27) = Sqr((Hoja1.Cells(11, 15) - Hoja1.Cells(13, 15)) ^
2 + (Hoja1.Cells(11, 16) - Hoja1.Cells(13, 16)) ^ 2)
        Hoja2.Cells(i + 10, 27) = Sqr((Hoja1.Cells(12, 15) - Hoja1.Cells(13, 15))
^ 2 + (Hoja1.Cells(12, 16) - Hoja1.Cells(13, 16)) ^ 2)
        Hoja2.Cells(i + 11, 27) = 0 'ESTACION CENTRAL
        Hoja2.Cells(i + 12, 27) = 0
    End If
    i = i + tiendas
Wend; End Sub

```

Anexo N° 10: Código macro "Elección".

```
Sub Eleccion(meses)
ni = Hoja2.Cells(2, 74)          ' número de clientes totales en la base de
datos
tiendas = 13                    ' 8 tiendas más la alternativa de no comprar.
'meses = 54
observaciones = 0
For i = 1 To ni                  ' número total de transacciones originales
    j = Hoja2.Cells(i + 1, 76)
    observaciones = observaciones + j
Next i
fila_original = 1
k = 0
For k = 1 To observaciones
    fila_original = fila_original + 1
    cliente = Hoja1.Cells(fila_original, 1)
    fecha = (Hoja1.Cells(fila_original, 12) & Hoja1.Cells(fila_original, 13))
    monto = Hoja1.Cells(fila_original, 11)
    tienda = Hoja1.Cells(fila_original, 10)
    For fila_nueva = 2 To meses * tiendas * ni + 1
        fecha1 = CStr(Hoja2.Cells(fila_nueva, 2))
        If Hoja2.Cells(fila_nueva, 3) = tienda And fecha1 = fecha And
Hoja2.Cells(fila_nueva, 1) = cliente Then
            If tienda = "ALTO LAS CONDES" Then
                Hoja2.Cells(fila_nueva + tiendas - 1, 5) = 0
                Hoja2.Cells(fila_nueva, 5) = 1
                Hoja2.Cells(fila_nueva, 6) = monto
                Hoja2.Cells(fila_nueva, 4) = 1
                Hoja2.Cells(fila_nueva + 1, 4) = 1
                Hoja2.Cells(fila_nueva + 2, 4) = 1
                Hoja2.Cells(fila_nueva + 3, 4) = 1
                Hoja2.Cells(fila_nueva + 4, 4) = 1
                Hoja2.Cells(fila_nueva + 5, 4) = 1
                Hoja2.Cells(fila_nueva + 6, 4) = 1
                Hoja2.Cells(fila_nueva + 7, 4) = 1
                Hoja2.Cells(fila_nueva + 8, 4) = 1
                Hoja2.Cells(fila_nueva + 9, 4) = 1
                Hoja2.Cells(fila_nueva + 10, 4) = 1
                Hoja2.Cells(fila_nueva + 11, 4) = 1
                Hoja2.Cells(fila_nueva + 12, 4) = 1
            Elseif tienda = "PARQUE ARAUCO" Then
                Hoja2.Cells(fila_nueva + tiendas - 2, 5) = 0
                Hoja2.Cells(fila_nueva, 5) = 1
                Hoja2.Cells(fila_nueva, 6) = monto
```

```

Hoja2.Cells(fila_nueva - 1, 4) = 2
Hoja2.Cells(fila_nueva, 4) = 2
Hoja2.Cells(fila_nueva + 1, 4) = 2
Hoja2.Cells(fila_nueva + 2, 4) = 2
Hoja2.Cells(fila_nueva + 3, 4) = 2
Hoja2.Cells(fila_nueva + 4, 4) = 2
Hoja2.Cells(fila_nueva + 5, 4) = 2
Hoja2.Cells(fila_nueva + 6, 4) = 2
Hoja2.Cells(fila_nueva + 7, 4) = 2
Hoja2.Cells(fila_nueva + 8, 4) = 2
Hoja2.Cells(fila_nueva + 9, 4) = 2
Hoja2.Cells(fila_nueva + 10, 4) = 2
Hoja2.Cells(fila_nueva + 11, 4) = 2
Elseif tienda = "LYON" Then
    Hoja2.Cells(fila_nueva + tiendas - 3, 5) = 0
    Hoja2.Cells(fila_nueva, 5) = 1
    Hoja2.Cells(fila_nueva, 6) = monto
    Hoja2.Cells(fila_nueva - 2, 4) = 3
    Hoja2.Cells(fila_nueva - 1, 4) = 3
    Hoja2.Cells(fila_nueva, 4) = 3
    Hoja2.Cells(fila_nueva + 1, 4) = 3
    Hoja2.Cells(fila_nueva + 2, 4) = 3
    Hoja2.Cells(fila_nueva + 3, 4) = 3
    Hoja2.Cells(fila_nueva + 4, 4) = 3
    Hoja2.Cells(fila_nueva + 5, 4) = 3
    Hoja2.Cells(fila_nueva + 6, 4) = 3
    Hoja2.Cells(fila_nueva + 7, 4) = 3
    Hoja2.Cells(fila_nueva + 8, 4) = 3
    Hoja2.Cells(fila_nueva + 9, 4) = 3
    Hoja2.Cells(fila_nueva + 10, 4) = 3
Elseif tienda = "CENTRO" Then
    Hoja2.Cells(fila_nueva + tiendas - 4, 5) = 0
    Hoja2.Cells(fila_nueva, 5) = 1
    Hoja2.Cells(fila_nueva, 6) = monto
    Hoja2.Cells(fila_nueva - 3, 4) = 4
    Hoja2.Cells(fila_nueva - 2, 4) = 4
    Hoja2.Cells(fila_nueva - 1, 4) = 4
    Hoja2.Cells(fila_nueva, 4) = 4
    Hoja2.Cells(fila_nueva + 1, 4) = 4
    Hoja2.Cells(fila_nueva + 2, 4) = 4
    Hoja2.Cells(fila_nueva + 3, 4) = 4
    Hoja2.Cells(fila_nueva + 4, 4) = 4
    Hoja2.Cells(fila_nueva + 5, 4) = 4
    Hoja2.Cells(fila_nueva + 6, 4) = 4
    Hoja2.Cells(fila_nueva + 7, 4) = 4
    Hoja2.Cells(fila_nueva + 8, 4) = 4
    Hoja2.Cells(fila_nueva + 9, 4) = 4
Elseif tienda = "TOBALABA" Then

```

```

Hoja2.Cells(fila_nueva + tiendas - 5, 5) = 0
Hoja2.Cells(fila_nueva, 5) = 1
Hoja2.Cells(fila_nueva, 6) = monto
Hoja2.Cells(fila_nueva - 4, 4) = 5
Hoja2.Cells(fila_nueva - 3, 4) = 5
Hoja2.Cells(fila_nueva - 2, 4) = 5
Hoja2.Cells(fila_nueva - 1, 4) = 5
Hoja2.Cells(fila_nueva, 4) = 5
Hoja2.Cells(fila_nueva + 1, 4) = 5
Hoja2.Cells(fila_nueva + 2, 4) = 5
Hoja2.Cells(fila_nueva + 3, 4) = 5
Hoja2.Cells(fila_nueva + 4, 4) = 5
Hoja2.Cells(fila_nueva + 5, 4) = 5
Hoja2.Cells(fila_nueva + 6, 4) = 5
Hoja2.Cells(fila_nueva + 7, 4) = 5
Hoja2.Cells(fila_nueva + 8, 4) = 5
Elseif tienda = "PLAZA VESPUCIO" Then
Hoja2.Cells(fila_nueva + tiendas - 6, 5) = 0
Hoja2.Cells(fila_nueva, 5) = 1
Hoja2.Cells(fila_nueva, 6) = monto
Hoja2.Cells(fila_nueva - 5, 4) = 6
Hoja2.Cells(fila_nueva - 4, 4) = 6
Hoja2.Cells(fila_nueva - 3, 4) = 6
Hoja2.Cells(fila_nueva - 2, 4) = 6
Hoja2.Cells(fila_nueva - 1, 4) = 6
Hoja2.Cells(fila_nueva, 4) = 6
Hoja2.Cells(fila_nueva + 1, 4) = 6
Hoja2.Cells(fila_nueva + 2, 4) = 6
Hoja2.Cells(fila_nueva + 3, 4) = 6
Hoja2.Cells(fila_nueva + 4, 4) = 6
Hoja2.Cells(fila_nueva + 5, 4) = 6
Hoja2.Cells(fila_nueva + 6, 4) = 6
Hoja2.Cells(fila_nueva + 7, 4) = 6
Elseif tienda = "PLAZA OESTE" Then
Hoja2.Cells(fila_nueva + tiendas - 7, 5) = 0
Hoja2.Cells(fila_nueva, 5) = 1
Hoja2.Cells(fila_nueva, 6) = monto
Hoja2.Cells(fila_nueva - 6, 4) = 7
Hoja2.Cells(fila_nueva - 5, 4) = 7
Hoja2.Cells(fila_nueva - 4, 4) = 7
Hoja2.Cells(fila_nueva - 3, 4) = 7
Hoja2.Cells(fila_nueva - 2, 4) = 7
Hoja2.Cells(fila_nueva - 1, 4) = 7
Hoja2.Cells(fila_nueva, 4) = 7
Hoja2.Cells(fila_nueva + 1, 4) = 7
Hoja2.Cells(fila_nueva + 2, 4) = 7
Hoja2.Cells(fila_nueva + 3, 4) = 7
Hoja2.Cells(fila_nueva + 4, 4) = 7

```

```

Hoja2.Cells(fila_nueva + 5, 4) = 7
Hoja2.Cells(fila_nueva + 6, 4) = 7
Elseif tienda = "PLAZA NORTE" Then
Hoja2.Cells(fila_nueva + tiendas - 8, 5) = 0
Hoja2.Cells(fila_nueva, 5) = 1
Hoja2.Cells(fila_nueva, 6) = monto
Hoja2.Cells(fila_nueva - 7, 4) = 8
Hoja2.Cells(fila_nueva - 6, 4) = 8
Hoja2.Cells(fila_nueva - 5, 4) = 8
Hoja2.Cells(fila_nueva - 4, 4) = 8
Hoja2.Cells(fila_nueva - 3, 4) = 8
Hoja2.Cells(fila_nueva - 2, 4) = 8
Hoja2.Cells(fila_nueva - 1, 4) = 8
Hoja2.Cells(fila_nueva, 4) = 8
Hoja2.Cells(fila_nueva + 1, 4) = 8
Hoja2.Cells(fila_nueva + 2, 4) = 8
Hoja2.Cells(fila_nueva + 3, 4) = 8
Hoja2.Cells(fila_nueva + 4, 4) = 8
Hoja2.Cells(fila_nueva + 5, 4) = 8
Elseif tienda = "PLAZA PUENTE" Then
Hoja2.Cells(fila_nueva + tiendas - 9, 5) = 0
Hoja2.Cells(fila_nueva, 5) = 1
Hoja2.Cells(fila_nueva, 6) = monto
Hoja2.Cells(fila_nueva - 8, 4) = 9
Hoja2.Cells(fila_nueva - 7, 4) = 9
Hoja2.Cells(fila_nueva - 6, 4) = 9
Hoja2.Cells(fila_nueva - 5, 4) = 9
Hoja2.Cells(fila_nueva - 4, 4) = 9
Hoja2.Cells(fila_nueva - 3, 4) = 9
Hoja2.Cells(fila_nueva - 2, 4) = 9
Hoja2.Cells(fila_nueva - 1, 4) = 9
Hoja2.Cells(fila_nueva, 4) = 9
Hoja2.Cells(fila_nueva + 1, 4) = 9
Hoja2.Cells(fila_nueva + 2, 4) = 9
Hoja2.Cells(fila_nueva + 3, 4) = 9
Hoja2.Cells(fila_nueva + 4, 4) = 9
Elseif tienda = "LA DEHESA" Then
Hoja2.Cells(fila_nueva + tiendas - 10, 5) = 0
Hoja2.Cells(fila_nueva, 5) = 1
Hoja2.Cells(fila_nueva, 6) = monto
Hoja2.Cells(fila_nueva - 9, 4) = 10
Hoja2.Cells(fila_nueva - 8, 4) = 10
Hoja2.Cells(fila_nueva - 7, 4) = 10
Hoja2.Cells(fila_nueva - 6, 4) = 10
Hoja2.Cells(fila_nueva - 5, 4) = 10
Hoja2.Cells(fila_nueva - 4, 4) = 10
Hoja2.Cells(fila_nueva - 3, 4) = 10
Hoja2.Cells(fila_nueva - 2, 4) = 10

```

```

Hoja2.Cells(fila_nueva - 1, 4) = 10
Hoja2.Cells(fila_nueva, 4) = 10
Hoja2.Cells(fila_nueva + 1, 4) = 10
Hoja2.Cells(fila_nueva + 2, 4) = 10
Hoja2.Cells(fila_nueva + 3, 4) = 10
Elseif tienda = "MANQUEHUE" Then
    Hoja2.Cells(fila_nueva + tiendas - 11, 5) = 0
    Hoja2.Cells(fila_nueva, 5) = 1
    Hoja2.Cells(fila_nueva, 6) = monto
    Hoja2.Cells(fila_nueva - 10, 4) = 11
    Hoja2.Cells(fila_nueva - 9, 4) = 11
    Hoja2.Cells(fila_nueva - 8, 4) = 11
    Hoja2.Cells(fila_nueva - 7, 4) = 11
    Hoja2.Cells(fila_nueva - 6, 4) = 11
    Hoja2.Cells(fila_nueva - 5, 4) = 11
    Hoja2.Cells(fila_nueva - 4, 4) = 11
    Hoja2.Cells(fila_nueva - 3, 4) = 11
    Hoja2.Cells(fila_nueva - 2, 4) = 11
    Hoja2.Cells(fila_nueva - 1, 4) = 11
    Hoja2.Cells(fila_nueva, 4) = 11
    Hoja2.Cells(fila_nueva + 1, 4) = 11
    Hoja2.Cells(fila_nueva + 2, 4) = 11
Elseif tienda = "ESTACION CENTRAL" Then
    Hoja2.Cells(fila_nueva + tiendas - 12, 5) = 0
    Hoja2.Cells(fila_nueva, 5) = 1
    Hoja2.Cells(fila_nueva, 6) = monto
    Hoja2.Cells(fila_nueva - 11, 4) = 12
    Hoja2.Cells(fila_nueva - 10, 4) = 12
    Hoja2.Cells(fila_nueva - 9, 4) = 12
    Hoja2.Cells(fila_nueva - 8, 4) = 12
    Hoja2.Cells(fila_nueva - 7, 4) = 12
    Hoja2.Cells(fila_nueva - 6, 4) = 12
    Hoja2.Cells(fila_nueva - 5, 4) = 12
    Hoja2.Cells(fila_nueva - 4, 4) = 12
    Hoja2.Cells(fila_nueva - 3, 4) = 12
    Hoja2.Cells(fila_nueva - 2, 4) = 12
    Hoja2.Cells(fila_nueva - 1, 4) = 12
    Hoja2.Cells(fila_nueva, 4) = 12
    Hoja2.Cells(fila_nueva + 1, 4) = 12
Else
    Hoja2.Cells(fila_nueva, 5) = 1
    Hoja2.Cells(fila_nueva, 6) = 0
    Hoja2.Cells(fila_nueva - 12, 4) = 13
    Hoja2.Cells(fila_nueva - 11, 4) = 13
    Hoja2.Cells(fila_nueva - 10, 4) = 13
    Hoja2.Cells(fila_nueva - 9, 4) = 13
    Hoja2.Cells(fila_nueva - 8, 4) = 13
    Hoja2.Cells(fila_nueva - 7, 4) = 13

```

```

        Hoja2.Cells(fila_nueva - 6, 4) = 13
        Hoja2.Cells(fila_nueva - 5, 4) = 13
        Hoja2.Cells(fila_nueva - 4, 4) = 13
        Hoja2.Cells(fila_nueva - 3, 4) = 13
        Hoja2.Cells(fila_nueva - 2, 4) = 13
        Hoja2.Cells(fila_nueva - 1, 4) = 13
        Hoja2.Cells(fila_nueva, 4) = 13
    End If
End If
Next fila_nueva
Next k
End Sub

```

Anexo N° 11: Código macro “Código”.

```

Sub Codigo(meses)
ni = Hoja2.Cells(2, 74)          ' número de clientes totales en la base de
datos
tiendas = 13                    ' 8 tiendas más la alternativa de no comprar.
For fila_nueva = 2 To meses * tiendas * ni + 1
    tiendas = Hoja2.Cells(fila_nueva, 3)
    If tiendas = "ALTO LAS CONDES" Then
        Hoja2.Cells(fila_nueva, 3) = 100021
    ElseIf tiendas = "PARQUE ARAUCO" Then
        Hoja2.Cells(fila_nueva, 3) = 100022
    ElseIf tiendas = "LYON" Then
        Hoja2.Cells(fila_nueva, 3) = 100023
    ElseIf tiendas = "CENTRO" Then
        Hoja2.Cells(fila_nueva, 3) = 100025
    ElseIf tiendas = "TOBALABA" Then
        Hoja2.Cells(fila_nueva, 3) = 100045
    ElseIf tiendas = "PLAZA VESPUCIO" Then
        Hoja2.Cells(fila_nueva, 3) = 100026
    ElseIf tiendas = "PLAZA OESTE" Then
        Hoja2.Cells(fila_nueva, 3) = 100027
    ElseIf tiendas = "PLAZA NORTE" Then
        Hoja2.Cells(fila_nueva, 3) = 100064
    ElseIf tiendas = "PLAZA PUENTE" Then
        Hoja2.Cells(fila_nueva, 3) = 100067
    ElseIf tiendas = "LA DEHESA" Then
        Hoja2.Cells(fila_nueva, 3) = 100065
    ElseIf tiendas = "MANQUEHUE" Then
        Hoja2.Cells(fila_nueva, 3) = 100071
    End If
End For
End Sub

```

```

Elseif tiendas = "ESTACION CENTRAL" Then
    Hoja2.Cells(fila_nueva, 3) = 100073
Elseif tiendas = "NO COMPRO" Then
    Hoja2.Cells(fila_nueva, 3) = 100000
End If
Next fila_nueva
End Sub

```

Anexo N° 12: Código macro "Identidad".

```

Sub Identidad(meses)
ni = Hoja2.Cells(2, 74)
Tiendas_Antiguas = 8
tiendas = 13
For i = 1 To Tiendas_Antiguas - 1
    For j = 2 To ni * meses * (tiendas) +
1
        Hoja2.Cells(j, i + 6) = 0
    Next j
Next i
For i = 1 To Tiendas_Antiguas - 1
    For j = 2 To ni * meses * (tiendas) +
1
        Hoja2.Cells(j + i - 1, i + 6) = 1
        j = j + tiendas - 1
    Next j
Next i
End Sub

```

Anexo N° 13: Código macro "Identidad_Entrada_PN".

```

Sub Identidad_Entrada_PN(meses)
ni = Hoja2.Cells(2, 74)
Tiendas_Antiguas = 8
tiendas = 13
For i = 1 To Tiendas_Antiguas
    For j = 2 To ni * meses * (tiendas) + 1

```

```

    Hoja2.Cells(j, i + 13) = 0
  Next j
Next i
For i = 1 To Tiendas_Antiguas
  For j = 2 To ni * meses * (tiendas) + 1
    fecha = Hoja2.Cells(j, 2)
    If fecha < 2003111 Then
      Hoja2.Cells(j + i - 1, i + 13) = 0
      j = j + tiendas - 1
    Else
      Hoja2.Cells(j + i - 1, i + 13) = 1
      j = j + tiendas - 1
    End If
  Next j
Next i
End Sub

```

Anexo N° 14: Código macro “Identidad_Entrada_PP”.

```

Sub Identidad_Entrada_PP(meses)
  ni = Hoja2.Cells(2, 74)
  Tiendas_Antiguas = 8
  tiendas = 13
  For i = 1 To Tiendas_Antiguas + 1
    For j = 2 To ni * meses * (tiendas) + 1
      Hoja2.Cells(j, i + 28) = 0
    Next j
  Next i
  For i = 1 To Tiendas_Antiguas + 1
    For j = 2 To ni * meses * (tiendas) + 1
      fecha = Hoja2.Cells(j, 2)
      If fecha < 2005101 Then
        Hoja2.Cells(j + i - 1, i + 28) = 0
        j = j + tiendas - 1
      Else
        Hoja2.Cells(j + i - 1, i + 28) = 1
        j = j + tiendas - 1
      End If
    Next j
  Next i
End Sub

```

Anexo N° 15: Código macro "Identidad_Entrada_DE".

```
Sub Identidad_Entrada_DE(meses)
ni = Hoja2.Cells(2, 74)
Tiendas_Antiguas = 8
tiendas = 13
For i = 1 To Tiendas_Antiguas + 2
  For j = 2 To ni * meses * (tiendas) + 1
    Hoja2.Cells(j, i + 37) = 0
  Next j
Next i
For i = 1 To Tiendas_Antiguas + 2
  For j = 2 To ni * meses * (tiendas) + 1
    fecha = Hoja2.Cells(j, 2)
    If fecha < 20051111 Then
      Hoja2.Cells(j + i - 1, i + 37) = 0
      j = j + tiendas - 1
    Else
      Hoja2.Cells(j + i - 1, i + 37) = 1
      j = j + tiendas - 1
    End If
  Next j
Next i
End Sub
```

Anexo N° 16: Código macro "Identidad_Entrada_MA".

```
Sub Identidad_Entrada_MA(meses)
ni = Hoja2.Cells(2, 74)
Tiendas_Antiguas = 8
tiendas = 13
For i = 1 To Tiendas_Antiguas + 3
  For j = 2 To ni * meses * (tiendas) + 1
    Hoja2.Cells(j, i + 47) = 0
  Next j
Next i
For i = 1 To Tiendas_Antiguas + 3
  For j = 2 To ni * meses * (tiendas) + 1
```

```

fecha = Hoja2.Cells(j, 2)
If fecha < 2007051 Then
    Hoja2.Cells(j + i - 1, i + 47) = 0
    j = j + tiendas - 1
Else
    Hoja2.Cells(j + i - 1, i + 47) = 1
    j = j + tiendas - 1
End If
Next j
Next i
End Sub

```

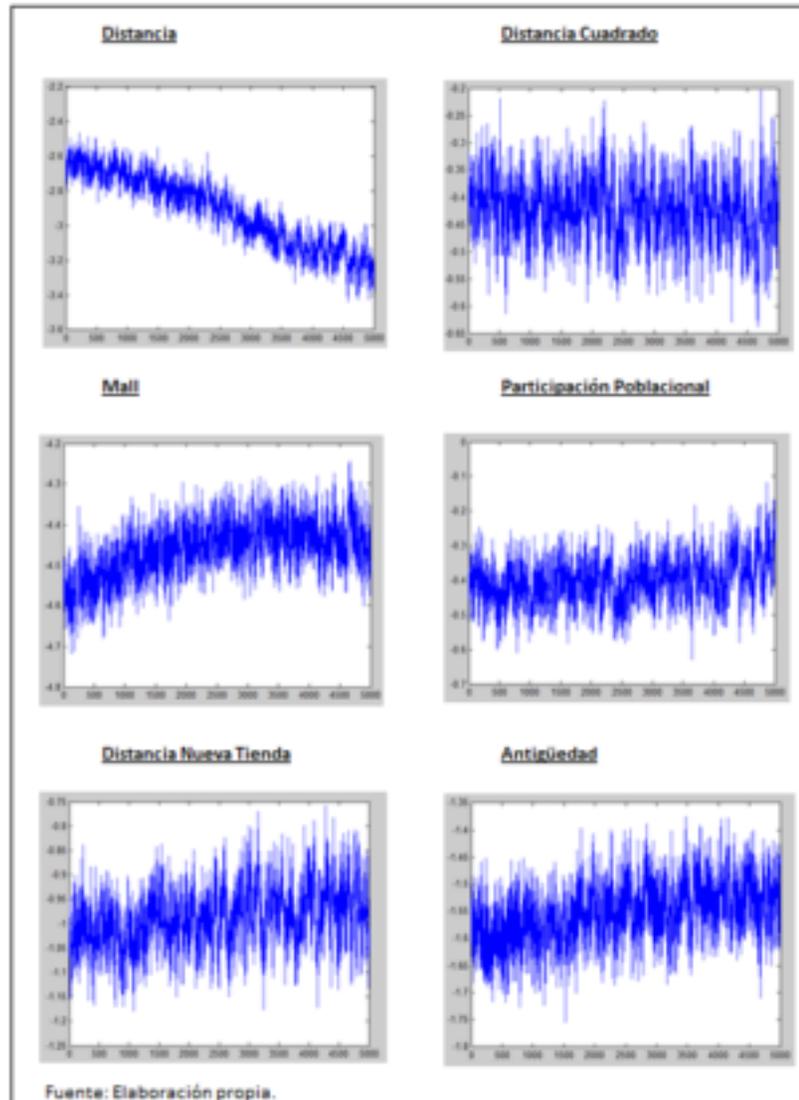
Anexo N° 17: Código macro “Identidad_Entrada_EC”.

```

Sub Identidad_Entrada_EC(meses)
ni = Hoja2.Cells(2, 74)
Tiendas_Antiguas = 8
tiendas = 13
For i = 1 To Tiendas_Antiguas + 4
    For j = 2 To ni * meses * (tiendas) + 1
        Hoja2.Cells(j, i + 58) = 0
    Next j
Next i
For i = 1 To Tiendas_Antiguas + 4
    For j = 2 To ni * meses * (tiendas) + 1
        fecha = Hoja2.Cells(j, 2)
        If fecha < 2008061 Then
            Hoja2.Cells(j + i - 1, i + 58) = 0
            j = j + tiendas - 1
        Else
            Hoja2.Cells(j + i - 1, i + 58) = 1
            j = j + tiendas - 1
        End If
    Next j
Next i
End Sub

```

Anexo N° 18: Gráficos de convergencia variables explicativas Mixed Logit.



Anexo N° 19: Intervalo confianza variables Mixed Logit.

	INTERVALO CONFIANZA		
	0.025	0.5	0.095
Intercepto Alto las Condes	-0,14	0,06	0,28

Intercepto Arauco	0,01	0,19	0,37
Intercepto Lyon	-3,71	-3,51	-3,35
Intercepto Centro	-2,52	-2,33	-2,15
Intercepto Tobalaba	0,1	0,31	0,52
Intercepto Vespucio	1,56	1,77	2
Intercepto Oeste	1,08	1,29	1,5
Distancia	-3,28	-2,9	-2,59
Distancia al Cuadrado	-0,54	-0,42	-0,31
Mall	-4,6	-4,46	-4,34
Participación Población	-0,53	-0,4	-0,27
Distancia Nueva Tienda	-1,1	-0,99	-0,86
Antigüedad	-1,66	-1,55	-1,45
PN/Alto las Condes	-0,3	-0,16	-0,01
PN/Arauco	-0,28	-0,14	0,01
PN/Lyon	-0,88	-0,74	-0,58
PN/Centro	-0,14	-0,01	0,14
PN/Tobalaba	-0,17	-0,03	0,1
PN/Vespucio	0,44	0,55	0,66
PN/Oeste	0,31	0,43	0,55
Plaza Norte (PN)	-1,27	-1,09	-0,89
PP/Alto las Condes	-0,8	-0,69	-0,58
PP/Arauco	-0,76	-0,59	-0,45
PN/Lyon	-0,69	-0,58	-0,47
PP/Centro	-1,1	-0,99	-0,89
PP/Tobalaba	-1,07	-0,95	-0,83
PP/Vespucio	-1,08	-0,96	-0,85
PP/Oeste	-1,33	-1,18	-1,04
PP/PN	-1,15	-1,03	-0,92
Plaza Puente (PP)	-0,94	-0,82	-0,69
DE/Alto las Condes	-0,6	-0,47	-0,34
DE/Arauco	-0,36	-0,24	-0,12
DE/Lyon	-0,51	-0,4	-0,27
DE/Centro	-0,54	-0,43	-0,31
DE/Tobalaba	-0,82	-0,69	-0,58
DE/Vespucio	-0,69	-0,56	-0,43
DE/Oeste	-0,56	-0,45	-0,33
DE/PN (21)	-1,17	-1,04	-0,91
DE/PP (37)	-1,46	-1,32	-1,19
La Dehesa (DE)	-2,57	-2,43	-2,3

MA/Alto las Condes	-0,49	-0,37	-0,25
MA/Arauco	-0,19	-0,07	0,04
MA/Lyon	-0,47	-0,35	-0,24
MA/Centro	-0,31	-0,2	-0,07
MA/Tobalaba	-0,59	-0,48	-0,35
MA/Vespucio	-0,33	-0,21	-0,09
MA/Oeste	-0,17	-0,04	0,08
MA/PN (21)	-0,48	-0,37	-0,24
MA/PP (37)	-0,31	-0,2	-0,08
MA/DE (47)	-1,21	-1,08	-0,94
Manquehue (MA)	-2,26	-2,1	-1,97
EC/Alto las Condes	-0,25	-0,13	0
EC/Arauco	0,09	0,22	0,34
EC/Lyon	0,09	0,21	0,32
EC/Centro	-0,08	0,04	0,15
EC/Tobalaba	-0,14	-0,03	0,08
EC/Vespucio	-0,24	-0,12	-0,01
EC/Oeste	-0,09	0,03	0,14
EC/PN	-0,02	0,1	0,22
EC/PP	-0,29	-0,17	-0,06
EC/DE	-0,46	-0,35	-0,22
EC/MA	-0,9	-0,77	-0,64
Estación Central (EC)	-1,36	-1,2	-1,07

Anexo N° 20: Modelo Monto de Compra (Regresión).

```

clc;
clear;

fprintf('Regression model\n');
data=load('DatosR.txt','-ascii');
n=1473;
k=4;      % set nobs and nvars
r = [0 0 0 0]';      % prior means
x=data(1:n,2:end);
b = ones(k+1,1);      % generate data set
y=data(1:n,1);
R = eye(k);
T = eye(k);      % prior variance

```

```

Q = chol(inv(T));
q = Q*r;
u=x'*x;
uu=x'*y;
b0 = (x'*x)\(x'*y);          % use ols as initial values
sige = (y-x*b0)'*(y-x*b0)/(n-k);
xpx = x'*x;
xpy = x'*y;                  % calculate x'x, x'y only once
qpq = Q'*Q;
qpv = Q'*q;                  % calculate Q'Q, Q'q only once
ndraw = 200000;
nomit = ndraw/2;             % set the number of draws
bsave = zeros(ndraw,k);     % allocate storage for results
ssave = zeros(ndraw,1);
tic;

% Start the sampling

for i=1:ndraw;
    t=sige*qpq;
    xpxi = inv(xpx + sige*qpq);
    media_b = xpxi*(xpy + sige*qpv);      % Paso 2.
    varianza_b=sige*xpxi;                % Paso 2.
    b = norm_rnd(varianza_b)+media_b;    % Paso 3.
% draw MV normal with mean(b), var(b)
    bsave(i,:) = b';                    % save b draws
    e = y-x*b;
    ssr = e'*e;                         % update sige
    chi = chis_rnd(1,n);                 % do chisquared(n) draw
    sige = ssr/chi;                      % Paso 4.
    ssave(i,1) = sige;                   % save sige draws
end;

% End the sampling

toc;
bhat = mean(bsave(nomit+1:ndraw,:)); % calculate means and
std deviations
bstd = std(bsave(nomit+1:ndraw,:));
tstat = bhat./bstd;
sighat = mean(ssave(nomit+1:ndraw,1));
tout = tdis_prb(tstat',n); % compute t-stat significance levels

% set up for printing results

```

```

in.cnames = strcat('Coefficient','t-statistic','t-probability');
in.rnames = strcat('Variable','variable 1','variable 2','variable 3');
in.fmt = '%16.6f';
fprintf(1,'Gibbs estimates \n');
fprintf(1,'Coefficient    t-statistic    t-probability \n');
tmp = [bhat' tstat' tout]
%mprint(tmp,in);
result = theil(y,x,r,R,T);    % compare to Theil-Goldberger
estimates
prt(result);

```

```

function y = norm_rnd(sig);
% PURPOSE: random multivariate random vector
based on

if nargin ~= 1
error('Wrong # of arguments to norm_rnd');
end;

h = chol(sig);
[nrow, ncol] = size(sig);
rv = randn(nrow,1);
y = h*rv;

```

```

function rn=chis_rnd(nn,v)
% PURPOSE: generates random chi-squared deviates

rn = rgamma(nn,v*0.5)*2;

function x = rgamma(nn,a)

if any(any(a<=0))
error('Parameter a is wrong')
end

n = prod(nn);
if length(nn) == 1
nn(2) = 1;
end

y0 = log(a)-1/sqrt(a);

```

```

c = a - exp(y0);
m = ceil(n*(1.7 + 0.6*(min(min(a))<2)));

y = log(rand(m,1)).*sign(rand(m,1)-0.5)/c + log(a);
f = a*y-exp(y) - (a*y0 - exp(y0));
g = c*(abs((y0-log(a))) - abs(y-log(a)));
reject = (log(rand(m,1)) + g) > f;
y(reject) = [];
x = zeros(n,1);
if length(y) >= n
    x = exp(y(1:n));
else
    tmp = rgamma(n - length(y), a);
    x = [exp(y)
        tmp];
end

```

```

function y = tdis_prb(x,n)
% PURPOSE: calculates t-probabilities for
elements in x-vector

if nargin ~= 2; error('Wrong # of arguments to
tdis_prb'); end;
if n <=0; error('dof is negative or zero in tdis_prb');
end;

x2 = n./(n+x.^2);
one = find(x2 >= 1);
if length(one) > 0
    x2(one,1) = 1-1e-12;
end;
zip = find(x2 <= 0);
if length(zip) > 0
    x2(zip,1) = 1e-12;
end;

tmp = 1.0 - 0.5*betainc(x2,0.5*n,0.5);
y = 2*(1-tmp);

```

```

function results=theil(y,x,rvec,rmat,v)
% PURPOSE: computes Theil-Goldberger mixed
estimator

```

```

if (nargin ~= 5); error('Wrong # of arguments to theil');
end;

[nobs nvar] = size(x);

results.meth = 'theil';
results.y = y;
results.nobs = nobs;
results.nvar = nvar;
[rsize junk] = size(rvec);
% fill in prior means and std deviations
results.pmean = rvec;
results.pstd = sqrt(diag(v));

vi = inv(v);

% do ols to get sig estimate;
bols = (x'*x)\(x'*y);
sige = ((y - x*bols)*(y-x*bols))/(nobs-nvar);
xpxi = inv(x'*x + sige*rmat'*vi*rmat);
xpy = (x'*y) + sige*rmat'*vi*rvec;
results.beta = xpxi*xpy;
results.yhat = x*results.beta;
results.resid = y - results.yhat;
sigu = results.resid'*results.resid;
results.sige = (sigu)/(nobs-nvar);
tmp = results.sige*diag(xpxi);
results.tstat = results.beta./sqrt(tmp);
ym = y - ones(nobs,1)*mean(y);
rsqr1 = sigu;
rsqr2 = ym'*ym;
results.rsqr = 1.0 - rsqr1/rsqr2; % r-squared
rsqr1 = rsqr1/(nobs-nvar);
rsqr2 = rsqr2/(nobs-1.0);
results.rbar = 1 - (rsqr1/rsqr2); % rbar-squared
ediff = results.resid(2:nobs) - results.resid(1:nobs-1);
results.dw = diag((ediff*ediff)./(sigu))'; % durbin-watson

```

```

function prt(results,vnames,fid)
% PURPOSE: Prints results structures returned by most functions
% error checking on inputs

```

```

if ~isstruct(results)
error('prt: requires a structure input');
elseif nargin == 3
arg = 0;
[vsize junk] = size(vnames); % user may supply a blank argument
if vsize > 0
arg = 3;
end;
elseif nargin == 2
arg = 2;
elseif nargin == 1
arg = 1;
else
error('Wrong # of inputs to prt');
end;
method = results(1).meth;

% call appropriate printing routine
switch method

case {'arma','boxcox','boxcox2','hwhite','lad','logit','mlogit','nwest','ols','olsc',...
'olsar1','olst','probit','ridge','robust','theil','tobit','tsls'}
% call prt_reg
if arg == 1
prt_reg(results);
elseif arg == 2
prt_reg(results,vnames);
elseif arg == 3
prt_reg(results,vnames,fid);
else
prt_reg(results,[],fid);
end;

case {'switch_em','hmarkov_em'}
% call prt_swm
if arg == 1
prt_swm(results);
elseif arg == 2
prt_swm(results,vnames);
elseif arg == 3
prt_swm(results,vnames,fid);
else
prt_swm(results,[],fid);
end;

```

```

case {'psem','semsfe','semtfe','semstfe'}
    % call prt_sp
    if arg == 1
        prt_sp(results);
    elseif arg == 2
        prt_sp(results,vnames);
    elseif arg == 3
        prt_sp(results,vnames,fid);
    else
        prt_sp(results,[],fid);
    end;

case
{'psem','semsfe','semtfe','semstfe','psar','sarsfe','sartfe','sarstfe','sarsre','semsre'}
    % call prt_spnew
    if arg == 1
        prt_spnew(results);
    elseif arg == 2
        prt_spnew(results,vnames);
    elseif arg == 3
        prt_spnew(results,vnames,fid);
    else
        prt_spnew(results,[],fid);
    end;

case {'thsls','sur'}
    % call prt_eqs
    if arg == 1
        prt_eqs(results);
    elseif arg == 2
        prt_eqs(results,vnames);
    elseif arg == 3
        prt_eqs(results,vnames,fid);
    else
        prt_eqs(results,[],fid);
    end;

case {'sem_gmm','sem2_gmm','sac_gmm','sar_gmm'}
    % call prt_gmm
    if arg == 1
        prt_gmm(results);
    elseif arg == 2
        prt_gmm(results,vnames);

```

```

elseif arg == 3
prt_gmm(results,vnames,fid);
else
prt_gmm(results,[],fid);
end;

case {'ar','vare','bvar','rvar','ecm','becm','recm'}
% call prt_var
if arg == 1
prt_var(results);
elseif arg == 2
prt_var(results,vnames);
elseif arg == 3
prt_var(results,vnames,fid);
else
prt_var(results,[],fid);
end;

case {'bvar_g','rvar_g','becm_g','recm_g'}
% call prt_varg
if arg == 1
prt_varg(results);
elseif arg == 2
prt_varg(results,vnames);
elseif arg == 3
prt_varg(results,vnames,fid);
else
prt_varg(results,[],fid);
end;

case {'johansen','adf','cadf','phillips'}
% call prt_coint
if arg == 1
prt_coint(results);
elseif arg == 2
prt_coint(results,vnames);
elseif arg == 3
prt_coint(results,vnames,fid);
else
prt_coint(results,[],fid);
end;

case {'coda','raftery','apm','momentg'}
% call prt_coda

```

```

if arg == 1
prt_coda(results);
elseif arg == 2
prt_coda(results,vnames);
elseif arg == 3
prt_coda(results,vnames,fid);
else
prt_coda(results,[],fid);
end;

case {'ar_g','ols_g','ols_gc','bma_g','tobit_g','probit_g','probit_gm'}
% call prt_gibbs
if arg == 1
prt_gibbs(results);
elseif arg == 2
prt_gibbs(results,vnames);
elseif arg == 3
prt_gibbs(results,vnames,fid);
else
prt_gibbs(results,[],fid);
end;

case {'sar','sar_g','sart_g','sarp_g','sar_c','sar_gv','sarp_gc','sar_gbma'}
% call prt_sar
if arg == 1
prt_sar(results);
elseif arg == 2
prt_sar(results,vnames);
elseif arg == 3
prt_sar(results,vnames,fid);
else
prt_sar(results,[],fid);
end;

case {'sem','sem_g','semt_g','semp_g'}
% call prt_sem
if arg == 1
prt_sem(results);
elseif arg == 2
prt_sem(results,vnames);
elseif arg == 3
prt_sem(results,vnames,fid);
else
prt_sem(results,[],fid);
end;

```

```

end;

case {'semip_g','semip_gc','semit_g'}
    % call prt_semip
    if arg == 1
        prt_sem(results);
    elseif arg == 2
        prt_semip(results,vnames);
    elseif arg == 3
        prt_semip(results,vnames,fid);
    else
        prt_semip(results,[],fid);
    end;

case {'sdm','sdm_g','sdmp_g','sdmt_g','sdm_gmm'}
    % call prt_sdm
    if arg == 1
        prt_sdm(results);
    elseif arg == 2
        prt_sdm(results,vnames);
    elseif arg == 3
        prt_sdm(results,vnames,fid);
    else
        prt_sdm(results,[],fid);
    end;

case {'far','far_g','far_gc'}
    % call prt_far
    if arg == 1
        prt_far(results);
    elseif arg == 2
        prt_far(results,vnames);
    elseif arg == 3
        prt_far(results,vnames,fid);
    else
        prt_far(results,[],fid);
    end;

case {'sac','sac_g','sacp_g','sact_g'}
    % call prt_sac
    if arg == 1
        prt_sac(results);
    elseif arg == 2
        prt_sac(results,vnames);

```

```

elseif arg == 3
prt_sac(results,vnames,fid);
else
prt_sac(results,[],fid);
end;

case{'moran','lmerror','lratios','walds','lmsar'}
% call prt_spat
if arg == 1
prt_spat(results);
elseif arg == 2
prt_spat(results,vnames);
elseif arg == 3
prt_spat(results,vnames,fid);
else
prt_spat(results,[],fid);
end;

case {'mess','mess_g','mess_g1','mess_g2','mess_g3','messv_g3','messt_g', ...
'messt_g1','messt_g2','messt_g3','messvt_g3','messp_g','messp_g1','messp_g2',
...
'messp_g3','messpv_g3'}
% call prt_mess
if arg == 1
prt_mess(results);
elseif arg == 2
prt_mess(results,vnames);
elseif arg == 3
prt_mess(results,vnames,fid);
else
prt_mess(results,[],fid);
end;

case {'gwr','bgwr','bgwrv','gwr_logit','gwr_probit'}
% call prt_gwr
if arg == 1
prt_gwr(results);
elseif arg == 2
prt_gwr(results,vnames);
elseif arg == 3
prt_gwr(results,vnames,fid);
else
prt_gwr(results,[],fid);
end;

```

```

case {'casetti','darp','bcasetti'}
    % call prt_cas
    if arg == 1
        prt_cas(results);
    elseif arg == 2
        prt_cas(results,vnames);
    elseif arg == 3
        prt_cas(results,vnames,fid);
    else
        prt_cas(results,[],fid);
    end;

case {'tvp','tvp_garch','tvp_markov'}
    % call prt_tvp
    if arg == 1
        prt_tvp(results);
    elseif arg == 2
        prt_tvp(results,vnames);
    elseif arg == 3
        prt_tvp(results,vnames,fid);
    else
        prt_tvp(results,[],fid);
    end;

case {'garch'}
    % call prt_garch
    if arg == 1
        prt_garch(results);
    elseif arg == 2
        prt_garch(results,vnames);
    elseif arg == 3
        prt_garch(results,vnames,fid);
    else
        prt_garch(results,[],fid);
    end;

case {'hamilton','hamilton_g'}
    % call prt_ham
    if arg == 1
        prt_ham(results);
    elseif arg == 2
        prt_ham(results,vnames);
    elseif arg == 3

```

```

    prt_ham(results,vnames,fid);
    else
    prt_ham(results,[],fid);
    end;
case {'felogit'}
    % call prt_felogit
    if arg == 1
    prt_felogit(results);
    elseif arg == 2
    prt_felogit(results,vnames);
    elseif arg == 3
    prt_felogit(results,vnames,fid);
    else
    prt_felogit(results,[],fid);
    end;

case {'multilogit'}
    arg = 1;
    [vsize junk] = size(vnames); % user may supply a blank argument
    if vsize > 0
    arg = 2;
    end;
    [vsize junk] = size(cnames); % user may supply a blank argument
    if vsize > 0
    arg = 3;
    end;
    [vsize junk] = size(fid); % user may supply a blank argument
    if vsize > 0
    arg = 4;
    end;

    % call prt_multilogit
    if arg == 1
    prt_multilogit(results);
    elseif arg == 2
    prt_multilogit(results,vnames);
    elseif arg == 3
    prt_multilogit(results,vnames,cnames);
    elseif arg == 4
    prt_multilogit(results,vnames,cnames,fid);
    else
    prt_multilogit(results,[],[],fid);
    end;
otherwise

```

```
error('results structure not known by prt function');
```

```
end;
```