



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

ESTUDIO DEL IMPACTO ECONÓMICO DE NIVELES DE SERVICIO
ASOCIADOS A ESPERAS EN CAJAS DE UNA CADENA DE
SUPERMERCADOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL
INDUSTRIAL

PABLO ANDRÉS JEREZ REYES

PROFESOR GUÍA:
MARCELO OLIVARES ACUÑA

MIEMBROS DE LA COMISIÓN:
RICARDO MONTOYA MOREIRA
ANDRÉS MUSALEM SAID

SANTIAGO DE CHILE
2015

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: Pablo Andrés Jerez Reyes
FECHA: Junio 2015
PROFESOR GUÍA: Marcelo Olivares Acuña

ESTUDIO DEL IMPACTO ECONÓMICO DE NIVELES DE SERVICIO ASOCIADOS
A ESPERAS EN CAJAS DE UNA CADENA DE SUPERMERCADOS

Para el estudio del efecto económico de los niveles de servicio, se utilizaron datos de una cadena de supermercados estadounidense, para lo cual se implementaron modelos probabilísticos de duración en tiempo continuo, los cuales definen como variable dependiente el tiempo en días entre transacciones. La estimación de este tiempo, se realizó en función de variables de diferente índole, asociadas a la última compra. Las cuales consideran montos transaccionales, niveles de servicio recibidos por los clientes y variables de control.

Se realizó un análisis de robustez de los modelos respecto a modelos nulos y a la inclusión de tiempos censurados, donde para ambos casos, se obtuvieron resultados consistentes a lo largo de las iteraciones realizadas. Lo que entrega indicios de un buen ajuste de los modelos, y significancia estadística de las variables, respecto al comportamiento de compra de los clientes.

Mediante modelos de teoría de colas, específicamente un modelo M/M/s, fue posible utilizar los impactos estimados por los modelos de duración y realizar una evaluación económica de los efectos negativos de las esperas en base a escenarios. La metodología desarrollada, permite justificar la apertura de una nueva caja, en base a las características particulares del sistema y a los beneficios esperados, generados por la disminución de los tiempos de espera en caja de los clientes.

Los resultados permitieron asignar un valor numérico e interpretable al efecto de variables de servicio, como tiempos de espera y largos de cola. Para las cuales se comprobó empíricamente el efecto negativo de una experiencia de servicio insatisfactoria, lo cual se traduce, en un aumento en el tiempo de retorno a las salas de venta. El efecto numérico de los niveles de servicio, se estimó en base a elasticidades de las tasas de compra, respecto a las variables estudiadas, donde para deterioros de un 1 % en los niveles de servicio promedio entregados, pueden implicar disminuciones en las tasas de compras entre un 1.03 % y un 4.26 % para el caso del total de personas en cola, y entre un 0.33 % y un 1.34 % para los tiempos de espera en cola.

Las metodologías de tratamiento de datos, modelamiento y evaluación económica documentadas en este trabajo, pueden ser aplicadas para la gestión de las salas de venta y sus clientes. De esta forma los niveles de servicio pueden ser utilizados como una herramienta estratégica, y generar ventajas competitivas.

A mis padres

Tabla de Contenido

1. Introducción	1
1.1. Objetivo General	3
1.2. Objetivos Específicos	3
1.3. Descripción del Negocio Estudiado	3
2. Fuentes de Datos	5
2.1. Base Transaccional	5
2.2. Base Registros de Niveles de Servicio	5
2.3. Club de Fidelización	9
2.4. Base Descriptiva de Tiendas	14
3. Introducción a los Modelos de Duración	15
3.1. Definiciones	15
3.1.1. Función de Supervivencia	16
3.1.2. Función de Distribución	16
3.1.3. Función de Verosimilitud	16
3.2. Modelos NBD	17
3.2.1. Modelo Exponencial	17
3.2.2. Modelo Gamma-Exponencial	17
3.2.3. Modelo Gamma-Weibull	18
3.3. Evaluación de los Modelos	18
4. Estimación de Modelos	20
4.1. Preparación de Datos	20
4.1.1. Filtros, Definiciones y Supuestos Iniciales	20
4.1.2. Metodología de Asociación de Niveles de Servicio	22
4.1.3. Descripción del Panel	23
4.2. Estimación de Modelos Nulos	24
4.3. Modelos con Covariables	25
4.3.1. Modelo Exponencial con Covariables	26
4.3.2. Modelo Gamma-Exponencial con Covariables	26
4.3.3. Modelo Gamma-Weibull con Covariables	27
4.4. Variables	27

4.4.1.	VARIABLES TRANSACCIONALES	27
4.4.2.	VARIABLES DE NIVELES DE SERVICIO	28
4.4.3.	VARIABLES DE CONTROL	29
4.4.4.	ESTADÍSTICOS DESCRIPTIVOS	30
4.5.	ESTIMACIÓN DE MODELOS CON COVARIABLES	32
4.5.1.	METODOLOGÍA DE SELECCIÓN DE MODELO	32
4.5.2.	MODELO ESCOGIDO	32
4.5.3.	ANÁLISIS DE RESULTADOS	34
4.6.	COMPARACIÓN DE MODELOS	36
5.	Modelos con Tiempo Censurado	37
5.1.	Tiempo Censurado	37
5.2.	Preparación de Datos	38
5.2.1.	Filtros, Supuestos y Definiciones Adicionales	38
5.2.2.	Descripción del Panel	39
5.3.	Estimación de Modelos Nulos	40
5.4.	Modelos con Covariables	40
5.5.	VARIABLES	42
5.5.1.	ESTADÍSTICOS DESCRIPTIVOS	42
5.6.	ESTIMACIÓN DE MODELOS CON COVARIABLES	44
5.6.1.	MODELO ESCOGIDO	44
5.6.2.	ANÁLISIS DE RESULTADOS	44
5.7.	COMPARACIÓN DE MODELOS	47
6.	Evaluación Económica	48
6.1.	Modelo de Colas	48
6.1.1.	Modelo M/M/s	48
6.1.2.	Elección de Escenarios	49
6.1.3.	Estimación de Tasa de Servicio	49
6.1.4.	Aumento de Dotación	50
6.2.	Evaluación Económica	50
6.2.1.	Cliente y Transacción Promedio	50
6.2.2.	Impacto Económico del Aumento de Dotación	51
7.	Conclusiones	55
7.1.	Limitaciones y Trabajo Futuro	57
	Bibliografía	58
	A. Iteraciones Modelos sin Tiempo Censurado	60
	B. Iteraciones Modelos con Tiempo Censurado	72

Índice de figuras

2.1. Venta Promedio por Hora	6
2.2. Venta Promedio por Día de la Semana	6
2.3. Ejemplo de Registro Total Cajas Abiertas	7
2.4. Ejemplo de Registro Total Clientes en Cola	7
2.5. Promedio de Cajas Regulares Abiertas por Día de la Semana	9
2.6. Promedio de Personas en Cola de Cajas Regulares por Día de la Semana . .	10
2.7. Promedio de Tiempo de Espera en Cajas Regulares por Día de la Semana . .	10
2.8. Promedio de Personas en Cola de Cajas Autoservicio por Día de la Semana .	11
2.9. Promedio de Cajas Regulares Abiertas por Hora	11
2.10. Promedio de Personas en Cola de Cajas Regulares por Hora	12
2.11. Promedio de Tiempo de Espera en Cajas Regulares por Hora	12
2.12. Promedio de Personas en Cola de Cajas Autoservicio por Hora	13
2.13. Histograma de Distancias Únicas en Kilómetros	13
4.1. Histograma de Tiempos entre Compras	24
5.1. Histograma de Frecuencias de Tiempos Censurados	39

Índice de tablas

2.1. Estadísticos Descriptivos de Niveles de Servicio	8
4.1. Estadísticos Descriptivos del Tiempo entre Compras	23
4.2. Modelos Nulos (Sin Covariables)	25
4.3. Fechas que conforman la variable Feriado	29
4.4. Estadísticos Descriptivos	30
4.5. Estadísticos Descriptivos Cajas Regulares	31
4.6. Estadísticos Descriptivos Cajas Autoservicio	31
4.7. Modelos Iteración 10	33
4.8. Elasticidades Modelos Iteración 10	34
4.9. Test Chi-Cuadrado de Ratios de Verosimilitud	36
5.1. Estadísticos Descriptivos del Tiempo Censurado	39
5.2. Modelos Censurados Nulos (Sin Covariables)	40
5.3. Modelos Nulos Actualizados(Sin Covariables)	41
5.4. Estadísticos Descriptivos	42
5.5. Estadísticos Descriptivos Cajas Regulares	43
5.6. Estadísticos Descriptivos Cajas Autoservicio	43
5.7. Modelos Tiempo Censurado Iteración 5	45
5.8. Elasticidades Modelos Tiempo Censurado Iteración 5	45
5.9. Test Chi-Cuadrado de Ratios de Verosimilitud	47
6.1. Escenarios Sistemas de Colas	50
6.2. Largos de Cola Promedio Estimados ante Aumento en Dotación	50
6.3. Coeficientes Asociados a Largos de Cola	51
6.4. Cambio Porcentual y Nueva Tasa Mensual Modelo 1	51
6.5. Cambio Porcentual y Nueva Tasa Mensual Modelo 2	52
6.6. Beneficio Económico por Cliente	52
6.7. Beneficio Total Neto por Escenario - Modelo 1	53
6.8. Beneficio Total Neto por Escenario - Modelo 2	53
6.9. Utilización por Escenario	54
A.1. Modelos Iteración 1	61
A.2. Modelos Iteración 2	62

A.3. Modelos Iteración 3	63
A.4. Modelos Iteración 4	64
A.5. Modelos Iteración 5	65
A.6. Modelos Iteración 6	66
A.7. Modelos Iteración 7	67
A.8. Modelos Iteración 8	68
A.9. Modelos Iteración 9	69
A.10. Modelos Iteración 11	70
A.11. Modelos Iteración 12	71
B.1. Modelos Tiempo Censurado Iteración 1	73
B.2. Modelos Tiempo Censurado Iteración 2	74
B.3. Modelos Tiempo Censurado Iteración 3	75
B.4. Modelos Tiempo Censurado Iteración 4	76

Capítulo 1

Introducción

Los tiempos de espera son un factor competitivo dentro del retail en general, donde las esperas y filas, implican una percepción negativa del servicio recibido por los clientes. En formatos como supermercados, donde la espera principal se encuentra al final de la experiencia de servicio, el efecto negativo asociado a las esperas en cajas, puede implicar una baja en la percepción de calidad, pese a los esfuerzos por mantener altos estándares de atención y venta dentro de la tienda (Kumar 2005). Mejorar los estándares de servicio, tiene un impacto en toda la competencia, ya que altera el punto de referencia de los clientes, los cuales se vuelven más exigentes (Mehta 2013). Por ello, la gestión de niveles de servicio más que ser una política de calidad para la empresa, es una herramienta para generar ventajas competitivas.

En la literatura se encuentran varios estudios que relacionan empíricamente niveles de servicio y las percepciones de los clientes, los que por lo general consisten en experimentos que combinan registros de video y/o contadores de tráfico, con encuestas de satisfacción. La disponibilidad de información es un aspecto clave en este tipo de estudios, ya que generar datos objetivos de niveles de servicio y percepciones de los clientes es costoso, más aun cuando se requiera realizar un análisis masivo de múltiples tiendas y largos intervalos temporales (Kumar 2005).

Aspectos visuales, como aglomeraciones y largos de cola, influyen en la percepción del servicio, debido a que los clientes utilizan esta información para proyectar sus tiempos de espera (Mehta 2013). Factores asociados a experiencias previas, tanto en la tienda como en la competencia, pueden influir la percepción del servicio, donde la satisfacción final, es afectada por las expectativas de espera de cada cliente (Kumar 2005).

Nuevas tecnologías en registros de video y reconocimiento de imágenes, han facilitado el desarrollo de aplicaciones para el monitoreo y gestión de salas de venta de retail. Estas aplicaciones, generan, almacenan y analizan información a gran escala de niveles de servicio para apoyar la toma de decisiones, a nivel de tienda y de cadena. Este estudio combina datos generados con aplicaciones de este tipo, con información transaccional de las ventas, lo cual tiene por objetivo analizar empíricamente efectos de los niveles de servicio sobre los

comportamientos de compra de los clientes. Se espera que una experiencia satisfactoria o insatisfactoria, en términos de servicio recibido, refleje cambios en los patrones habituales de compra de cada cliente.

Para este estudio se utilizaron datos de una cadena de supermercado, la cual posee una visión de calidad de servicio en múltiples aspectos. En lo que respecta a niveles de servicio en cajas, se tiene una noción del efecto negativo de las esperas, por lo que existen políticas asociadas a minimizar las esperas de los clientes, mediante cajas especializadas como las autoserivicio y para baja cantidad de productos.

Gran parte de la oferta de valor de la empresa está asociada a un programa de membresía, el cual da acceso a descuentos, cupones y ofertas a sus clientes mediante el uso de una tarjeta de identificación en sus compras. Estos programas tienen por objetivo fidelizar a los clientes mediante beneficios, pero también permiten asociar información transaccional a clientes particulares a lo largo del tiempo. Este aspecto es crucial en este estudio, ya que permite realizar un seguimiento de las compras realizadas a nivel de cliente asociando las compras con el identificador del programa.

Los niveles de servicio abordados en este estudio, son los asociados a las esperas y largos de cola en cajas de supermercados. El principal factor innovador del trabajo, consiste en la asociación de las mediciones de estos niveles a cada transacción, es decir, se analizarán las ventas en el tiempo a nivel de cliente, lo que permitirá concluir su efectos empíricos de las esperas ante cambios en los patrones habituales de compra de los clientes, ya que se espera que el cliente demuestre su satisfacción con el servicio aumentando o disminuyendo su nivel de compras.

Los primeros capítulos de este trabajo, se enfocan en el tratamiento de datos necesario para la estimación de impactos mediante modelos de duración, para lo cual es necesario realizar una serie de definiciones, supuestos y filtros. Bajo esta lógica se logró consolidar datos de fuentes transaccionales y de registros de video, de forma tal que se puede aproximar los niveles de servicio recibidos por los clientes al momento de cada transacción. Los siguientes capítulos analizan los efectos y significancias de las variables estudiadas, mediante modelos de duración, para lo cual se realiza una comparación con modelos nulos además de la inclusión de tiempos censurados. Esto permitió encontrar efectos significativos en las variables de servicio y encontrar evidencia empírica de su efecto negativo sobre el comportamiento de compra de los clientes. Finalmente se utilizaron modelos de teoría de colas, para estimar variaciones en los largos de cola producidos por aumentos de dotación, de esta forma se logró realizar una metodología de evaluación costo-beneficio de la apertura de nuevas cajas, respecto al beneficio esperado en compras, generados por la disminución de las esperas. Los resultados obtenidos, permiten justificar al aumento de dotación ante situaciones particulares de cada sistema.

1.1. Objetivo General

- El objetivo general de este trabajo es cuantificar económicamente, en términos monetarios, impactos de los niveles de servicio asociados a esperas en cajas de supermercados, con el fin de generar herramientas de apoyo a la gestión de dotación en tiendas.

1.2. Objetivos Específicos

Se identifican los siguientes objetivos específicos:

- Elaborar una metodología de tratamiento de datos, para la construcción de un panel de clientes, que permita realizar un seguimiento y estudio del comportamiento de compra de este.
- Estudiar los impactos de los niveles de servicio sobre el comportamiento de compra de los clientes mediante modelos de duración.
- Evaluar económicamente medidas sobre dotación para gestionar los niveles de servicio asociados a esperas en cajas de supermercados.

1.3. Descripción del Negocio Estudiado

El estudio se realizó sobre una cadena de supermercados estadounidense, presente en 8 estados de la costa este del país. Su formato de tienda más común es del tipo vecinal, que apunta a clientes que realizan varias visitas en el mes e incluso semanales. La empresa ha implementado varias iniciativas para integrarse con las comunidades y localidades donde está presente, como por ejemplo, alianzas con mercados locales, como agricultores y criadores de animales, con el fin de posicionar sus productos en el mercado. Su propuesta de valor incluye una serie de productos y servicios entre los cuales destacan los siguientes:

- Club de fidelización: Programa de membresía, el cual permite acceder a una serie de beneficios, descuentos y servicios preferenciales disponibles solo para clientes los registrados. El programa asigna un identificador único y una tarjeta física la cual permite identificar al cliente en cada una de sus compras.
- Tiendas 24 Horas: Ciertas tiendas de la cadena ofrecen un horario de atención continuado, las 24 horas del día. Este servicio se encuentra disponible en la mayoría de las salas de venta de la cadena.
- Farmacia: Ciertas tiendas de la cadena poseen farmacias al interior de las salas de venta. Estas pertenecen a la cadena, y están asociadas a la misma marca. En ellas se ofrecen los servicios estándares del rubro, además de ciertos beneficios exclusivos para los clientes del club de fidelización, como por ejemplo tiempos preferenciales en la elaboración de fármacos.

- Self-Checkout: Corresponden a cajas autoservicio, donde los clientes registran y empaacan sus propios productos. Estas están supervisadas por personal de la tienda que se encarga de ayudar a las personas que presenten problemas en sus compras o que necesiten instrucciones para su utilización. Este tipo de cajas también permite la identificación del los cliente pertenecientes al programa de membresía para la aplicación de los beneficios.
- Servicios de Comida: Dentro de las tiendas existen secciones de comida rápida, las cuales ofrecen promociones especiales para los clientes pertenecientes al club de fidelización.

En síntesis, la mayoría de la oferta de valor de la empresa está asociada al mismo programa, por lo que se pretenda abordar múltiples necesidades del cliente, y fomentar la identificación de estos en los sistemas transaccionales. Esto hace que la empresa sea buen candidato para un estudio de minería de datos.

La información descrita anteriormente, es relevante al momento de trabajar los datos de la cadena, y fue abordada en el modelamiento estadístico mediante filtros, variables y exclusiones. Los cuales se encuentran descritos en el Capítulo de Estimación de Modelos.

Capítulo 2

Fuentes de Datos

2.1. Base Transaccional

Corresponde a todos a los registros transaccionales, de los puntos de venta de las diferentes tiendas de la cadena de supermercados. En particular se utilizaron datos de las cajas finales, también denominadas *Checkout*, para las cuales se cuenta información detallada asociada cada compra. Los datos de encuentran indexados mediante el identificador único de cada cliente en del club de membresía.

Se registran 94,864,603 compras realizadas en 215 tiendas entre los meses de Diciembre 2012 y Agosto 2013, donde el 86 % de las transacciones se encuentra asociadas a una cliente mediante su número identificador. En el periodo de observación, cada cliente realiza en promedio 11.9 visitas, en 1.1 tiendas y por un monto de 29.483 dólares.

La Figura 2.1 ilustra la venta total promedio por hora, en la cual se observan dos horarios punta, uno cercano a las 12:00 horas y otro a las 17:00. Las ventas nocturnas son considerablemente inferiores a las registradas durante el día. La Figura 2.2, por su parte, ilustra la venta total promedio por día, donde se aprecia una mayor venta los fines de semana y los días jueves.

2.2. Base Registros de Niveles de Servicio

Esta fuente posee información de niveles de servicio, generados con tecnologías de reconocimiento de imágenes y registros de video para 32 tiendas de la cadena de supermercados. Los datos se encuentran indexados por tienda y caja, y se identifican los siguientes cuatro tipos de mediciones asociados a esperas en cajas:

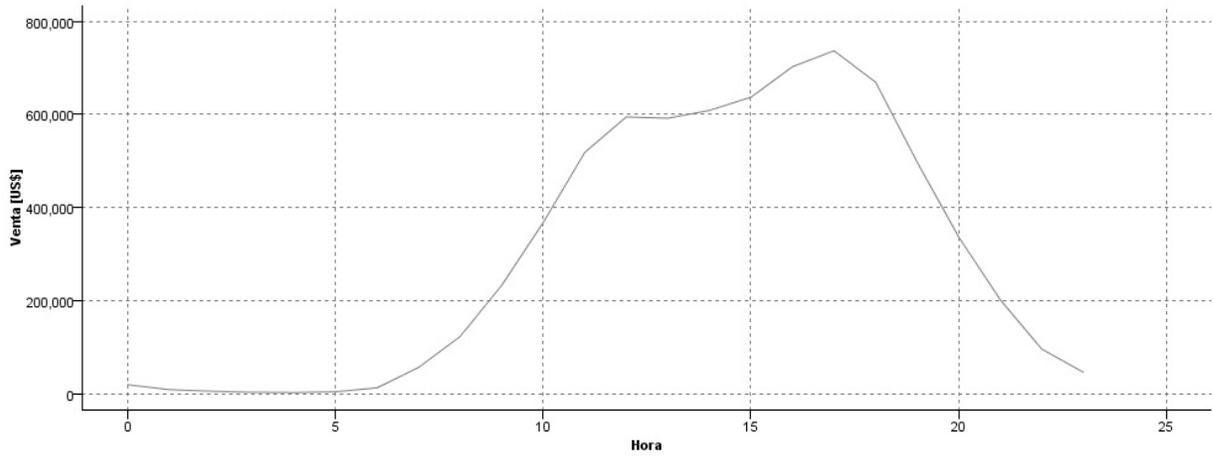


Figura 2.1: Venta Promedio por Hora

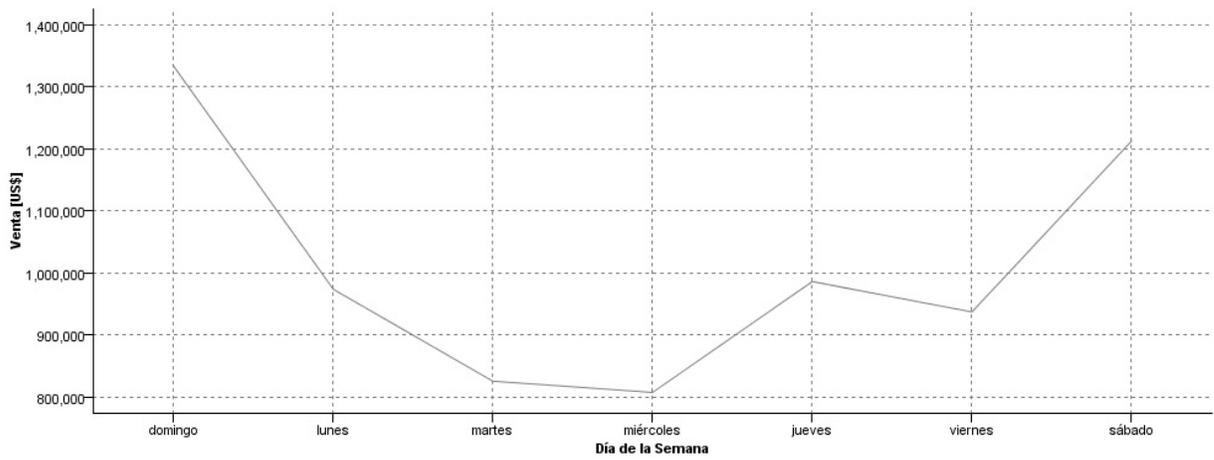


Figura 2.2: Venta Promedio por Día de la Semana

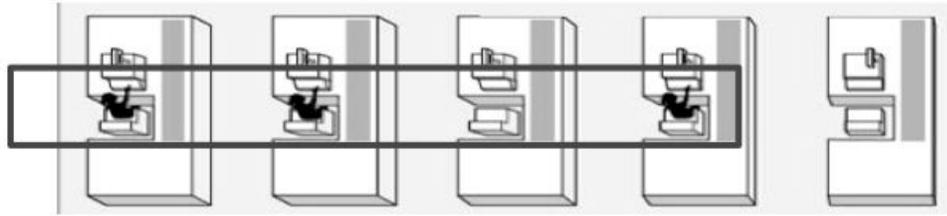


Figura 2.3: Ejemplo de Registro Total Cajas Abiertas

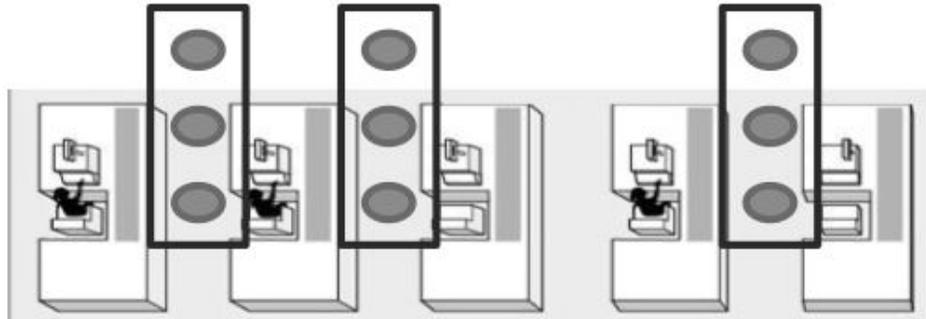


Figura 2.4: Ejemplo de Registro Total Clientes en Cola

1. **Total de cajas regulares abiertas:** Contabiliza la dotación total de cajas regulares habilitadas, para un tiempo específico. La Figura 2.3 ejemplifica una medición en la cual se presentan tres cajas regulares disponibles, para el registro de productos y pago por parte de los clientes.
2. **Total de clientes en colas de cajas regulares:** Contabiliza el total de clientes en cola de cajas regulares, para un tiempo específico, es decir, corresponde a la suma todas las filas de cajas regulares habilitadas. La Figura 2.4 ejemplifica un registro de 9 personas en cola en la tienda.
3. **Tiempo de espera en colas regulares:** Contabiliza el tiempo de espera en segundos de clientes en cajas regulares. La metodología de medición contabiliza el tiempo desde que el cliente se detiene frente a la fila de una caja regular, hasta que registra el primer producto, o entrega su credencial para asociar la compra con el club de membresía.
4. **Total de clientes en colas autoservicio:** Es una medición análoga a la que contabiliza clientes en cajas regulares, pero para las cajas del tipo autoservicio.

La Tabla 2.1 contiene los estadísticos descriptivos de los niveles de servicio descritos anteriormente. Como se observa en los valores medios, las esperas de los clientes asociadas a las cajas poseen valores bajos, confirmando la declaración de calidad de servicio de la empresa estudiada.

Medición	Min	Max	Media	Desv. Est.
Total Personas en Cola Checkout	0	37	5.669	4.102
Total Cajas Abiertas	0	22	4.424	2.184
Total Personas en Cola SelfCheckout	0	12	2.19	1.797
Tiempo de Espera en Cola (Seg)	1	631	30.597	46.043

Tabla 2.1: Estadísticos Descriptivos de Niveles de Servicio

Por la naturaleza de la metodología con la que se recopilan las mediciones, existen asociaciones en los registros, como es el caso de las mediciones del **Total de cajas regulares abiertas** y **Total de clientes en colas de cajas regulares**, las cuales se generan a partir de la misma captura del estado de las cajas regulares en la tienda. Las mediciones de **Tiempo de espera en colas regulares** corresponden a un tipo de medición diferente, por lo que posee tiempos de registro independientes. Estas diferencias también se ven reflejadas en la periodicidad de los registros diarios, donde las dos primera mediciones mencionadas, presentan 26.4 registros diarios en promedio. Por su parte las mediciones de los tiempos de espera, sólo presentan 3.3 mediciones diarias en promedio. Esto se debe en parte al esfuerzo requerido para transformar un registro de video en dato numérico, ya que para los tiempos de espera es necesario observar fragmentos completos de video, identificar eventos específicos y contabilizar el tiempo transcurrido.

Las mediciones del **Total de clientes en colas autoservicio** son independientes y análogas a las de cajas regulares por ello también poseen una alta periodicidad, presentando 28.9 registros diarios en promedio. Un punto a considerar, es la disponibilidad de este tipo de medición en las tiendas estudiadas, ya que sólo 30 de las 32 tiendas poseen registros de niveles de servicio en cajas autoservicio.

Al graficar los niveles de servicio promedio, por día de la semana (Figura 2.5, 2.6, 2.7 y 2.8), se aprecia una relación directa con el patron de ventas observado en la Figura 2.2, donde las mediciones promedio reflejan el comportamiento esperado ante un mayor flujo de clientes, es decir un aumento en el número de cajas habilitadas, tiempos esperas y largos de cola. Se observa además, el mismo patron en los valores máximos, con niveles superiores los fines de semana y días jueves.

Los largos de cola de las cajas autoservicio, tienen un comportamiento ligeramente diferente a las regulares, con un niveles promedio más altos desde jueves a lunes como de aprecia en la Figura 2.8. El rango de los niveles promedio es más estable que las cajas regulares, presentando valores entre 2.05 y 2.36. Rango comparativamente menor al de las cajas regulares, las cuales presentan valores entre 4.68 y 7.30, lo que da indicios de un comportamiento más estable del nivel de servicio a lo largo de la semana.

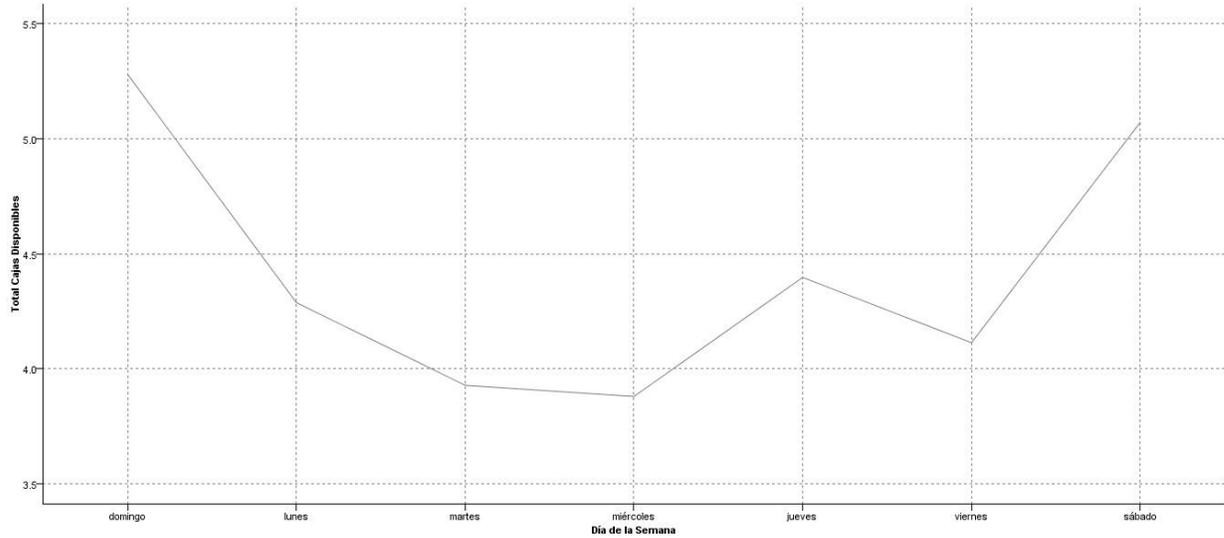


Figura 2.5: Promedio de Cajas Regulares Abiertas por Día de la Semana

Al graficar los niveles de servicio promedio por hora del día (Figura 2.9, 2.10, 2.11 y 2.12,) se mantiene la relación con el patron de ventas por hora ilustrado en la Figura 2.1, para los niveles de servicio asociados largos de cola y cajas habilitadas. Los tiempos de espera, reflejan un patrón diferente y aleatorio, el cual puede ser explicado por la baja periodicidad de las mediciones por día, particularmente en los horarios extremos, lo que impiden realizar una comparación respecto a las ventas.

2.3. Club de Fidelización

El club de beneficios, posee información que permite aproximar la localización del los hogares de los clientes, mediante códigos Zip los cuales se encuentran asociados a cada identificador único de cliente. Estos datos, en conjunto con la localización de las tiendas, permiten generar una variable demográfica de distancia entre el hogar y las tiendas. Para esto se realizó una transformación de las direcciones de los clientes a coordenadas mediante una API de Google Maps, y luego se calculó la distancia a las coordenadas de la tienda. La Figura ilustra las distancias únicas de los hogares a tiendas en las cuales se realizaron transacciones. La API permite un número limitado de consultas, por ello el gráfico representa sólo las combinaciones únicas de coordenadas, de todas formas se aprecia una tendencia a distancias cortas entre los hogares y tiendas, además de casos extremos con distancias superiores a los 4,000 kilómetros.

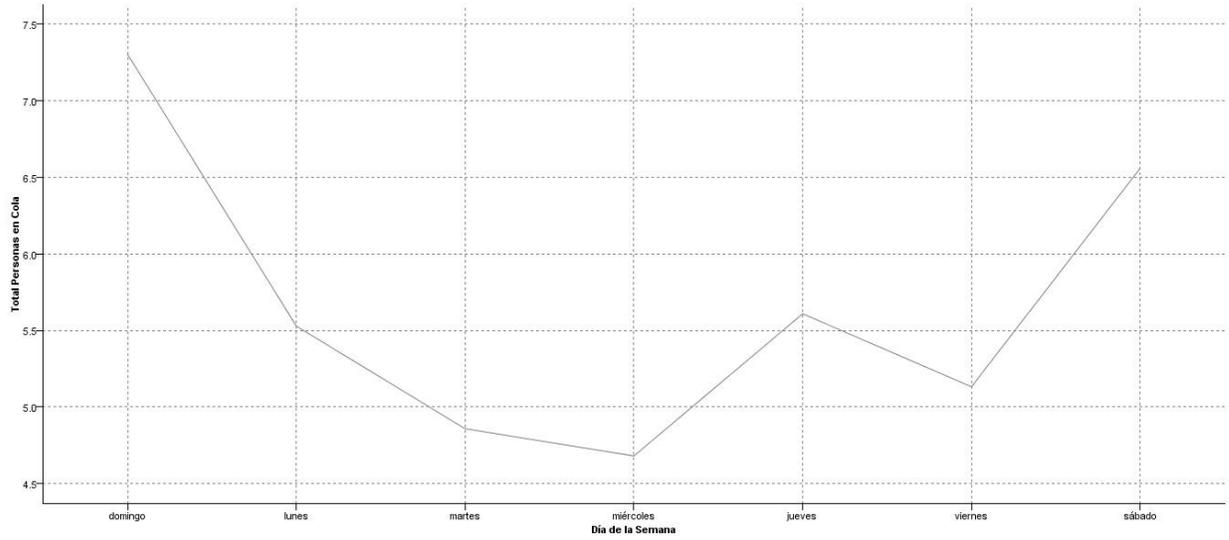


Figura 2.6: Promedio de Personas en Cola de Cajas Regulares por Día de la Semana

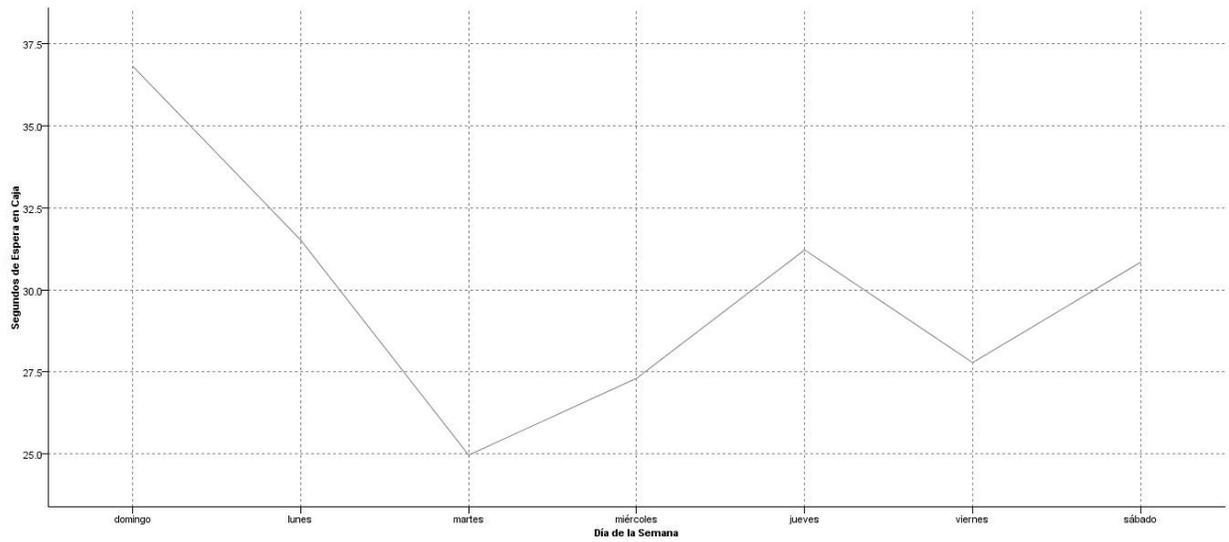


Figura 2.7: Promedio de Tiempo de Espera en Cajas Regulares por Día de la Semana

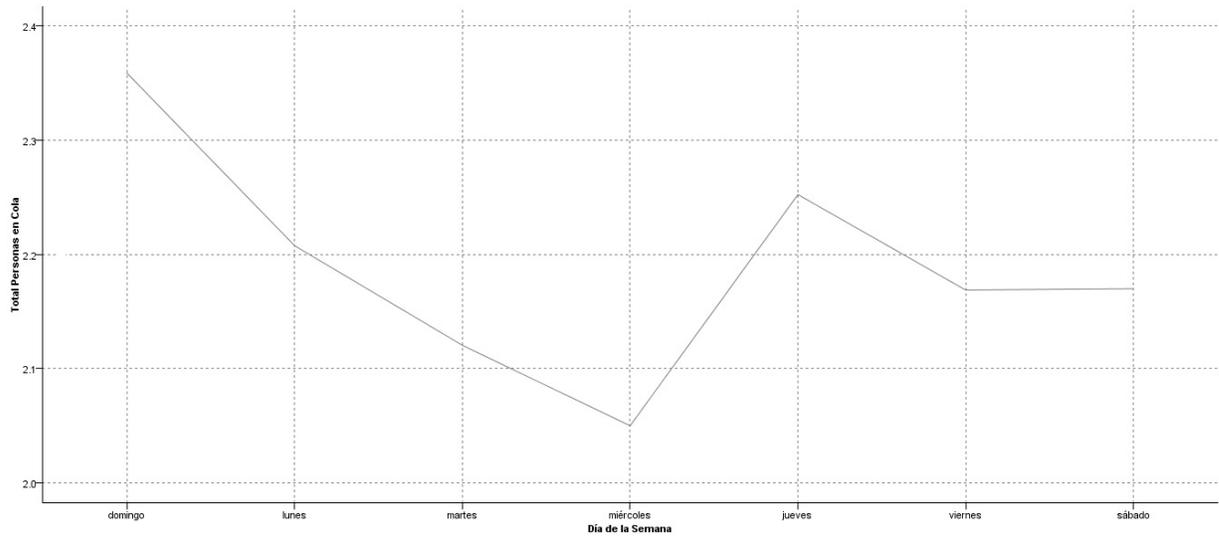


Figura 2.8: Promedio de Personas en Cola de Cajas Autoservicio por Día de la Semana

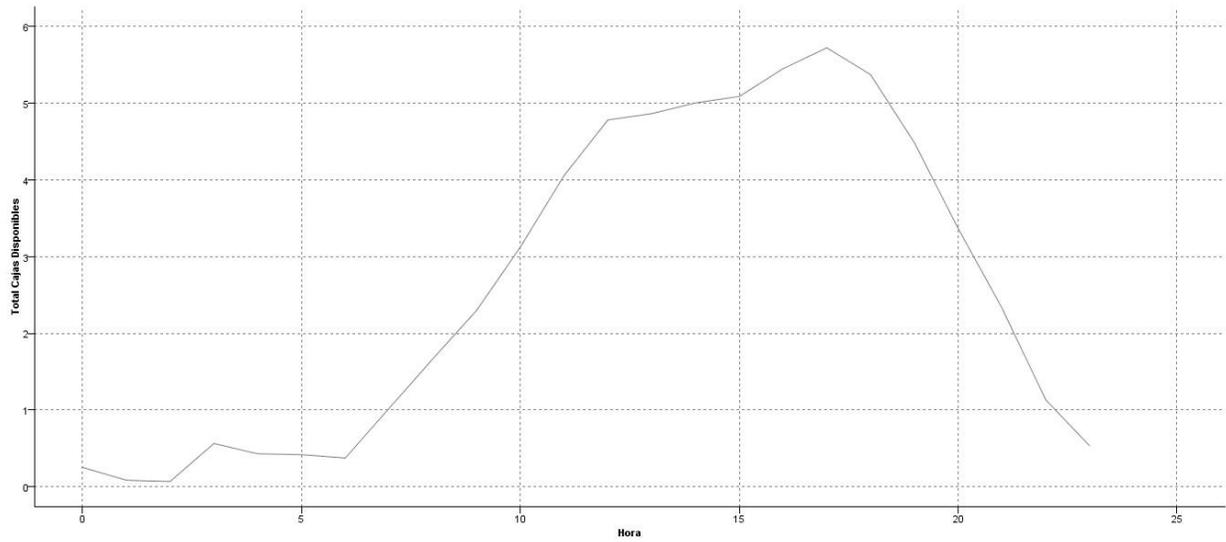


Figura 2.9: Promedio de Cajas Regulares Abiertas por Hora

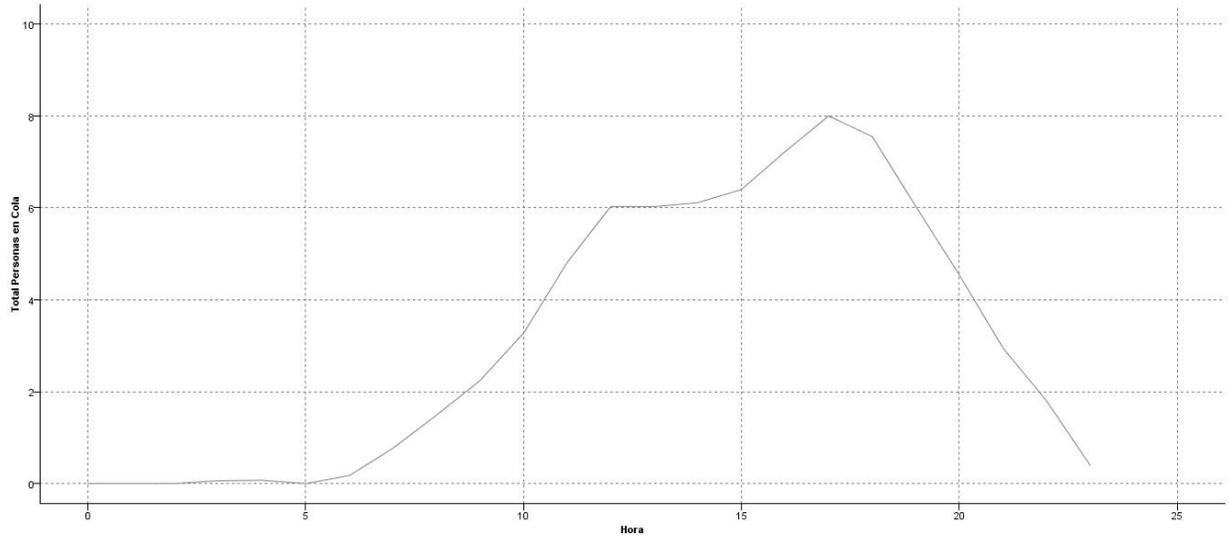


Figura 2.10: Promedio de Personas en Cola de Cajas Regulares por Hora

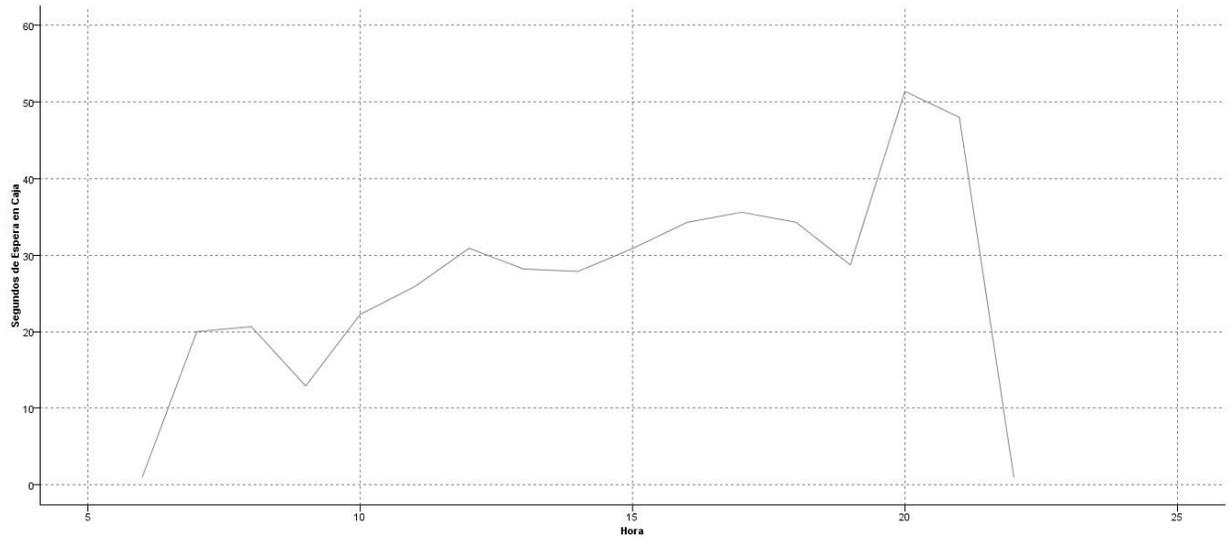


Figura 2.11: Promedio de Tiempo de Espera en Cajas Regulares por Hora

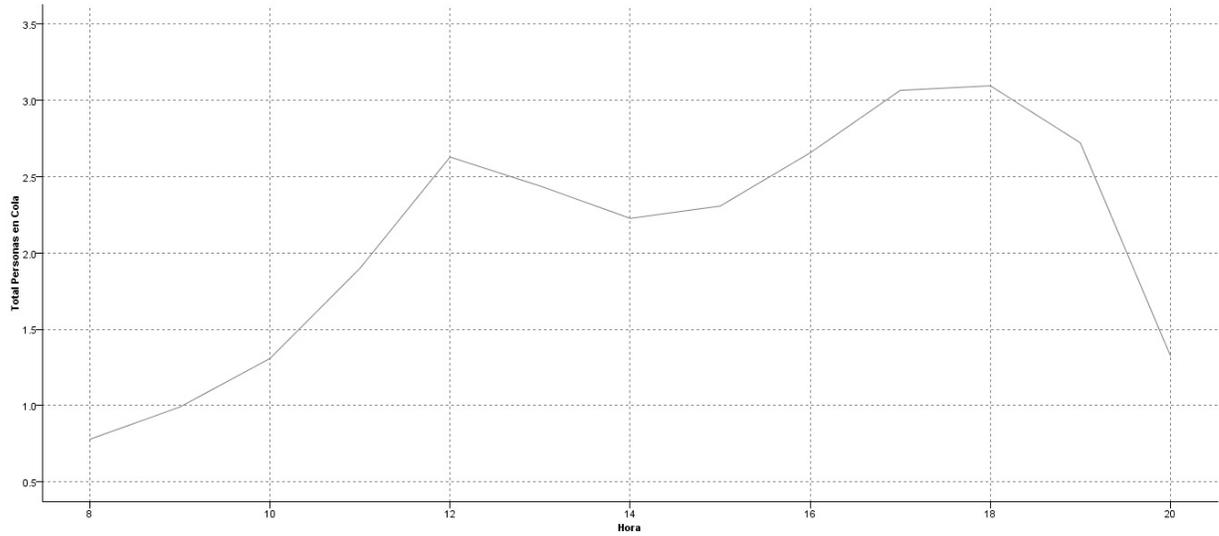


Figura 2.12: Promedio de Personas en Cola de Cajas Autoservicio por Hora

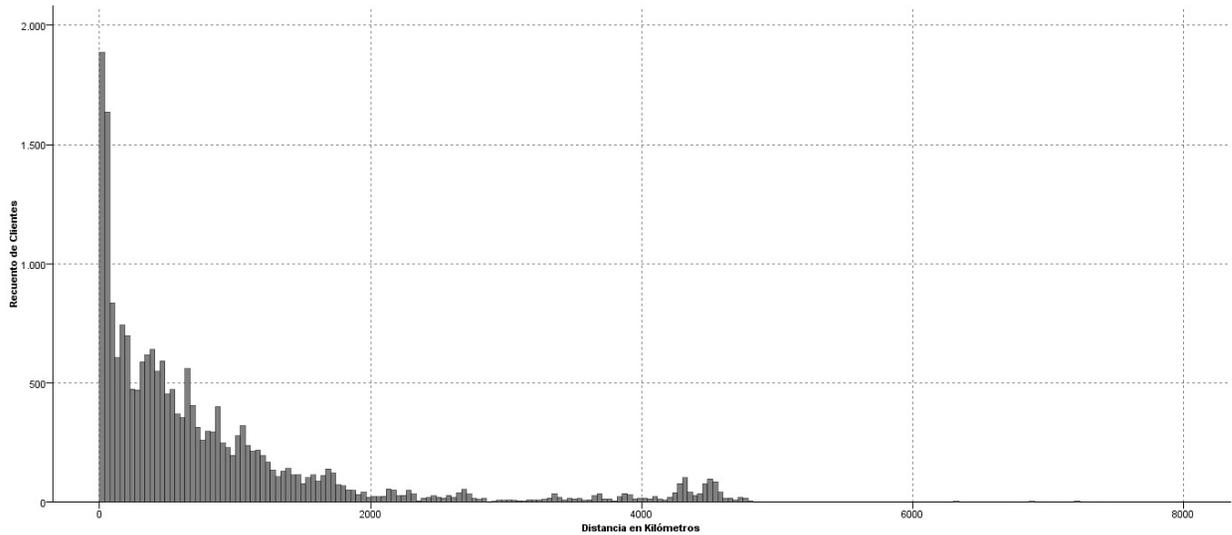


Figura 2.13: Histograma de Distancias Únicas en Kilómetros

2.4. Base Descriptiva de Tiendas

Posee información asociada a los productos y servicios disponibles en cada tienda, además de los horarios de atención y la ubicación geográfica. Se utilizó principalmente para identificar la disponibilidad de cajas autoservicio y la disponibilidad del servicio de atención 24 horas.

Capítulo 3

Introducción a los Modelos de Duración

También conocidos como modelos de supervivencia, son herramientas estadísticas que utilizan distribuciones matemáticas e información histórica para estimar el tiempo que toma la ocurrencia de un evento en el futuro. Herramientas de este tipo son ampliamente utilizadas en distintas disciplinas como Medicina, Ciencias Sociales y Marketing (Fader y Hardie 2009) (Zaman y Pfeiffer 2011). Dentro de esta última se han aprovechado los avances tecnológicos en almacenamiento y procesamiento de datos, para estudiar comportamientos de negocios a nivel transaccional, permitiendo desarrollar estimaciones tanto a nivel agregado como a nivel de clientes individuales. Esto ha permitido dotar a los negocios de mayor información, lo que permite implementar políticas y gestionar las carteras de clientes a niveles cada vez más personalizados.

Estos modelos poseen una naturaleza probabilística, la cual asume que el comportamiento observado se encuentra caracterizado por un proceso estocástico, el cual asume que información de comportamientos previos, no es un reflejo exacto del comportamiento futuro, y da la posibilidad a variaciones con una respectiva probabilidad asociada (Fader y Hardie 2009). Por ejemplo, si en el pasado un evento demoró dos semanas en ocurrir, los modelos consideran posible que el mismo evento en el futuro demore una o tres semanas, asociando una probabilidad de ocurrencia para cada caso.

3.1. Definiciones

Las distribuciones de probabilidad permiten un análisis de distintas perspectivas relativas al tiempo de ocurrencia del evento. A continuación se definen y describen funciones asociadas a modelos de duración, relevantes para este estudio.

3.1.1. Función de Supervivencia

Se define como la probabilidad que un evento, caracterizado por una función de densidad de probabilidad $f(x)$, ocurra después de un determinado tiempo T .

$$S(t) = P(T > t) = \int_t^{\infty} f(u)du$$

3.1.2. Función de Distribución

El la función complementaria a la de Supervivencia, y se define la probabilidad de que un evento caracterizado por una función de densidad de probabilidad $f(x)$, ocurra en un tiempo menor o igual a T .

$$F(t) = P(T \leq t) = 1 - S(t)$$

3.1.3. Función de Verosimilitud

Corresponde a la función de densidad de probabilidad conjunta de muestras independientes, utilizada en estadística, para comparar estimaciones de parámetros de las funciones. Para el caso de una función continua, caracterizada por un parámetro θ y n muestras independientes, la función de verosimilitud puede ser expresada de la siguiente forma:

$$L = \prod_{i=1}^n f(x_i|\theta)$$

Esta función permite comparar dos estimadores y evaluar el mejor ajuste a las muestras, caracterizado por un mayor valor en la función de verosimilitud.

La estimación de parámetros por lo general se realiza sobre el logaritmo de esta función, la cual se denomina Log-Verosimilitud. Esta función posee la propiedad de crecimiento monótono, lo cual disminuye la complejidad de los problemas de optimización, asociados a la estimación de parámetros.

$$LL = \sum_{i=1}^n f(x_i|\theta)$$

3.2. Modelos NBD

Introducidos por Ehrenberg (1959) y desarrollados por Morrison y Schmittlein (1980), utilizan mixturas de distribuciones de probabilidad para modelar heterogeneidad y dependencia en la duración. Gupta (1991) introduce variables explicativas a estos modelos, permitiendo el estudio de su impacto sobre la tasa de eventos. La principal ventaja principales de estos modelos son la forma cerrada de la función de verosimilitud, lo que permite estimar los parámetros mediante métodos de optimización.

Desde el punto de vista del problema que aborda este estudio, los tiempos entre transacciones pueden modelados como una variable dependiente en función de información relativa al comportamiento de compra de los clientes y los niveles de servicio recibidos. La utilización de estos modelos permite la incorporación de variables asociadas a compras anteriores para explicar cambios en los tiempos de retorno a las tiendas. Las siguientes secciones describen las especificaciones de los modelos, en función de las mixturas de distribuciones utilizadas.

3.2.1. Modelo Exponencial

Corresponde a la especificación más simple, la cual asume un tiempo entre compras t distribuido exponencialmente con tasa λ . La función de distribución de probabilidad de la distribución exponencial es la siguiente:

$$f(t|\lambda) = \lambda e^{-\lambda t}$$

A partir de esta, se obtiene la función de distribución acumulada para el tiempo, relativa al parámetro λ .

$$F(t|\lambda) = 1 - e^{-\lambda t}$$

La logverosimilitud, para un cliente i , que registra V_i visitas se puede expresar de la siguiente forma:

$$LL = \sum_{i=1}^n \sum_{j=1}^{V_i} (\text{Log}(\lambda) - \lambda t_{ij})$$

3.2.2. Modelo Gamma-Exponencial

Este modelo introduce heterogeneidad no observada en la especificación, asumiendo una distribución *Gamma* en la tasa λ .

$$g(\lambda|\alpha, r) = \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}$$

La función de densidad acumulada es la siguiente:

$$F(t|\alpha, r) = \int_0^\infty F(t|\lambda)g(\lambda)d\lambda = 1 - \left(\frac{\alpha}{\alpha + t}\right)^r$$

La función de distribución de probabilidad se puede calcular a partir de la función de densidad acumulada.

$$f(t|\alpha, r) = \frac{\partial F(t|\alpha, r)}{\partial t} = \frac{r\alpha^r}{(\alpha + t)^{r+1}}$$

Luego la logverosimilitud para un cliente i , que registra V_i visitas se puede expresar de la siguiente forma:

$$LL = \sum_{i=1}^n \sum_{j=1}^{V_i} (\text{Log}(r) + r\text{Log}(\alpha) - (r + 1)\text{Log}(\alpha + t_{ij}))$$

3.2.3. Modelo Gamma-Weibull

Este modelo introduce dependencia en la duración, el cual asume que el tiempo entre compras t , posee una distribución *Weibull* de parámetros λ y c .

$$f(t|\lambda, c) = c\lambda t^{c-1}$$

$$F(t|\lambda, c) = 1 - e^{-\lambda t^c}$$

La función de densidad acumulada es la siguiente:

$$F(t|\alpha, r, c) = \int_0^\infty F(t|\lambda, c)g(\lambda)d\lambda = 1 - \alpha^r(\alpha + t^c)^{-r}$$

La función de distribución de probabilidad se puede calcular a partir de la función de densidad acumulada.

$$f(t|\alpha, r, c) = \frac{\partial F(t|\alpha, c, r)}{\partial t} = \frac{r\alpha^r c t^{c-1}}{(\alpha + t^c)^{r+1}}$$

Luego la logverosimilitud para un cliente i , que registra V_i visitas se puede expresar de la siguiente forma:

$$LL = \sum_{i=1}^n \sum_{j=1}^{V_i} (\text{Log}(r) + r\text{Log}(\alpha) + \text{Log}(c) + (c - 1)\text{Log}(t_{ij}) - (r + 1)\text{Log}(\alpha + t_{ij}^c))$$

3.3. Evaluación de los Modelos

Los modelos propuestos estiman los parámetros en función de una maximización de la función de Log-Verosimilitud. Para realizar comparaciones robustas entre modelos es necesario definir indicadores complementarios de bondad de ajuste además de test estadísticos.

Criterio de Información de Akaike (AIC)

Es un indicador de calidad relativa, el cual penaliza por la inclusión de parámetros en los modelos, con la finalidad de controlar el sobreajuste. Para un modelo con Log-Verosimilitud LL , estimado a partir de k parámetros, el Criterio de Información de Akaike se calcula de la siguiente forma:

$$2k - 2LL$$

Criterio de Información Bayesiano (BIC)

Es similar al criterio AIC, ya que considera el sobreajuste generado por la inclusión de parámetros además del producido por los tamaños muestrales. Para un modelo con Log-Verosimilitud LL , estimado a partir de k parámetros y una muestra de tamaño n , el Criterio de Información Bayesiano se calcula de la siguiente forma:

$$k \ln(n) - 2LL$$

Ratio de Verosimilitud (LR)

Este estadístico, se utiliza para comparar dos modelos en función de sus Log-Verosimilitudes y la cantidad de parámetros incluidos.

$$LR = 2(LL_A - LL_B)$$

Luego es estadístico se comprara con una distribución χ^2 de k grados de libertad, donde k es la cantidad de parámetros adicionales del modelo A por sobre B. Si $LR > \chi^2(0,05, k)$ se concluye que el modelo A es mejor B.

Capítulo 4

Estimación de Modelos

4.1. Preparación de Datos

4.1.1. Filtros, Definiciones y Supuestos Iniciales

Para la estimación de los modelos, es necesario elaborar un panel datos que permita estudiar el comportamiento de compra de los clientes a lo largo del tiempo mediante las transacciones realizadas en las tiendas de supermercado. El resto de la información a estudiar debe ser consolidada y asociada a transacciones puntuales, de forma que capture de forma más certera las compras realizadas por cada cliente. Con este objetivo presente y en función de características de cada fuente de datos, se realizaron supuestos, filtros y transformaciones para la consolidación de la información. A continuación se describen los filtros y supuestos realizados para la construcción del panel.

- **Definición 1: Universo de Tiendas de Estudio**

La base de niveles de servicio posee información sólo para 30 de las tiendas registradas en la base transaccional. Por lo tanto el universo de tiendas del estudio se define para estas tiendas, ya que en ellas se podrá asociar el servicio recibido a las compras de cada cliente.

- **Definición 2: Universo de Clientes de Estudio**

El comportamiento de compra de un cliente sólo puede ser estudiado si se pueden identificar sus compras, por lo que el universo de clientes de estudio se define como los clientes que presenten compras en el universo de tiendas de estudio, los cuales posean compras identificadas mediante la tarjeta del club de fidelización.

- **Definición 3: Tiempo entre Transacciones a Nivel de Cadena**

La definición del tiempo entre transacciones a nivel de cadena permitirá estudiar el valor del cliente en términos globales para la empresa, independiente de la sala de venta en la

cual realice sus transacciones. En consecuencia, la variable dependiente de los modelos de duración, la cual es el tiempo entre transacciones, se define como el tiempo transcurrido entre dos compras realizadas en cualquiera de las tiendas de la cadena de supermercados.

- **Filtro 1: Compras Fuera del Universo de Tiendas de Estudio**

En función de la definición de tiempo entre transacciones, se filtran del panel de clientes todos aquellos que posean compras en alguna tienda diferente a las 30 que conforman el universo de estudio. Esto permitirá mitigar sesgos en el análisis de la variable temporal, generados por las tiendas filtradas.

- **Supuesto 1: Compras Nocturnas son Diferentes a las realizadas en Horario Regular**

Las mediciones de niveles de servicio solo se realizan durante el día, por lo que no es posible asociarlos a transacciones realizadas durante la noche, por ello y con el mismo objetivo de la **Definición 1**, lo ideal sería poder prescindir de los datos de compras nocturnas sin que la exclusión de estos afecte las conclusiones de los modelos, para ello se plantea el supuesto de que las compras en horario nocturno son diferentes a las realizadas en horario regular, lo que permitiría incluir el horario de la compra en la definición del tiempo entre transacciones.

Para comprobar la validez de este supuesto se realizó una segmentación mediante K-Medias de los montos y cantidad de ítem promedio por hora, donde se espera que los valores promedio representen el tipo de compra promedio de cada horario. La técnica clasificó dos segmentos horarios correlativos, los cuales pueden ser interpretados como un horario regular (9:00 a 21:59) y uno nocturno (22:00 a 8:59). Esto permite concluir que en promedio las compras en horario regular son diferentes a las nocturnas.

- **Filtro 2: Compras Nocturnas**

En función del **Supuesto 1** se filtran todas las compras realizadas entre las 22:00 y 08:59 horas del panel de clientes.

■ **Definición 4: Unidad Temporal del Análisis**

Se define la unidad temporal en días, ya que el 90.36 % de los clientes realiza como máximo una compra diaria en intervalo temporal estudiado.

■ **Filtro 3: Compras Diarias por Cliente**

En función de la **Definición 4**, se filtran del panel de datos todos los clientes que registren más de una compra en la cadena, para un día cualquiera del intervalo temporal estudiado. Este filtro, también tiene por objetivo mitigar errores en la asociación de los niveles de servicio a las transacciones, donde dos compras muy cercanas, podrían tener niveles de servicio iguales, e inducir conclusiones erróneas.

■ **Filtro 4: Días sin mediciones de niveles de servicio**

Para realizar un correcto análisis del efecto de niveles de servicio sobre el comportamiento de compra, es necesario consolidar un panel de datos con toda la información disponible, esto implica que no pueden existir registros nulos para las variables en estudio. Por ello se filtran del panel de datos todos los clientes que para alguna de sus compras no puedan ser asociados todos los niveles de servicio en estudio.

■ **Filtro 5: Mínimo de Transacciones**

Por definición, para el cálculo del tiempo entre transacciones se requiere un mínimo de dos transacciones realizadas en el intervalo temporal en estudio. Por ello se filtran del análisis todos los clientes que posean una única compra registrada.

4.1.2. Metodología de Asociación de Niveles de Servicio

Los niveles de servicio se registran una cantidad limitada de veces en el día, los cuales dependiendo del tipo de medición, poseen una periodicidad entre un par de minutos hasta horas. El panel de datos, debe representar lo más certeramente posible los niveles de servicio recibidos por cada cliente, para ello se desarrolló una metodología, que busca y establece prioridad de asociación entre las mediciones realizadas en un horario cercano al de la transacción. En primer lugar se consolidan las siguientes tres mediciones con las que se establecerá la regla de asociación:

1. Registro del nivel de servicio en estudio que coincida exactamente con el registro temporal de la compra.
2. Registro del nivel de servicio en estudio más cercano, tal que el registro temporal de la compra sea posterior a este.
3. Registro del nivel de servicio en estudio más cercano, tal que el registro temporal de la compra sea anterior a este.

Media	Desviación Estándar	Mínimo	Máximo
17.343	21.702	1	228

Tabla 4.1: Estadísticos Descriptivos del Tiempo entre Compras

Con las mediciones consolidadas, se tienen en el mejor de los casos tres registros, correspondientes a niveles de servicio capturados en tiempos anteriores, iguales y posteriores a las compras de los clientes. Dependiendo de la hora, es posible que no exista alguno de los tres registros, ya sea por que no coinciden temporalmente con los tiempos de las compras, o por que estas se realizaron a principios o a finales del día, lo que no implica que no existe una medición anterior o posterior. Para la asociación final de la medición de nivel de servicio a los registros transaccionales se utiliza la siguiente regla de prioridad:

1. En caso que exista un registro de nivel de servicio coincidente en tiempo con la transacción, este es asociado. Este caso representa la situación ideal, donde se capturó exactamente el estado de las cajas al momento de una compra.
2. Si existe un registro anterior y uno posterior, se asocia el más cercano, y en caso de que la distancia temporal sea la misma para ambos, se asocia el promedio de las mediciones. Este caso representa una aproximación de los niveles recibidos por el cliente a la situación de cajas más cercana capturada.
3. En caso de que sólo exista un registro, este es asociado. Este caso aplica para los horarios extremos, donde por ejemplo en las primeras horas del día es posible que no existan mediciones anteriores al tiempo de la transacción, caso análogo a mediciones posteriores a una transacción realizada al final del día.

4.1.3. Descripción del Panel

Una vez aplicado el tratamiento de datos descrito en las secciones anteriores, se consolida un panel de 41.300 transacciones de 15.618 clientes. La Figura 4.1 ilustra un histograma de frecuencias de los registros de tiempo entre transacciones, donde se aprecia una gran proporción de registros presentan tiempos menores a un mes. La tabla 4.1 posee los estadísticos descriptivos de la variable dependiente, a partir de la cual se puede concluir que en promedio, los clientes del panel asisten aproximadamente cada dos semanas a alguna de las tiendas de la cadena (17.3 días), comportamiento que coincide con la descripción del formato vecinal, donde se esperan más casos de clientes que asisten múltiples veces en el mes. Esta intuición de estudiará con mas detalle en la sección que describe las variables transaccionales incluidas en los modelos.

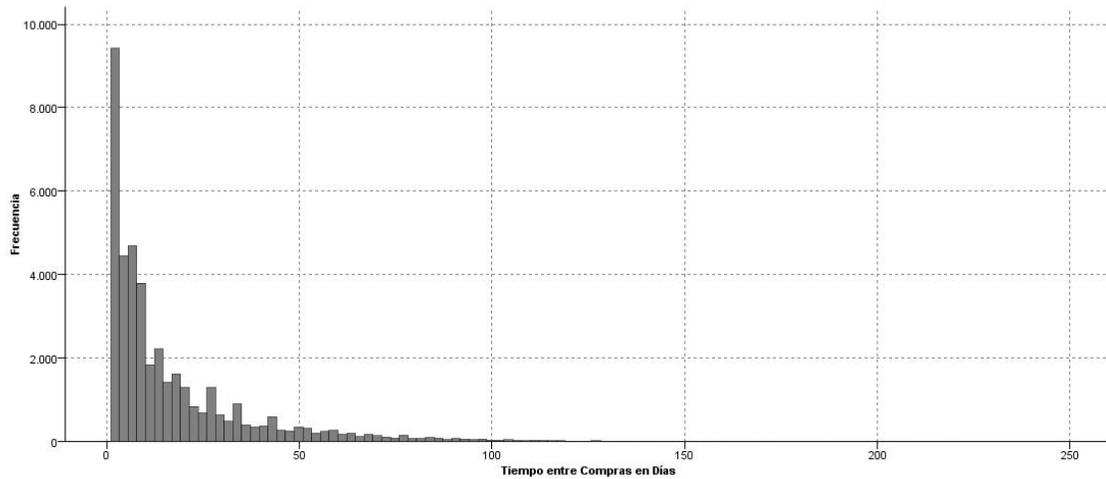


Figura 4.1: Histograma de Tiempos entre Compras

4.2. Estimación de Modelos Nulos

Los modelos nulos, corresponden a ajustes de los tiempos entre compras, utilizando únicamente parámetros de distribuciones de probabilidad. Estos permitirán generar una referencia de comparación para los modelos que incluyen variables, y evaluar cambios en los parámetros de las distribuciones e indicadores de ajuste.

La Tabla 4.2 posee la estimación de los parámetros de las distribuciones, en la cual se observa una clara tendencia y mejora en el ajuste a medida que se agrega complejidad y flexibilidad en los modelos. En particular el modelo Gamma-Weibull posee el mejor ajuste entre los tres modelos. La heterogeneidad de la muestra se ve reflejada en el parámetro r , donde el modelo Gamma-Weibull la ajusta mejor que el Gamma-Exponencial, con un valor menor y cercano a la unidad. En cuanto a significancia estadística, todos los parámetros poseen un p-valor inferior al 0.1% lo que refleja una alta significancia de los modelos.

La notación relativa a la significancia de los parámetros utilizada en todos los modelos descritos en este estudio es la siguiente:

- * : pvalor < 0,05
- ** : pvalor < 0,01
- *** : pvalor < 0,001

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0577*** (0.0049)		
α		53.3519*** (0.0323)	32.4339*** (0.0007)
r		4.0461*** (0.0263)	1.8014*** (0.0010)
c			1.2009*** (0.0001)
$-LL$	159,137	158,027	157,800
AIC	318,276	316,058	315,607
BIC	318,285	316,075	315,633
<hr/>			
$n = 41,300$			

Tabla 4.2: Modelos Nulos (Sin Covariables)

4.3. Modelos con Covariables

Las variables de estudio o covariables, se incluyen en la especificación de la tasa entre compras (λ), donde variaciones en ellas, implican un aumento o disminución de la tasa entre compras, lo que en consecuencia genera variaciones en el tiempo de retorno a alguna de las tiendas de la cadena.

Los modelos propuestos consideran que existe una tasa base λ_0 la cual sufre cambios en función de las variables de la última compra compra X_{t-1} , las cuales poseen un efecto β el cual explica cambios en el tiempo que demorará el cliente en volver a realizar una transacción.

$$\lambda(t) = \lambda_0 \exp(\beta X_{t-1})$$

Los modelos con covariables estiman además de los parámetros de las distribuciones, un vector β , de coeficientes asociados al impacto individual de cada una variables incluidas en el modelo. Las siguientes secciones describen la especificación de los modelos NBD, ante la inclusión de variables explicativas.

4.3.1. Modelo Exponencial con Covariables

Este modelo asume que el tiempo del cliente i entre las visitas $j - 1$ y j (t_{ij}), está exponencialmente distribuido con tasa λ_{ij-1} , la cual se modela a partir de una tasa base λ y variables explicativas de la última visita X_{ij-1} .

$$f(t|\lambda_{ij-1}) = \lambda_{ij}e^{-\lambda_{ij}t}$$

$$\lambda_{ij-1} = \lambda \exp(\beta X_{ij-1})$$

La logverosimilitud, para un cliente i que registra V_i visitas, y un comportamiento de compra X_i , se puede expresar de la siguiente forma:

$$LL = \sum_{i=1}^n \sum_{j=1}^{V_i} (\text{Log}(\lambda) + \beta X_{ij-1} - \lambda \exp(\beta X_{ij-1}) t_{ij})$$

4.3.2. Modelo Gamma-Exponencial con Covariables

Este modelo introduce heterogeneidad no observada en la especificación asumiendo que la tasa λ distribuye *Gamma*, es decir la tasa base es diferente entre los clientes, y esta se modela mediante una distribución de probabilidad.

$$f(t|\lambda_{ij-1}) = \lambda_{ij}e^{-\lambda_{ij-1}t}$$

$$\lambda_{ij-1} = \lambda \exp(\beta X_{ij-1})$$

$$\lambda \sim \text{Gamma}(r, \alpha)$$

La función de densidad acumulada es la siguiente:

$$F(t|\lambda, \alpha, r, \beta) = \int_0^\infty F(t|\lambda)g(\lambda)d\lambda = 1 - \alpha^r(\alpha + te^{\beta X})^{-r}$$

La función de distribución de probabilidad se puede calcular a partir de la función de densidad acumulada.

$$f(t|\alpha, r, \beta) = \frac{\partial F(t|\alpha, r, \beta)}{\partial t} = \frac{r\alpha^r e^{\beta X}}{(\alpha + te^{\beta X})^{r+1}}$$

La logverosimilitud, para un cliente i que registra V_i visitas, y un comportamiento de compra X_i , se puede expresar de la siguiente forma:

$$LL = \sum_{i=1}^n \sum_{j=1}^{V_i} (\text{Log}(r) + r\text{Log}(\alpha) + \beta X_{ij-1} - (r + 1)\text{Log}(\alpha + t_{ij}e^{\beta X_{ij-1}}))$$

4.3.3. Modelo Gamma-Weibull con Covariables

Este modelo introduce dependencia en la duración, asumiendo que el tiempo entre compras t_{ij} del cliente i entre las visitas $j-1$ tiene una distribución *Weibull*.

$$F(t|\lambda_{ij-1}, c) = 1 - e^{-\lambda_{ij-1}t^c}$$

$$\lambda_{ij-1} = \lambda \exp(\beta X_{ij-1})$$

$$\lambda \sim \text{Gamma}(r, \alpha)$$

La función de densidad acumulada es la siguiente:

$$F(t|\alpha, r, c) = \int_0^\infty F(t|\lambda_{ij}, c)g(\lambda)d\lambda = 1 - \alpha^r (\alpha + e^{\beta X} t^c)^{-r}$$

La función de distribución de probabilidad se puede calcular a partir de la función de densidad acumulada.

$$f(t|\alpha, r, c, \beta) = \frac{\partial F(t|\alpha, c, r)}{\partial t} = \frac{r\alpha^r e^{\beta X} c t^{c-1}}{(\alpha + e^{\beta X} t^c)^{r+1}}$$

Luego la logverosimilitud para un panel donde el cliente i registra V_i visitas se puede expresar de la siguiente forma:

$$LL = \sum_{i=1}^n \sum_{j=1}^{V_i} (\text{Log}(r) + r \text{Log}(\alpha) + \beta X_{ij-1} + \text{Log}(c) + (c-1) \text{Log}(t_{ij-1}) - (r+1) \text{Log}(\alpha + e^{\beta X_{ij-1}} t_{ij-1}^c))$$

4.4. Variables

4.4.1. Variables Transaccionales

En esta sección se describen las variables obtenidas y/o generadas, a partir de las bases transaccionales, las que poseen información de las compras y cajas donde estas se realizaron.

- $Dollars_{ij-1}$: Monto total en dólares de la compra del cliente i en la visita $j-1$.
- $Item_{ij-1}$: Total de productos diferentes comprados por el cliente i en la visita $j-1$. Esta variable hace referencia al identificador único de cada producto.
- $RatioDollarItem_{ij-1}$: Valor promedio de los productos comprados por el cliente i en la visita $j-1$. Esta variable se construye a partir del ratio entre las variables descritas anteriormente $(\frac{Dollars_{ij-1}}{Item_{ij-1}})$.
- $Checkout_{ij-1}$: Variable binaria que toma el valor 1 si el cliente i realizó una compra en una caja regular en su visita $j-1$ y 0 si se realizó en otro tipo de caja.
- $SelfCheckout_{ij-1}$: Variable binaria que toma el valor 1 si el cliente i realizó una compra en una caja autoservicio en su visita $j-1$ y 0 si se realizó en otro tipo de caja.

4.4.2. Variables de Niveles de Servicio

En esta sección se describen las variables generadas a partir de las bases de niveles de servicio, utilizando la metodología de asociación descrita en el capítulo de preparación de datos, es decir variables construidas con mediciones de nivel de servicio asociadas a cada transacción. Las variables de este tipo incluidas en los modelos son las siguientes:

- $WaitTime_{it-1}$ =Tiempo de espera en cola asociado al cliente i en la visita $j - 1$.
- $QueueLenght_{ij-1}$ =Total de personas en cola asociadas al cliente i en la visita $j - 1$.
- $OpenQueue_{ij-1}$ =Total de cajas abiertas asociadas al cliente i en la visita $j - 1$.
- $QueueLenghtInd_{ij-1}$ = Corresponde a una aproximación de la cola individual asociada al cliente i en su visita $j - 1$, obtenida a partir del ratio entre el total de personas en cola y las cajas disponibles ($\frac{QueueLenght_{ij-1}}{OpenQueue_{ij-1}}$).

Mediante las variables transaccionales $Checkout_{ij-1}$ y $SelfCheckout_{ij-1}$, es posible generar variables interactivas que asocian los niveles de servicio al tipo de caja en el cual se realizaron las transacciones:

- $COpenQueue_{it-1}$ = Variable que toma el valor $OpenQueue_{ij-1}$ si el cliente i realizó una compra en una caja regular en su visita $j - 1$. La variable toma el valor cero 0 si la compra se realizó en otro tipo de caja.
- $CQueueLenght_{it-1}$ =Variable que toma el valor $QueueLenght_{ij-1}$ si el cliente i realizó una compra en una caja regular en su visita $j - 1$. La variable toma el valor cero 0 si la compra se realizó en otro tipo de caja.
- $CQueueLenghtInd_{it-1}$ =Variable que toma el valor $QueueLenghtInd$ si el cliente i realizó una compra en una caja regular en su visita $j - 1$. La variable toma el valor cero 0 si la compra se realizó en otro tipo de caja.
- $SCQueueLenght_{it-1}$ =Variable que toma el valor $QueueLenght_{ij-1}$ si el cliente i realizó una compra en una caja autoservicio en su visita $j - 1$. La variable toma el valor cero 0 si la compra se realizó en otro tipo de caja.
- $CWaitTime_{it-1}$ =Variable que toma el valor $WaitTime_{it-1}$ si el cliente i realizó una compra en una caja regular en su visita $j - 1$. La variable toma el valor cero 0 si la compra se realizó en otro tipo de caja.
- $OpenQueueZero_{ij-1}$ =Variable binaria que toma el valor 1 si el cliente i se enfrentó a una situación de cero cajas regulares abiertas en su visita $j - 1$. La variable toma el valor 0 en otro caso.

Fecha	Feriado
25 Dic 2012	Navidad
1 Ene 2013	Año Nuevo
21 Ene 2013	Día de Martin Luther King
18 Feb 2013	Día del Presidente
27 May 2013	Día de la Memoria
4 Jul 2013	Día de la Independencia

Tabla 4.3: Fechas que conforman la variable Feriado

4.4.3. Variables de Control

En esta sección se describen las variables de control, las cuales tienen por objetivo incluir en los modelos comportamientos esperados, de forma de controlar el efecto estimado del resto de las variables en estudio. Por lo general estas variables están asociadas a estacionalidades y poseen una alta significancia. Las variables de este tipo incluidas en los modelos son las siguientes:

- $Weekend_{it-1}$ =Variable binaria que toma el valor 1 si la visita $j - 1$ del cliente i se realizó un fin de semana y 0 si se realizó en un día de semana.
- $CWeekend_{it-1}$ =Variable binaria que toma el valor 1 si la visita $j - 1$ del cliente i se realizó un fin de semana y además en una caja regular. La variable toma el valor cero 0 en otro caso.
- $Feriado_{it-1}$ =Variable binaria que toma el valor 1 si la visita $j - 1$ del cliente i se realizó un día festivo y 0 en otro caso. Las fechas que conforman esta variables se encuentran detalladas en la Tabla 4.3.
- $AnteFeriado_{it-1}$ =Variable binaria que toma el valor 1 si la visita $j - 1$ del cliente i se realizó el día anterior a un festivo y 0 en otro caso. Las fecha que conforman esta variable son los días inmediatamente anteriores a los que conforman la variable Feriado.
- $SuperBowl_{it-1}$ =Variable binaria que toma el valor 1 si la visita $j - 1$ del cliente i se realizó el día del Super Bowl (3 Feb 2013) y 0 en otro caso.
- $Distancia_{it-1}$ =Kilómetros desde el hogar del cliente i a la tienda en la que realizó la transacción $j - 1$.

Variable	Media	Desviación Estándar	Mínimo	Máximo
$Item_{ij-1}$	8.871	10.089	1	140
$SelfCheckout_{ij-1}$	-	-	0	1
$Dollars_{ij-1}$	40.275	47.938	0.140	812.150
$Weekend_{it-1}$	-	-	0	1
$COpenQueue_{ij-1}$	3.216	3.057	0	13
$CQueueLenght_{ij-1}$	4.465	5.101	0	37
$SCQueueLenght_{ij-1}$	1.140	1,882	0	12
$CQueueLenghtInd_{ij-1}$	0.835	0.810	0	8
$OpenQueueZero_{ij-1}$	-	-	0	1
$CWaitTime_{ij-1}$	23.836	44.118	0	455
$Feriado_{ij-1}$	-	-	0	1
$AnteFeriado_{ij-1}$	-	-	0	1
$SuperBowl_{ij-1}$	-	-	0	1
$CWeekend_{ij-1}$	-	-	0	1
$Distancia_{ij-1}$	192.305	583.318	0.663	7,420.364
$RatioDollarItem_{ij-1}$	4.995	4.092	0.140	250

Tabla 4.4: Estadísticos Descriptivos

4.4.4. Estadísticos Descriptivos

El conjunto de variables a estudiar, está conformado por todas las variables descritas en la sección anterior, con la consideración, de que los niveles de servicio están asociados al tipo de caja correspondiente, es decir, se estudiará solamente el uso de las variables interactivas que utilizan el tipo de caja, ya que representan los niveles asociados a esperas recibidos por los clientes en sus visitas. La Tabla 4.4 posee los estadísticos descriptivos de las variables a estudiar utilizando el panel completo de datos.

El tipo de caja es un factor importante a estudiar en la diferenciación de las variables, ya que un 38 % de las compras registradas se realizaron en una caja autoservicio. Las Tablas 4.5 y 4.6, corresponden a una segmentación de los estadísticos descriptivos en función del tipo de caja en la que se realizó la transacción. Al comparar las tablas, se aprecia una diferencia en las transacciones según el tipo de caja, donde las cajas regulares presentan mayores compras con mayores montos e items promedio que las cajas autoservicio, lo mismo ocurre con los niveles de servicio, donde las cajas regulares presentan mayores lagos de cola promedio.

Variable	Media	Desviación Estándar	Mínimo	Máximo
<i>Item</i> _{<i>ij-1</i>}	11.510	11.574	1	140
<i>SelfCheckout</i> _{<i>ij-1</i>}	-	-	-	-
<i>Dollars</i> _{<i>ij-1</i>}	52.273	55.355	0.250	812.150
<i>Weekend</i> _{<i>ij-1</i>}	-	-	0	1
<i>COpenQueue</i> _{<i>ij-1</i>}	5.196	2,193	0	13
<i>CQueueLenght</i> _{<i>ij-1</i>}	7.214	4,712	0	37
<i>SCQueueLenght</i> _{<i>ij-1</i>}	-	-	-	-
<i>CQueueLenghtInd</i> _{<i>ij-1</i>}	1,348	0.606	0	8
<i>OpenQueueZero</i> _{<i>ij-1</i>}	-	-	0	1
<i>CWaitTime</i> _{<i>ij-1</i>}	38.508	50.790	1	455
<i>Feriado</i> _{<i>ij-1</i>}	-	-	0	1
<i>AnteFeriado</i> _{<i>ij-1</i>}	-	-	0	1
<i>SuperBowl</i> _{<i>ij-1</i>}	-	-	0	1
<i>CWeekend</i> _{<i>ij-1</i>}	-	-	0	1
<i>Distancia</i> _{<i>ij-1</i>}	217.806	638.645	0.663	7,420.364
<i>RatioDollarItem</i> _{<i>ij-1</i>}	4.935	4.086	0.250	250

Tabla 4.5: Estadísticos Descriptivos Cajas Regulares

Variable	Media	Desviación Estándar	Mínimo	Máximo
<i>Item</i> _{<i>ij-1</i>}	4.583	4.452	1	53
<i>SelfCheckout</i> _{<i>ij-1</i>}	-	-	1	1
<i>Dollars</i> _{<i>ij-1</i>}	20.783	20.970	0.140	246.270
<i>Weekend</i> _{<i>ij-1</i>}	-	-	0	1
<i>COpenQueue</i> _{<i>ij-1</i>}	-	-	-	-
<i>CQueueLenght</i> _{<i>ij-1</i>}	-	-	-	-
<i>SCQueueLenght</i> _{<i>ij-1</i>}	2.993	1.937	0	12
<i>CQueueLenghtInd</i> _{<i>ij-1</i>}	-	-	-	-
<i>OpenQueueZero</i> _{<i>ij-1</i>}	-	-	-	-
<i>CWaitTime</i> _{<i>ij-1</i>}	-	-	-	-
<i>Feriado</i> _{<i>ij-1</i>}	-	-	0	1
<i>AnteFeriado</i> _{<i>ij-1</i>}	-	-	0	1
<i>SuperBowl</i> _{<i>ij-1</i>}	-	-	0	1
<i>CWeekend</i> _{<i>ij-1</i>}	-	-	-	-
<i>Distancia</i> _{<i>ij-1</i>}	150.877	477.151	0.663	4,755,050
<i>RatioDollarItem</i> _{<i>ij-1</i>}	5.091	4.098	0.140	129.990

Tabla 4.6: Estadísticos Descriptivos Cajas Autoservicio

4.5. Estimación de Modelos con Covariables

4.5.1. Metodología de Selección de Modelo

La metodología de selección del modelo final utiliza una combinación de Backward Selection, con los siguientes cuatro criterios:

1. **Significancia Estadística:** La significancia mínima exigida para las variables es de un 95 % de confianza.
2. **Bondad de Ajuste:** Los modelos son comparados mediante los criterios Log-Verosimilitud (LL), Akaike (AIC) y Bayesiano (BIC), a partir de estos indicadores se podrá seleccionar entre modelos que posean todas sus variables significativas.
3. **Interpretabilidad de Coeficientes:** Los coeficientes deben hacer sentido en base a la intuición del tipo de variables, es decir, se espera que las variables transaccionales representen abastecimiento, por lo que deberían tener un efecto negativo en la tasa. Para el caso de las variables de niveles de servicio también se espera un efecto negativo sobre la tasa, donde una mala experiencia impliquen mayores tiempos de retorno a las salas de venta de la cadena.
4. **Diversificación de Tipos de Variables:** Para controlar mitigar errores en las conclusiones, los modelos deben combinar variables transaccionales, de niveles de servicio y controles, de esta forma se equilibran los impactos asignados a cada variable.

En el capítulo de Anexos A, se encuentran documentados los modelos correspondientes a las iteraciones realizadas, donde se comienza con todo el conjunto de variables y se eliminan o agregan variables en función de los criterios descritos. Para la estimación de los modelos se programaron las Log-Verosimilitudes, y se estimaron los parámetros mediante funciones de optimización en MATLAB.

4.5.2. Modelo Escogido

La Tabla 4.7 corresponde a los parámetros del modelo escogido en función de la metodología de selección definida. El modelo estima todos los coeficientes con valor negativo, lo cual implica que un aumento en cualquiera de las variables incluidas en el modelo implica una disminución de las tasas entre compras, lo que en consecuencia se traduce en un aumento de los tiempos entre transacciones.

El efecto numérico del impacto de las variables puede ser estudiado a través de la elasticidad de la tasa, la cual puede ser calculada a partir de los coeficientes y los valores promedio de las variables $\beta_i \bar{X}_i$. Gupta (1991) representa el aumento o disminución porcentual de la tasa ante un aumento o disminución porcentual de las variables del modelo de la forma descrita anteriormente.

Modelos con Covariates

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0685*** (0.0099)		
α		45.4988*** (0.0339)	27.5014*** (0.0300)
r		4.0851*** (0.0264)	1.8626*** (0.0328)
c			1.1981*** (0.0084)
$\beta_{Weekend}$	-0.0185*** (0.0106)	-0.0209*** (0.0128)	-0.0252*** (0.0152)
$\beta_{QueueLenght}$	-0.0057*** (0.0013)	-0.0052*** (0.0015)	-0.0059*** (0.0018)
$\beta_{OpenQueueCheckCero}$	-0.2585*** (0.0521)	-0.2870*** (0.0590)	-0.3521*** (0.0611)
$\beta_{WaitTime}$	-0.0001** (0.0001)	-0.0002** (0.0001)	-0.0003*** (0.0002)
$\beta_{SuperBowl}$	-0.2759** (0.1082)	-0.3087** (0.1220)	-0.3869** (0.1232)
$\beta_{RatioDollarItem}$	-0.0075*** (0.0014)	-0.0076*** (0.0016)	-0.0093*** (0.0019)
$-LL$	158,892	157,858	157,631
AIC	317,798	315,732	315,280
BIC	317,858	315,801	315,357
$n = 41,300$			

Tabla 4.7: Modelos Iteración 10

Los modelos estimados en este estudio incluyen variables binarias, e interactivas que dependen del tipo de caja, esto fue considerado en el cálculo de elasticidades para este tipo de variables, donde para las binarias se utilizó directamente el coeficiente, con el fin de calcular la elasticidad ante la ocurrencia del evento. Para el caso de las interactivas se utilizó el valor promedio, sujeto a compras realizadas en el tipo de caja respectivo, de esta forma se puede distinguir el efecto asociado de los niveles de servicio asociados a de cada tipo de caja. La tabla 4.8 ilustra las elasticidades calculadas para las variables del modelo final escogido.

Variable	Exponencial	Gamma-Exponencial	Gamma-Weibull
<i>Weekend</i>	-1.85 %	-2.09 %	-2.52 %
<i>QueueLenght</i>	-4.10 %	-3.75 %	-4.26 %
<i>OpenQueueCheckCero</i>	-25.85 %	-28.70 %	-35.21 %
<i>WaitTime</i>	-0.50 %	-0.84 %	-1.17 %
<i>SuperBowl</i>	-27.59 %	-30.87 %	-38.69 %
<i>RatioDollarItem</i>	-3.74 %	-3.81 %	-4.63 %

Tabla 4.8: Elasticidades Modelos Iteración 10

4.5.3. Análisis de Resultados

Variables Transaccionales

Las variables transaccionales asociadas a montos de compras y cantidad de items presentan un comportamiento opuesto en los modelos cuando se incluyen en conjunto, por lo general la variable $Dollars_{ij-1}$ posee el efecto esperado asociado al abastecimiento, donde mayores compras implican un aumento en los tiempos entre transacciones, pero la variable $Item_{ij-1}$ posee el efecto contrario, lo cual da indicios de una interacción no lineal entre ambas variables.

Al estimar modelos con la inclusión separada de alguna de las dos variables, como en el caso de las Iteraciones 8 y 11 (Tablas A.8 y A.10 de los Anexos), ambas mantienen el efecto contrario al esperado, por ello se construye la variable $RatioDollarItem_{ij-1}$ para incluir el efecto de ambas variables en un comportamiento de abastecimiento, el cual está asociado el monto promedio de los items comprados, donde montos promedio mayores pueden ser asociados a productos compuestos por múltiples unidades y a productos con precio asociado según su peso. Como se ilustra en la Tabla 4.8, la variable posee una elasticidad de hasta un -4.63 % para el caso del Modelo Gamma-Weibull, lo que implica que captura un efecto positivo sobre los tiempos entre compras.

Variables de Niveles de Servicio

El tipo de caja por si solo no posee significancia estadística en los modelos estudiados, pero si algunas de las variables interactivas construidas para diferencias los niveles de servicio por tipo de caja. Para el caso de las cajas regulares tanto el tiempo de espera en cola como los largos de cola son significativos con un efecto negativo sobre la tasa de compras. Como se ilustra en la Tabla 4.8 de elasticidades, la variable $WaitTime$ posee un efecto de hasta un -1.17 % y la variable $QueueLenght$ una elasticidad de hasta un -4.26 %, ambas reflejan una percepción negativa de las esperas en cajas de supermercados, donde la percepción del total de personas en cola posee un mayor efecto que los tiempos de espera individuales recibidos por un cliente. La variable asociada a total de personas en cola en cajas autoservicio, posee un efecto positivo sobre la tasa entre compras como se aprecia en todas las iteraciones que incluyen la variable $SCQueueLenght_{ij-1}$, esta diferencia con respecto a las cajas regulares

podría ser explicada por la configuración distinta que poseen las cajas autoservicio, donde una cola única es servida por múltiples cajas. Esta diferencia podría implicar que los efectos de las variables no sean comparables.

La variable $OpenQueueZero_{ij-1}$ posee un gran impacto sobre la tasa de compras, alcanzando hasta un -35.21 % de elasticidad como se ilustra en la Tabla 4.8. La interpretación de este efecto se puede asociar a una espera forzada en función de la disponibilidad de recursos, o un cambio forzado del tipo de caja en el cual el cliente tenía definido realizar su compra. Ambos eventos son percibidos negativamente implicando un aumento en los tiempos de retorno a las salas de venta de la cadena.

Las variables asociadas al número de cajas regulares abiertas y a tiempo estimado de largo de cola individual resultaron no ser estadísticamente significativas desde la primera iteración, por lo cual fueron excluidas del resto de las estimaciones (Tabla A.1 de los Anexos).

Variables de Control

Los controles que mantuvieron su significancia son las variables asociadas a fines de semana y al día del Super Bowl, ambas con un efecto negativo en la tasa de compras. Este comportamiento es esperable, ya que en ambos casos se esperan compras de magnitud mayor y por ende un mayor abastecimiento de los clientes, lo que implica un aumento en el tiempo de regreso a las tiendas de supermercado. Como se ilustra en la Tabla 4.8 la variable $Weekend_{ij-1}$ posee una elasticidad de hasta un -1.17 %, y la variable $SuperBowl_{ij-1}$ hasta un -38.69 %.

Las variables asociadas a feriados y los días inmediatamente anteriores resultaron no ser estadísticamente significativas. Es posible que esto se deba a la estimación agregada de múltiples fechas festivas, y que en caso de dividir cada festivo de forma independiente alguno resulte estadísticamente significativo.

La variable $Distancia_{ij-1}$ fue incluida en los modelos por ser el único aspecto demográfico disponible en las bases de datos. Los coeficientes estimados evidencian un efecto positivo en la tasa, lo cual no es fácil de interpretar, ya que esto implicaría que mientras mayores son las distancias recorridas a la tienda, menor es el tiempo de retorno. El efecto esperado es exactamente el contrario, donde distancias menores evidencien cercanía a tienda, y por ende una mayor cantidad de visitas en el intervalo temporal estudiado. De todas formas el efecto es cercano a cero, por lo que la variable fue excluida en la iteración 3 (Tabla A.3 de los Anexos).

Modelo	Parámetros	χ^2	p-valor
Exponencial Nulo	1	-	-
Exponencial Iteración 10	7	245.11	0.0000
Gamma-Exponencial Nulo	2	-	-
Gamma-Exponencial Iteración 10	8	168.76	0.0000
Gamma-Weibull Nulo	3	-	-
Gamma-Weibull Iteración 10	9	169.64	0.0000

Tabla 4.9: Test Chi-Cuadrado de Ratios de Verosimilitud

4.6. Comparación de Modelos

La validación de los modelos se realiza mediante el Ratio de Verosimilitud, el cual compara si la inclusión de las variables es estadísticamente significativa con respecto a los modelos nulos estimados inicialmente. Esta conclusión se realiza mediante un test Chi-Cuadrado con grados de libertad asociados a los parámetros del modelo. La Tabla 4.9 posee el resultado de los test aplicados donde se concluye que la inclusión de covariables es estadísticamente significativa.

Capítulo 5

Modelos con Tiempo Censurado

5.1. Tiempo Censurado

El tiempo transcurrido entre la última transacción registrada y la fecha límite del intervalo temporal de estudio también posee información relevante sobre el comportamiento de compra de los clientes. Este tiempo incompleto, denominado censurado, puede ser incorporado en los modelos como la fracción del tiempo transcurrido entre la última compra y la siguiente la cual podría ser realizada en alguna fecha posterior a la máxima incluida en las fuentes de datos.

Este tiempo se incluye en la función de verosimilitud, mediante la Función de Supervivencia, es decir la probabilidad de que la transacción se realice en un tiempo superior a la diferencia entre la fecha límite de la base transaccional y la fecha de la última compra. En las siguientes secciones se describen las especificaciones y las funciones de Log-Verosimilitud asociadas a los modelos de duración en estudio con la incorporación de tiempos censurados. Van den Bulte e Iyengar 2011, sugieren que esta es la mejor forma de minimizar sesgos, y obtener conclusiones correctas de las variables en estudio. Un análisis comparativo entre los modelos estimados en el capítulo anterior y los con tiempos censurados, permitirá obtener un resultado más robusto en términos de variables significativas y sus impactos, los cuales se ven afectados en función de la magnitud de datos truncados del panel.

A continuación se describen las especificaciones de los modelos estudiados en el Capítulo 4 con la incorporación de tiempos censurados en las funciones de Log-Verosimilitud, para un panel donde cada cliente i registra V_i visitas, con tiempo censurado c_i .

Modelo Exponencial

$$P[t > T] = 1 - F[T|\lambda] = e^{-\lambda T}$$
$$LL = \sum_{i=1}^n \sum_{j=1}^{V_i} (\text{Log}(\lambda) - \lambda t_{ij}) + \sum_{i=1}^n (-\lambda c_i)$$

Modelo Gamma-Exponencial

$$P[t > T] = 1 - F[T|\alpha, r] = \left(\frac{\alpha}{\alpha + r} \right)^r$$
$$LL = \sum_{i=1}^n \sum_{j=1}^{V_i} (\text{Log}(r) + r\text{Log}(\alpha) - (r + 1)\text{Log}(\alpha + t_{ij})) + \sum_{i=1}^n r(\text{Log}(\alpha) - \text{Log}(\alpha + c_i))$$

Modelo Gamma-Weibull

$$P[t > T] = 1 - F[T|\alpha, r, c] = \alpha^r (\alpha + T^c)^{-r}$$
$$LL = \sum_{i=1}^n \sum_{j=1}^{V_i} (\text{Log}(r) + r\text{Log}(\alpha) + \text{Log}(c) + (c - 1)\text{Log}(t_{ij}) - (r + 1)\text{Log}(\alpha + t_{ij}^c))$$
$$+ \sum_{i=1}^n r\text{Log}(\alpha) - r\text{Log}(\alpha + c_i^c)$$

5.2. Preparación de Datos

5.2.1. Filtros, Supuestos y Definiciones Adicionales

La inclusión de los tiempos incompletos o censurados requiere de filtros, supuestos y definiciones adicionales a los descritos en la sección de Preparación de datos del Capítulo 4.

▪ Supuesto 2: Variables Explicativas del Tiempo Censurado

Para mantener la estructura del panel de datos y facilitar la estimación de parámetros, se asume que las variaciones del tiempo censurado son explicadas a partir de las mismas variables que explican los cambios en los tiempos entre transacciones. Esto implica que ambos tiempos poseen la misma especificación en los modelos.

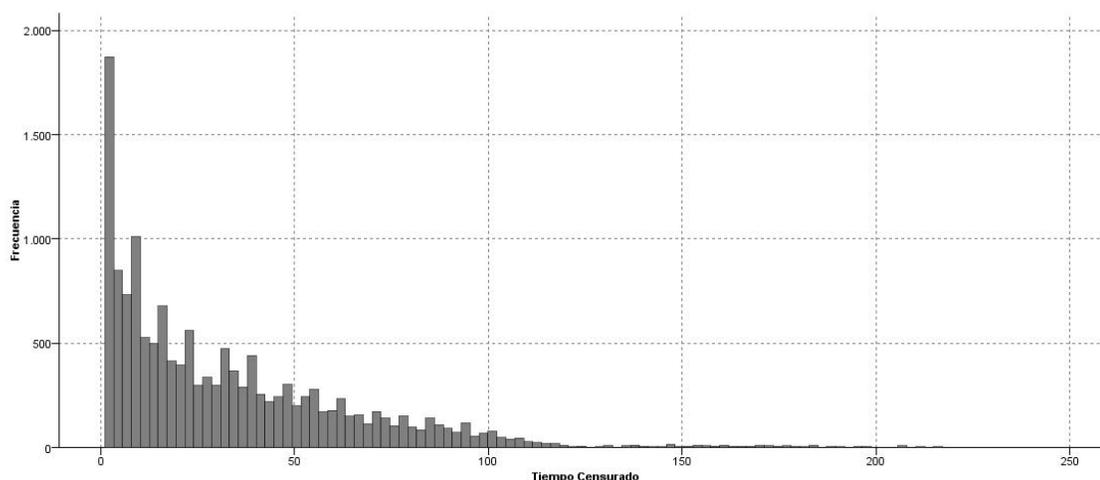


Figura 5.1: Histograma de Frecuencias de Tiempos Censurados

	Media	Desviación Estándar	Mínimo	Máximo
Tiempo Censurado	33,134	32,109	1	231
Tiempo entre Compras	17.301	21.637	1	228

Tabla 5.1: Estadísticos Descriptivos del Tiempo Censurado

■ Filtro 6: Niveles de Servicio en la Última Compra

De forma análoga al **Filtro 4**, se excluyen del panel de datos los clientes que no posean alguna de las mediciones de niveles de servicio el día de la última transacción registrada. Para los modelos anteriores este filtro no era necesario, ya que estos utilizan variables de la compra anterior. En este caso las variables que explican el tiempo censurado son las de la última compra registrada por ello se requiere la existencia de mediciones de las variables estudiadas.

5.2.2. Descripción del Panel

La aplicación del **Filtro 6** disminuye el universo de clientes y transacciones, ya que se exige que la última compra presente mediciones de niveles de servicio. El panel final está conformado por 40,975 transacciones pertenecientes a 15,454 clientes, lo que representa una reducción de un 1.06 % del universo de clientes. La Figura 5.1 posee la distribución de tiempos censurados, los cuales poseen una distribución similar a los tiempos entre transacciones, pero con media y desviación estándar de mayor magnitud que los tiempos entre transacciones como se aprecia en la Tabla 5.1.

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0342*** (0.0049)		
α		23.7007*** (0.0231)	14.9312*** (0.0008)
r		1.3857*** (0.0164)	0.5001*** (0.0001)
c			1.4693*** (0.0000)
$-LL$	179,232	175,980	167,772
AIC	358,465	351,963	335,550
BIC	358,474	351,981	335,576
<hr/>			
$n = 40,975$			

Tabla 5.2: Modelos Censurados Nulos (Sin Covariables)

5.3. Estimación de Modelos Nulos

Las Tabla 5.2 y 5.3, corresponden a las estimaciones de los modelos nulos, con y sin la inclusión de tiempos censurados respectivamente. La estimación sin censura se realizó con fines comparativos para estudiar el efecto de los filtros adicionales sobre el panel de clientes, donde se aprecian cambios mínimos en los parámetros, y un leve aumento en indicadores de bondad de ajuste, provocado por la disminución de los clientes que conforman el panel.

La inclusión de tiempos censurados implica cambios en la estimación de los parámetros, los cuales ajustan tasas considerablemente menores de compras. Esto tiene sentido ya que los tiempos censurados poseen una media de aproximadamente el doble de los tiempos entre transacciones. La bondad de ajuste también disminuye y mantiene la tendencia de mejora con la inclusión de complejidad y flexibilidad a los modelos.

5.4. Modelos con Covariables

De forma análoga a los modelos estimados en el Capítulo 4, tanto el tiempo entre compras t , como el tiempo censurado c pueden ser modelados a partir de una distribución exponencial, la cual posee como parámetro la tasa de compras λ_{ij} . La cual a su vez puede ser modelada a partir de una tasa base λ , y variables explicativas de la última visita X_{ij-1} .

A continuación se describen las especificaciones de los modelos con la incorporación de tiempos censurados en las funciones de Log-Verosimilitud, para un panel donde cada cliente i registra V_i visitas, con tiempo censurado c_i y variables explicativas X_{ij} .

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0578*** (0.0049)		
α		53.5916*** (0.0325)	32.6971*** (0.0007)
r		4.0679*** (0.0265)	1.8043*** (0.0010)
c			1.2031*** (0.0001)
$-LL$	157,785	156,693	156,468
AIC	315,572	313,389	312,941
BIC	315,581	313,407	312,967

$n = 40, 975$

Tabla 5.3: Modelos Nulos Actualizados(Sin Covariables)

Modelo Exponencial con Covariables

$$P[t > T] = 1 - F[T|\lambda_{it} = \lambda e^{\beta X}] = e^{-\lambda e^{\beta X} T}$$

$$LL = \sum_{i=1}^n \sum_{j=1}^{V_i} (\text{Log}(\lambda) + \beta X_{ij} - \lambda \exp(\beta X_{ij}) t_{ij}) - \sum_{i=1}^n \lambda e^{\beta X_{ij} c_i}$$

Modelo Gamma-Exponencial con Covariables

$$P[t > T] = 1 - F[T|\alpha, r] = \left(\frac{\alpha}{\alpha + T e^{\beta X}} \right)^r$$

$$LL = \sum_{i=1}^n \sum_{j=1}^{V_i} (\text{Log}(r) + r \text{Log}(\alpha) + \beta X_{ij} - (r+1) \text{Log}(\alpha + t_{ij} e^{\beta X_{ij}}))$$

$$+ \sum_{i=1}^n r \text{Log}(\alpha) - r \text{Log}(\alpha + c_i e^{\beta X_{ij}})$$

Modelo Gamma-Weibull con Covariables

$$P[t > T] = 1 - F[T|\alpha, r, c] = \alpha^r (\alpha + e^{\beta X} T^c)^{-r}$$

Variable	Media	Desviación Estándar	Mínimo	Máximo
$Item_{ij-1}$	8,754	10,006	1	140
$SelfCheckout_{ij-1}$	-	-	0	1
$Dollars_{ij-1}$	40,016	47,683	0.140	812,150
$Weekend_{ij-1}$	-	-	0	1
$COpenQueue_{ij-1}$	3,248	3,039	0	13
$CQueueLenght_{ij-1}$	4,491	5,062	0	37
$SCQueueLenght_{ij-1}$	1,114	1,862	0	12
$CQueueLenghtInd_{ij-1}$	0.842	0.805	0	8
$OpenQueueZero_{ij-1}$	-	-	0	1
$CWaitTime_{ij-1}$	24.113	44.286	0	455
$Distancia_{ij-1}$	176.710	553.885	0.663	7,420
$AnteFeriado_{ij-1}$	-	-	0	1
$Feriado_{ij-1}$	-	-	0	1
$SuperBowl_{ij-1}$	-	-	0	1
$CWeekend_{ij-1}$	-	-	0	1
$RatioDollarItem_{ij-1}$	5,063	4,291	0.058	250

Tabla 5.4: Estadísticos Descriptivos

$$\begin{aligned}
LL = \sum_{i=1}^n \sum_{j=1}^{V_i} & (\text{Log}(r) + r \text{Log}(\alpha) + \beta X_{ij} + \text{Log}(c) + (c-1) \text{Log}(t_{ij}) - (r+1) \text{Log}(\alpha + e^{\beta X_{ij}} t_{ij}^c)) \\
& + \sum_{i=1}^n r \text{Log}(\alpha) - r \text{Log}(\alpha + e^{\beta X_{v_{ij}}} c_i^c)
\end{aligned}$$

5.5. Variables

Las variables a estudiar son las mismas utilizadas en los modelos sin inclusión de tiempos censurados, esto en función del **Supuesto 2** el cual tiene por objetivo facilitar las especificaciones de los modelos.

5.5.1. Estadísticos Descriptivos

La Tabla 5.4 ilustra los estadísticos descriptivos del panel de clientes, por su parte las Tablas 5.5 y 5.6 diferencian los estadísticos por tipo de caja. En general se obtienen valores muy similares al panel original, donde para las cajas regulares se aprecian transacciones de mayor magnitud y niveles de servicio asociados a mayores esperas. Esta situación da indicios de que la incorporación de la última transacción por cliente al panel, no altera las características de las variables estudiadas.

Variable	Media	Desviación Estándar	Mínimo	Máximo
$Item_{ij-1}$	11.276	11.447	1	140
$SelfCheckout_{ij-1}$	-	-	-	-
$Dollars_{ij-1}$	51.552	54.865	0.250	812.150
$Weekend_{ij-1}$	-	-	0	1
$COpenQueue_{ij-1}$	5.179	2.175	0	13
$CQueueLenght_{ij-1}$	7.161	4.662	0	37
$SCQueueLenght_{ij-1}$	-	-	-	-
$CQueueLenghtInd_{ij-1}$	1.343	0.601	0	8
$OpenQueueZero_{ij-1}$	-	-	0	1
$CWaitTime_{ij-1}$	38.452	50.756	1.000	455
$Distancia_{ij-1}$	196.365	602.017	0.663	7,420
$AnteFeriado_{ij-1}$	-	-	0	1
$Feriado_{ij-1}$	-	-	0	1
$SuperBowl_{ij-1}$	-	-	0	1
$CWeekend_{ij-1}$	0.305	0.460	0	1
$RatioDollarItem_{ij-1}$	5.010	4.330	0.058	250

Tabla 5.5: Estadísticos Descriptivos Cajas Regulares

Variable	Media	Desviación Estándar	Mínimo	Máximo
$Item_{ij-1}$	4.512	4.410	1	53
$SelfCheckout_{ij-1}$	-	-	1	1
$Dollars_{ij-1}$	20.616	20.857	0.140	246.270
$Weekend_{ij-1}$	-	-	0	1
$COpenQueue_{ij-1}$	-	-	-	.
$CQueueLenght_{ij-1}$	-	-	-	-
$SCQueueLenght_{ij-1}$	2.987	1.924	0.000	12,000
$CQueueLenghtInd_{ij-1}$	-	-	-	-
$OpenQueueZero_{ij-1}$	-	-	-	-
$CWaitTime_{ij-1}$	-	-	-	-
$Distancia_{ij-1}$	143,657	459,890	0.663	4,755
$AnteFeriado_{ij-1}$	-	-	0	1
$Feriado_{ij-1}$	-	-	0	1
$SuperBowl_{ij-1}$	-	-	0	1
$CWeekend_{ij-1}$	-	-	-	-
$RatioDollarItem_{ij-1}$	5.153	4.223	0.100	137

Tabla 5.6: Estadísticos Descriptivos Cajas Autoservicio

5.6. Estimación de Modelos con Covariables

La metodología utilizada para la selección del modelo final mantiene los Significancia Estadística, Bondad de Ajuste, Interpretabilidad de Coeficientes y Deversificación de Tipos de Variables utilizados en los modelos sin inclusión de tiempos censurados.

5.6.1. Modelo Escogido

La Tabla 5.7 ilustra los parámetros del modelo final escogido, en función de la metodología definida. Las variables significativas del modelo son las mismas incluidas en el modelo final del Capítulo 4, con la diferencia de que la variable asociada a compras en cajas autoservicio ahora mantiene su significancia.

En cuanto a bondad de ajuste, el modelo Exponencial es el que mejor se ajusta a los datos del panel, seguido por el Gamma-Weibull y finalmente el Gamma Exponencial. Esta tendencia se aprecia en todas las iteraciones de los modelos que incluyen tiempos censurados, y se debe en gran parte a clientes los cuales poseen una probabilidad muy baja de regresar a las tiendas, los cuales reflejan grandes tiempos censurados en comparación con los tiempos entre compras de panel original. Fader y Hardie (2002) capturan comportamientos de este tipo, diferenciando en los modelos clientes del tipo *one-time buyers* (luego de una compra salen del mercado), del resto de los clientes, para lo cual agregan un parámetro adicional el cual corresponde a la probabilidad de ser un cliente de este tipo. En su trabajo ajustan los tiempos entre compras a distribuciones Exponenciales sin la aplicación de distribuciones NBD en los parámetros para los clientes *one-time buyers*, modelo que es exactamente el que posee mejor ajuste en este estudio.

El efecto numérico de las variables puede ser estudiado a partir de las elasticidades de los tiempos entre compras y tiempos censurados respecto a las variables del modelo. Estas se encuentran calculadas en la Tabla 5.8, con las consideraciones descritas en el Capítulo 4 respecto a las variables binarias y tipos de cajas.

5.6.2. Análisis de Resultados

Variables Transaccionales

Al igual que en los modelos sin inclusión de tiempos censurados, se estima un efecto contrario entre las variables asociadas a montos e items de las transacciones, donde la variable $Dollars_{ij-1}$ presenta un efecto negativo sobre la tasa de compras, y la variable $Items_{ij-1}$ un comportamiento positivo. Esta situación se presenta sólo cuando ambas variables son incluidas en la especificación, ya que al mantener sólo el monto en dólares como en la Iteración 3, se estima un efecto positivo de esta sobre la tasa, y por consecuencia positivo sobre los tiempos entre transacciones.

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0660*** (0.0139)		
α		22.1272*** (0.0369)	14.4092*** (0.0342)
r		1.4312*** (0.0271)	0.6315*** (0.0336)
c			1.3381*** (0.0084)
$\beta_{SelfCheckout}$	0.0256*** (0.0138)	0.0547*** (0.0203)	0.0522*** (0.0264)
$\beta_{Weekend}$	-0.0197*** (0.0086)	-0.0207** (0.0155)	-0.0307*** (0.0202)
$\beta_{QueueLenght}$	-0.0041*** (0.0011)	-0.0014 (0.0019)	-0.0030** (0.0025)
$\beta_{OpenQueueCheckCero}$	-0.2436*** (0.0841)	-0.1315* (0.1307)	-0.2096** (0.1842)
$\beta_{WaitTime}$	-0.0001* (0.0001)	-0.0002** (0.0002)	-0.0003*** (0.0002)
$\beta_{SuperBowl}$	-0.2877*** (0.1461)	-0.3212* (0.2990)	-0.3776* (0.4032)
$\beta_{RatioDollarItem}$	-0.0075*** (0.0011)	-0.0102*** (0.0019)	-0.0131*** (0.0023)
$-LL$	129,359	175,802	156,300
AIC	258,733	351,623	312,619
BIC	258,802	351,700	312,705

$n = 40,975$

Tabla 5.7: Modelos Tiempo Censurado Iteración 5

Variable	Exponencial	Gamma-Exponencial	Gamma-Weibull
<i>SelfCheckout</i>	2.56 %	5.47 %	5.22 %
<i>Weekend</i>	-1.97 %	-2.07 %	-3.07 %
<i>QueueLenght</i>	-2.94 %	-1.03 %	-2.12 %
<i>OpenQueueCheckCero</i>	-24.36 %	-13.15 %	-20.96 %
<i>WaitTime</i>	-0.33 %	-0.84 %	-1.34 %
<i>SuperBowl</i>	-28.77 %	-32.12 %	-37.76 %
<i>RatioDollarItem</i>	-3.75 %	-5.11 %	-6.58 %

Tabla 5.8: Elasticidades Modelos Tiempo Censurado Iteración 5

La posible interacción no lineal entre las variables asociadas a montos e items, se incluye en la variable $RatioDollarItem_{ij-1}$ la cual captura el monto promedio de los items comprados. El efecto estimado de esta variable es negativo sobre la tasa, y presenta un efecto de hasta un -6.58 % como se expresa en la Tabla 5.8. Comparativamente con los modelos estimados en el Capítulo 4 el efecto creciente se mantiene con la incorporación de complejidad a los modelos, pero la magnitud del crecimiento es mayor para los modelos que incluyen tiempos censurados.

Para la variable $SelfCheckout_{ij-1}$, asociada a compras realizadas en cajas autoservicio, se estima un efecto positivo sobre la tasa de compras y en consecuencia negativo sobre los tiempos entre compras, lo cual puede ser interpretado como un abastecimiento de corto plazo, y en consecuencia un pronto regreso a las tiendas de la cadena. Como se ilustra en la tabla de elasticidades, efecto alcanza hasta un 5.47 % en el modelo Gamma-Exponencial.

Variables de Niveles de Servicio

La variable asociada a total de personas en colas de cajas regulares, mantiene su significancia y signo respecto a lo estimado en los modelos del Capítulo 4, sin embargo presenta una disminución en la magnitud de los coeficientes y un cambio en el patrón creciente de este asociado a la complejidad de los modelos, donde la variable $CQueueLenght_{it-1}$ alcanza un efecto de hasta un -2.94 % sobre la tasa en el modelo exponencial, como se ilustra en la Tabla 5.8. La variable $CWaitTime_{it-1}$ en cambio, asociada a los tiempos de espera en cajas regulares, además de mantener la significancia y signo del efecto, mantiene el patrón creciente del impacto de la variable asociado a la complejidad de los modelos, con una leve disminución en el efecto sobre la tasa en todos los modelos, alcanzando un impacto de hasta un -1.34 % para el modelo Gamma-Weibull.

La variable $OpenQueueZero_{ij-1}$, asociada a una situación de cero cajas disponibles, presenta un cambio similar al estimado en la variable $CQueueLenght_{it-1}$, donde existe un cambio en el patrón magnitud-complejidad, y una disminución en el efecto sobre la tasa. Como se ilustra en la Tabla 5.8, la variable alcanza un impacto negativo de hasta un -24.36 % para el modelo Exponencial.

El resto de las variables estudiadas asociadas a niveles de servicio resultaron no ser significativas a lo largo de las iteraciones realizadas, lo cual coincide con lo estimado en los modelos sin inclusión de tiempos censurados. En general las estimaciones finales de ambos tipos de modelos coinciden en gran parte en significancia y efecto de las variables, lo que permite concluir de forma robusta el efecto negativo de las esperas asociadas a cajas de supermercado sobre el comportamiento de compra de los clientes, donde mayor niveles de esperas o personas en cola implican aumentos empíricos en los tiempos entre transacciones.

Modelo	Parámetros	χ^2	p-valor
Exponencial Nulo	1	-	-
Exponencial Iteración 5	8	49,873	0.0000
Gamma-Exponencial Nulo	2	-	-
Gamma-Exponencial Iteración 5	9	177	0.0000
Gamma-Weibull Nulo	3	-	-
Gamma-Weibull Iteración 5	10	11,473	0.0000

Tabla 5.9: Test Chi-Cuadrado de Ratios de Verosimilitud

Variables de Control

Los controles asociados a compras realizadas los fines de semana y en la fecha del Super Bowl mantienen el signo y el patron de crecimiento asociado a la complejidad de los modelos, sin embargo la variable $Weekend_{it-1}$ presenta un leve aumento en la magnitud de los coeficiente respecto a los modelos estimados en el Capítulo 4, al contrario de la variable $SuperBowl_{it-1}$ la cual presenta una leve disminución en los coeficientes estimados.

En general, el comportamiento de las variables de control, es bastante robusto en las iteraciones respecto a los modelos sin inclusión de tiempos de censura. Los resultados son similares en interpretación, magnitud, signo y significancia de los coeficientes, a o largo de todas las iteraciones, lo que en consecuencia implicó la exclusión de las mismas variables.

5.7. Comparación de Modelos

Para probar si la inclusión de parámetros, es estadísticamente significativa respecto al modelo nulo con inclusión de tiempos censurados, se calcula el Ratio de Verosimilitud, el cual aplica un test Chi-Cuadrado con grados de libertad asociados a la diferencia de parámetros entre el modelo final y el nulo. Como se ilustra en la Tabla 5.9, todos los modelos estimados, con inclusión de tiempo censurado y variables explicativas, son estadísticamente significativos.

Capítulo 6

Evaluación Económica

Los modelos estudiados en los capítulos anteriores, cuantifican numéricamente el efecto negativo de disminuciones en los niveles de servicio asociados a esperas en cajas. Los efectos estimados, pueden ser utilizados para evaluar económicamente los costos y beneficios asociados a la aplicación de medidas que disminuyan las esperas en cajas. Para efectos de este estudio, se evaluarán los costos y beneficios asociados a la habilitación de cajas adicionales, y los efectos que esta medida genera sobre las esperas y comportamientos de compra de los clientes de la cadena de supermercados.

6.1. Modelo de Colas

El sistema de cajas de una sala de ventas de supermercado, puede ser estudiado mediante el modelo M/M/s, también conocido como Erlang-C. Este modelo asume un proceso Markoviano en la llegada de clientes, y que el sistema es operado por c cajas independientes, las cuales poseen un tiempo de atención que distribuye exponencialmente.

Las tiendas de supermercado, por lo general combinan diferentes configuraciones de sistemas de cajas, como sistemas M/M/1 y M/M/s, los cuales a su vez pueden poseer o no características especiales, como por ejemplo las cajas preferenciales o con límite de productos. Por simplicidad, se asumirá una única cola, servida por múltiples cajas. Esta definición también permite utilizar directamente los registros de largos de colas utilizados en este estudio, los cuales contabilizan el total de personas en cola, y no de cajas específicas.

6.1.1. Modelo M/M/s

El modelo Erlang-C, permite obtener el largo de cola promedio del sistema partir de la tasa de llegada (λ), tasa de servicio (μ) y el número de servidores (s). Otras definiciones a utilizar corresponden a la probabilidad de no hacer fila (P_0) y la utilización del sistema (ρ).

Las especificaciones del modelo se describen a continuación.

$$L_q = P_0 \frac{(s\rho)^s}{s!} \frac{1}{1-\rho}$$

$$\frac{1}{P_0} = \sum_{k=0}^{s-1} \frac{s\rho^k}{k!} + \frac{s\rho^s}{s!} \frac{1}{1-\rho}$$

$$\rho = \frac{\lambda}{s\mu}$$

6.1.2. Elección de Escenarios

Se definió como unidad temporal de los parámetros del modelo, ventanas de una hora cronológica, donde se espera capturar comportamientos estables asociados al nivel de transacciones, dotación y largos de cola. Se seleccionaren 7 ventanas de la totalidad de tiendas y días registrados en el panel de datos, diferenciadas por hora y si corresponden a un día de semana o fin de semana, con esto se espera estudiar diferencias en función de las estacionalidades que afectan el número de clientes que visitan las tiendas de supermercado. Para la elección de las ventanas se verifico que la dotación de cajas habilitadas se mantuviera estable, con el fin de mitigar errores en las estimaciones. También se incluyeron escenarios donde se reflejen altas tasas de llegada en los clientes y varios servidores disponibles, para evaluar el impacto ante situaciones de alto flujo en el sistema.

6.1.3. Estimación de Tasa de Servicio

De los registros transaccionales es posible calcular la cantidad de transacciones realizadas en una unidad de tiempo. Este valor puede ser utilizado como la tasa de llegadas de clientes (λ) bajo un estado estacionario del sistema. El número de servidores (s) y los largos de cola promedio (L) pueden ser obtenidos a partir de los registros de niveles de servicio. La tasa de servicio de las cajas (μ) y por ende la utilización del sistema (ρ), no son posibles de obtener o estimar directamente de los datos. Para esto se utilizó el modelo Erlang-C, con la finalidad de obtener la tasa que minimice la diferencia entre el largo de cola teórico del modelo y el empírico obtenido de los registros de video.

Los valores asociados a cada escenario se encuentran detallados en la Tabla 6.1, donde las tasas de servicio estimadas, prestan valores valores más altos ante mayores largos de cola promedio, esto implicaría que los cajeros aceleran el ingreso de compras ante una mayor cantidad de personas esperando por ser atendidas. Este efecto también se observa en el trabajo de Yu, Olivares et al (2013).

Escenario	Hora	Fin de Semana	λ	s	L	μ
A	10	NO	9	3	3	3.6742
B	17	NO	19	5	7.5	4.1934
C	16	SI	20	6	8	3.6541
D	10	SI	12	4	4	3.5202
E	17	SI	60	9	18	6.9815
F	13	SI	44	10	16.5	4.6219
G	19	NO	63	8	30	8.1131

Tabla 6.1: Escenarios Sistemas de Colas

Escenario	s_1	L_1	$s_2 = s_1 + 1$	L_2
A	3	3	4	0.4748
B	5	7.5	6	1.3296
C	6	8	7	1.6088
D	4	4	5	0.7486
E	9	18	10	3.3904
F	10	16.5	11	3.5565
G	8	30	8	3.6673

Tabla 6.2: Largos de Cola Promedio Estimados ante Aumento en Dotación

6.1.4. Aumento de Dotación

La habilitación de cajas adicionales, genera una disminución en las esperas de los clientes, ya que se cuenta con mayores recursos para que estos sean atendidos. El efecto sobre los niveles de servicio, en particular el largo de cola promedio, puede ser estimado mediante el modelo Erlang-C, modificando la cantidad de servidores, y utilizando la tasa de servicio obtenida para cada escenario en las sección anterior. Los valores estimados se encuentran ilustrados en la Tabla 6.2.

6.2. Evaluación Económica

6.2.1. Cliente y Transacción Promedio

La evaluación económica se realizará sobre comportamientos promedios de clientes perteneciente al panel de datos con el cual se estimaron los modelos de duración, el monto en dólares de una compra promedio realizada en una caja regular es similar en magos paneles, por lo que se utilizará la media aritmética entre ambos valores, igual a \$51.913. De forma análoga para los tiempos entre compras promedio, se utilizará la media entre los valores de ambos paneles, obteniendo un valor igual a 17.332 días.

Modelo	Descripción	β_L
1	Gamma-Weibull sin Censura	-0.0059
2	Exponencial con Censura	-0.0041

Tabla 6.3: Coeficientes Asociados a Largos de Cola

Escenario	β_{L_1}	β_{L_2}	Cambio Porcentual ($\beta_{L_2} - \beta_{L_1}$)	Nueva Tasa Mensual
A	-1.77 %	-0.28 %	1.49 %	1.7577
B	-4.43 %	-0.78 %	3.64 %	1.7950
C	-4.72 %	-0.95 %	3.77 %	1.7972
D	-2.36 %	-0.44 %	1.92 %	1.7651
E	-10.62 %	-2.00 %	8.62 %	1.8812
F	-9.74 %	-2.10 %	7.64 %	1.8642
G	-17.70 %	-2.16 %	15.54 %	2.0010

Tabla 6.4: Cambio Porcentual y Nueva Tasa Mensual Modelo 1

Utilizando los valores promedio y definiendo como unidad temporal un mes de 30 días, se puede fijar una tasa promedio de visitas a la cadena de supermercados igual a 1.732 visitas al mes, y por ende un nivel mensual de transacciones promedio igual a \$89.907. Este último valor servirá como nivel base para evaluar el efecto que poseen los niveles de servicio sobre las compras de los clientes.

6.2.2. Impacto Económico del Aumento de Dotación

Para la evaluación económica se utilizan los modelos para los casos con y sin inclusión de censura. Los coeficientes estimados asociados a los largos de cola se encuentran ilustrados en la Tabla 6.3. Estos valores agregarán otra dimensión a los escenarios, definiendo un umbral en el impacto que posee esta variable sobre el comportamiento de compra.

Asumiendo que el resto de las variables permanece constante, se puede calcular el efecto que posee el cambio en el largo de cola promedio de cada escenario, sobre la tasa promedio de compras (1.732), para cada escenario y modelo, como se describe en las Tablas 6.4 y 6.5. Aplicando los porcentajes correspondientes se obtiene una nueva tasa esperada de compras, ante una disminución de los largos de cola en caja de los clientes, estos valores se encuentran descritos en la última columna de ambas tablas.

De forma análoga al cálculo realizado en la sección de Cliente y Transacción Promedio, se puede cuantificar un beneficio económico mensual esperado por mejoras en los niveles de

Escenario	β_{L_1}	β_{L_2}	Cambio Porcentual ($\beta_{L_2} - \beta_{L_1}$)	Nueva Tasa Mensual
A	-1.23 %	-0.19 %	1.04 %	1.7498
B	-3.08 %	-0.55 %	2.53 %	1.7757
C	-3.28 %	-0.66 %	2.62 %	1.7773
D	-1.64 %	-0.31 %	1.33 %	1.7550
E	-7.38 %	-1.39 %	5.99 %	1.8356
F	-6.77 %	-1.46 %	5.31 %	1.8238
G	-12.30 %	-1.50 %	10.80 %	1.9189

Tabla 6.5: Cambio Porcentual y Nueva Tasa Mensual Modelo 2

Escenario	Modelo 1	Modelo 2
A	\$1.3395	\$0.9308
B	\$3.2731	\$2.2745
C	\$3.3902	\$2.3559
D	\$1.7247	\$1.1985
E	\$7.7497	\$5.3854
F	\$6.8659	\$4.7712
G	\$13.9683	\$9.7068

Tabla 6.6: Beneficio Económico por Cliente

servicio. En este caso el monto de transacción promedio debe ser multiplicado por el diferencial en el número de compras mensuales esperadas, lo cual puede ser interpretado como un aumento en las compras de cada cliente en el corto plazo, generado por la experiencia de servicio de la última compra. Se definió un intervalo temporal de un mes para capturar el beneficio de la mejora en la tasa, ya que las eventuales próximas compras, también podrían modificar la tasa, ante grandes cambios en las variables significativas del modelo. La Tabla 6.6 describe el beneficio económico en el corto plazo por cliente, para cada escenario y modelo.

Con los beneficios monetarios por cliente calculados, se puede calcular el beneficio total del aumento de dotación, el cual corresponde al beneficio individual, multiplicado por la cantidad de clientes que realizaron transacciones en cada escenario. Este valor representa al aumento en ventas en el corto plazo generado por la disminución de los largos de cola. Para realizar un análisis en términos de utilidad, es necesario considerar los costos asociados a los productos. Para esto se definen tres escenarios de posibles utilidad sobre la venta, en base a los valores esperados para la industria de los supermercados (3 %, 5 % y 7 %). Las Tablas 6.7 y 6.8 describen el beneficio neto esperado en términos de ventas para los diferentes escenarios y porcentajes de utilidad.

Escenario	Beneficio Total	3 %	5 %	7 %
A	\$12.0555	\$0.3617	\$0.6028	\$0.8439
B	\$62.1890	\$1.8657	\$3.1095	\$4.3532
C	\$67.8046	\$2.0341	\$3.3902	\$4.7463
D	\$20.6966	\$0.6209	\$1.0348	\$1.4488
E	\$464.9822	\$13.9495	\$23.2491	\$32.5488
F	\$302.1003	\$9.0630	\$15.1050	\$21.1470
G	\$880.0001	\$26.4000	\$44.0000	\$61.6000

Tabla 6.7: Beneficio Total Neto por Escenario - Modelo 1

Escenario	Beneficio Total	3 %	5 %	7 %
A	\$8.3775	\$0.2513	\$0.4189	\$0.5864
B	\$43.2161	\$1.2965	\$2.1608	\$3.0251
C	\$47.1185	\$1.4136	\$2.3559	\$3.2983
D	\$14.3824	\$0.4315	\$0.7191	\$1.0068
E	\$323.1232	\$9.6937	\$16.1562	\$22.6186
F	\$209.9341	\$6.2980	\$10.4967	\$14.6954
G	\$611.5255	\$18.3458	\$30.5763	\$42.8068

Tabla 6.8: Beneficio Total Neto por Escenario - Modelo 2

Escenario	λ	s	μ	ρ
A	9	3	3.6742	0.8165
B	19	5	4.1934	0.9062
C	20	6	3.6541	0.9122
D	12	4	3.5202	0.8522
E	60	9	6.9815	0.9549
F	44	10	4.6219	0.9520
G	63	8	8.1131	0.9707

Tabla 6.9: Utilización por Escenario

El costo asociado a habilitar una caja adicional puede ser aproximado al salario por hora de un cajero en la industria. Considerando que el salario mínimo bordea los \$10 dólares por hora, se puede comparar este valor con el beneficio total neto, para evaluar si se justifica la apertura de una caja adicional en la tienda. Para el caso de los escenarios evaluados, se puede observar que aquellos donde sería rentable abrir una nueva caja son los E,F y G, ya que poseen un beneficio superior al costo de un cajero.

Los escenarios rentables tienen un factor común en comparación al resto, y es que representan escenarios más estresados. En la Tabla 6.9 se encuentran calculadas las utilidades de los sistemas correspondientes, y se puede apreciar que los escenarios E,F y G están más cercanos a una utilización del 100% lo cual se traduciría en un colapso del sistema.

La metodología desarrollada permite tomar decisiones sobre la dotación en función de las características del sistema en análisis. Como se aprecia en los resultados anteriores, mientras más alta es la utilización del sistema, se justifica en mayor medida la apertura de nuevas cajas, lo cual puede ser evaluado en términos monetarios utilizando los efectos estimados por los modelos de duración.

Capítulo 7

Conclusiones

A raíz de los resultados obtenidos, se puede concluir que los modelos probabilísticos Exponencial, Gamma-Exponencial y Gamma-Weibull permiten la estimación de impactos de variables de niveles de servicio sobre el comportamiento de compra de los clientes. La metodología de preparación de datos para el modelado permitió la elaboración de un panel de clientes con información consolidada a nivel de transacción para el análisis de impactos y significancia estadística de las diferentes variables propuestas.

Todos los modelos seleccionados, poseen significancia estadística respecto a los modelos nulos estimados en primera instancia, testados con ratios de verosimilitud (LR), el cual corrobora estadísticamente la intuición de que la inclusión de mayor información, permite realizar mejores estimaciones.

Los modelos estimados, comprueban empíricamente un efecto negativo de los niveles de servicio asociados a esperas en cajas de supermercados, afectando negativamente la tasa de compras de los clientes, en función de la experiencia de servicio recibida en la última compra. Estos efectos fueron controlados con otras variables asociadas a montos transaccionales, tipos de cajas y variables estacionales. De esta forma se mitigan sesgos en la estimación de los parámetros mediante una diversificación de las variables incluidas.

Los largos de cola resultaron tener un mayor efecto que los tiempos de espera, lo cual puede ser asociado al factor visual de las aglomeraciones, lo cual también es descrito por Mehta (2013), quien plantea que las personas son susceptibles a proyectar sus tiempos de espera en función de la cantidad de personas que observan en las filas, más allá, de los posibles tiempos esperados reales a los que serían expuestas.

Un hallazgo relevante es el efecto de presenciar un cierre de todas las cajas regulares. Esta situación también puede ser considerado un nivel de servicio, el cual genera un impacto de hasta un -35.21 % sobre la tasa de compras, y en consecuencia un aumento del tiempo promedio en un tercio para la próxima visita del cliente.

La inclusión de tiempos censurados en los modelos, correspondientes al tiempo transcurrido entre la última transacción y la fecha límite disponible en las fuentes de datos transaccionales, permitió robustecer las estimaciones, y corroborar la significancia de las variables seleccionadas. La gran diferencia de estos modelos, respecto a los que no incluyen tiempos censurados, resultó ser el cambio en el patrón de mejor ajuste en medida que se agrega complejidad y flexibilidad a los modelos, ya que originalmente el Modelo Gamma-Weibull poseía los mejores índices de bondad, y para los censurados, el mejor ajuste se obtiene mediante el modelo Exponencial. La explicación a esta situación, es la incorporación en el panel de datos, de posibles salidas de clientes del mercado, donde estos pasan por largos periodos de inactividad antes de volver a realizar transacciones en la cadena de supermercados, generando un impacto en la variable dependiente de los modelos, incluyendo mayores tiempos entre compras, y en consecuencia menores tasas esperadas de compras.

Las iteraciones de los modelos presentan un comportamiento robusto para todos los modelos, donde la mayoría de las variables mantienen relativamente estables sus efectos sobre la tasa, ante modificaciones en el número de variables incluidas en las especificaciones. Este efecto también se puede apreciar en la significancia de las variables más importantes, las cuales mantiene su poder productivo en las diferentes iteraciones.

El objetivo general de este trabajo consistía en estimar impactos de las variables de servicio en términos económicos, para esto se seleccionaron diferentes escenarios para realizar una evaluación económica de la habilitación de nuevas cajas. En otras palabras, se realizó un análisis costo beneficio, de los costos asociados a la dotación de personal, en comparación con los beneficios esperados, generados a partir de la disminución en las esperas, coaccionados por la apertura de una caja adicional.

A partir de las fuentes de datos era posible obtener información asociada a el flujo de clientes, dotación y tiempos de espera, no así de los tiempos de servicio de los cajeros. Por ello se aplicaron modelos de teoría de colas, específicamente un modelo M/M/s, el cual permitió caracterizar completamente los sistemas de cada escenario, y analizar impactos en las esperas generados por la apertura de cajas.

Mediante un análisis de los modelos estimados, características de los sistemas y escenarios de utilidad de la empresa, se llegó a la conclusión de que es factible evaluar económicamente el impacto de los niveles de servicio, y en particular de los largos de cola promedio. Donde se logro generar una metodología para la toma de decisión de apertura de una nueva caja en función de los costos y beneficios asociados.

Los escenarios correspondientes a sistemas más estresados, evidencian una mayor justificación de apartara de cajas en función de mayor beneficios económicos esperados, a raíz de mejorar la experiencia de espera de los clientes. Para los escenarios más estabilizados, la metodología justifica la manutención de dotación disponible. El uso de los hallazgos obtenidos

a partir de este trabajo, podrían apoyar la gestión de las salas de ventas y sus clientes, para utilizar los niveles de servicio, para generar ventajas competitivas.

7.1. Limitaciones y Trabajo Futuro

La metodología desarrollada en este trabajo posee varios supuestos declarados, los cuales podrían ser abordados mediante información complementaria o de mejor calidad. En particular se identifican los registros de tiempos de espera, los cuales por su costo asociado poseen una baja periodicidad de registros. Generar mayor información de este nivel de servicio podría permitir la elaboración de mejores estimaciones, y de prescindir del supuesto de colas M/M/s, y modelar el sistema más acorde a la realidad, como un conjunto de sistemas M/M/1. Otra información que podría ayudar a la elaboración de mejores estimaciones sería contar con registros de las tasas de llegadas de clientes a las cajas, de esta forma se podría generar un modelo de colas en base a simulaciones, y poder estimar de mejor forma las tasas de servicio en función de características del sistema.

Los modelos con tiempo censurado podrían ser mejorados mediante la incorporación de un parámetro relativo al estado del cliente, el cual permita identificar estados de inactividad, para no afectar los efectos de las variables sobre las tasas de compras. Jerath, Fader y Hardie (2011) incluyen esta diferenciación, la cual podría mejorar el ajuste de las especificaciones.

Existe evidencia empírica de que variables demográficas pueden explicar el comportamiento de compra, en este estudio se incluyó la distancia aproximada de los hogares a las tiendas, la cual resultó no ser significativa. Complementar este análisis con otras variables podía mejorar los ajustes de los modelos (Fox y Thomas 2006).

Las herramientas estadísticas y de gestión de operaciones utilizadas en este trabajo, permitieron elaborar una metodología que permite justificar niveles de dotación en términos monetarios, estas estimaciones podrían ser adaptadas a realidades diferenciadas, como por ejemplo tiendas y horarios particulares, y utilizadas para la gestión diaria de las salas de venta, en función de las características de sus clientes.

Bibliografía

- [1] Sunil Gupta, 1992. *Stochastic Models of inter purchase Time With Time-Dependent covariates*, Journal of Marketing Research, Vol. XXVIII: 1-15.
- [2] P.N. Seetharaman, Pradeep K. Chintagunta, 2003. *The Proportional Hazards Model for Purchase Timing: A Comparison of Alternative Specifications*, Journal of Business & Economic Statistics, Vol.21, No.3.
- [3] Yina Lu, Marcelo Olivares, Andrés Musalem, Ariel Schilkrut, 2013. *Measuring the Effect of Queues on Customer Purchases*, Management Science, Vol.59, No.8, pp. 1743-1763.
- [4] Van den Bulte, Iyengar, 2011. *Tricked by Truncation: Spurious Duration Dependence and Social Contagion in Hazard Models* Marketing Science 30(2) : 233-248
- [5] Cox, D.R. & Oakes, 1984. *Analysis of Survival Data*, London: Chapman and Hall.
- [6] Cox, D.R., 1972. *The competitive impact of service process improvement: Examining customers' waiting experiences in retail markets*, Journal of the Royal Statistical Society, Series B 34: 187-220.
- [7] Piyush Kumar, 2005. *Regression Models and Life Tables (with Discussion)*, Journal of Retailing 81 (3, 2005) 171–180.
- [8] Ritu Mehta, 2013. *Understanding perceived retail crowding: A critical review and research agenda*, Journal of Retailing and Consumer Services 20: 642–649
- [9] Fader, Hardie, 2002. *A note on an integrated model of customer buying behavior*, European Journal of Operational Research 139: 682–687
- [10] Fader, Hardie, 2009. *Probability Models for Customers-Base Analysis*, European Journal of Operational Research 139: 682–687
- [11] Jerath, Fader y Hardie, 2011. *New Perspectives on Customer “Death” Using a Generalization of the Pareto/NBD Model*, Marketing Science 30, No. 5: 866–880
- [12] Morrison, Schmittlein, 1980. *Jobs, Strikes, and Wars: Probability Models for Duration*, Journal of Interactive Marketing 23: 61–69

- [13] Fox, Thomas, 2006. *Predicting Retail Customers Share of Wallet Using Shopper Loyalty Card Data*, Working paper, Marketing Department, Southern Methodist University, Dallas
- [14] Ehrenberg, 1959. *The Pattern of Consumer Purchases*, Journal of the Royal Statistical Society 8: 26-4
- [15] Zaman, Pfeiffer, 2011. *Survival Analysis in Medical Research*, Interstat. 17:1-36

Apéndice A

Iteraciones Modelos sin Tiempo Censurado

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0580 (0.0379)		
α		60.0947*** (0.0568)	34.7866*** (0.0612)
r		4.4083*** (0.0280)	1.9149*** (0.0331)
c			1.2024*** (0.0083)
β_{Item}	0.0042*** (0.0011)	0.0041*** (0.0014)	0.0045*** (0.0017)
$\beta_{SelfCheckout}$	0.0244 (0.0430)	0.0212 (0.0513)	0.0009 (0.0587)
$\beta_{Dollars}$	-0.0007*** (0.0002)	-0.0007*** (0.0003)	-0.0008*** (0.0004)
$\beta_{Weekend}$	-0.0334*** (0.0168)	-0.0349** (0.0202)	-0.0413** (0.0238)
$\beta_{OpenQueue}$	-0.0029 (0.0072)	-0.0007 (0.0087)	-0.0029 (0.0103)
$\beta_{QueueLenght}$	-0.0038 (0.0046)	-0.0049* (0.0056)	-0.0045 (0.0067)
$\beta_{QueueLenghtSelfCheck}$	0.0001 (0.0041)	0.0028 (0.0050)	0.0046 (0.0060)
$\beta_{QueueLenghtInd}$	-0.0016 (0.0248)	0.0035 (0.0302)	-0.0049 (0.0356)
$\beta_{OpenQueueCheckCero}$	-0.2615*** (0.0555)	-0.2768*** (0.0640)	-0.3636*** (0.034)
$\beta_{WaitTime}$	-0.0001 (0.0001)	-0.0002** (0.0002)	-0.0003** (0.0002)
$\beta_{Distancia}$	0.0001*** (0.0000)	0.0001*** (0.0000)	0.0002*** (0.0000)
$\beta_{Anteferiado}$	0.0019 (0.0358)	0.0099 (0.0439)	0.0073 (0.0517)
$\beta_{Feriado}$	0.0349* (0.0386)	0.0282 (0.0456)	0.0255 (0.0541)
$\beta_{SuperBowl}$	-0.2855** (0.1061)	-0.3020** (0.1227)	-0.3661** (0.1287)
$\beta_{WeekendCheckout}$	0.0158 (0.0239)	0.0181 (0.0289)	0.0216 (0.0349)
$-LL$	158,460	157,504	157,249
AIC	316,952	315,042	314,535
BIC	317,090	315,189	314,690

$n = 41,300$

Tabla A.1: Modelos Iteración 1

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0586*** (0.0103)		
α		56.213*** (0.0348)	35.0175*** (0.0300)
r		4.2560*** (0.0271)	1.8973*** (0.0328)
c			1.2042*** (0.0083)
β_{Item}	0.0041*** (0.0011)	0.0041*** (0.0014)	0.0044*** (0.0017)
$\beta_{Dollars}$	-0.0007*** (0.0002)	-0.0007*** (0.0003)	-0.0008*** (0.0004)
$\beta_{Weekend}$	-0.0274*** (0.0166)	-0.0309*** (0.0199)	-0.0382*** (0.0233)
$\beta_{QueueLenght}$	-0.0059*** (0.0013)	-0.0055*** (0.0015)	-0.0063*** (0.0018)
$\beta_{QueueLenghtSelfCheck}$	0.0045** (0.0032)	0.0053** (0.0040)	0.0062** (0.0047)
$\beta_{OpenQueueCheckCero}$	-0.2638*** (0.0516)	-0.2816*** (0.0592)	-0.3497*** (0.0612)
$\beta_{WaitTime}$	-0.0001* (0.0001)	-0.0002** (0.0002)	-0.0003*** (0.0002)
$\beta_{Distancia}$	0.0001*** (0.0000)	0.0001*** (0.0000)	0.0002*** (0.0000)
$\beta_{Feriado}$	0.0350* (0.0386)	0.0278 (0.0458)	0.0265 (0.0543)
$\beta_{SuperBowl}$	-0.2767** (0.1081)	-0.2988** (0.1242)	-0.3892** (0.1226)
$\beta_{WeekendCheckout}$	0.0072 (0.0227)	0.0122 (0.0277)	0.0182 (0.0335)
$-LL$	158,468	157,508	157,250
AIC	316,961	315,041	314,528
BIC	317,064	315,153	314,649

$n = 41,300$

Tabla A.2: Modelos Iteración 2

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0586*** (0.0103)		
α		55.0512*** (0.0343)	35.0795*** (0.0300)
r		4.1857*** (0.0266)	1.8994*** (0.0329)
c			1.2044*** (0.0083)
β_{Item}	0.0041*** (0.0011)	0.0041*** (0.0014)	0.0044*** (0.0017)
$\beta_{Dollars}$	-0.0007*** (0.0002)	-0.0007*** (0.0003)	-0.0008*** (0.0004)
$\beta_{Weekend}$	-0.0237*** (0.0111)	-0.0245*** (0.0135)	-0.0288*** (0.0160)
$\beta_{QueueLenght}$	-0.0058*** (0.0012)	-0.0054*** (0.0015)	-0.0060*** (0.0018)
$\beta_{QueueLenghtSelfCheck}$	0.0044** (0.0031)	0.0046** (0.0038)	0.0057** (0.0046)
$\beta_{OpenQueueCheckCero}$	-0.2611*** (0.0944)	-0.2859*** (0.1138)	-0.3550*** (0.1369)
$\beta_{WaitTime}$	-0.0001* (0.0001)	-0.0002** (0.0001)	-0.0003** (0.0002)
$\beta_{Distancia}$	0.0001*** (0.0000)	0.0001*** (0.0000)	0.0002*** (0.0000)
$\beta_{SuperBowl}$	-0.2786** (0.2047)	-0.2778** (0.2453)	-0.3676** (0.2993)
$-LL$	158,471	157,510	157,251
AIC	316,962	315,042	314,527
BIC	317,048	315,137	314,630

$n = 41,300$

Tabla A.3: Modelos Iteración 3

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0624*** (0.0077)		
α		48.3964*** (0.0325)	30.5647*** (0.0284)
r		3.9727*** (0.0258)	1.8572*** (0.0329)
c			1.1967*** (0.0084)
$\beta_{Dollars}$	0.0001* (0.0001)	0.0001 (0.0001)	0.0000 (0.0002)
$\beta_{Weekend}$	-0.0217*** (0.0105)	-0.0224*** (0.0129)	-0.0256*** (0.0152)
$\beta_{QueueLenght}$	-0.0059*** (0.0011)	-0.0053*** (0.0013)	-0.0058*** (0.0016)
$\beta_{OpenQueueCheckCero}$	-0.2703*** (0.0507)	-0.2957*** (0.0592)	-0.3690*** (0.0612)
$\beta_{WaitTime}$	-0.0001** (0.0001)	-0.0002** (0.0001)	-0.0003*** (0.0002)
$\beta_{SuperBowl}$	-0.2912*** (0.1044)	-0.3183** (0.1200)	-0.3902** (0.1223)
$-LL$	158,976	157,922	157,696
AIC	317,966	315,869	315,410
BIC	318,026	315,929	315,487

$n = 41,300$

Tabla A.4: Modelos Iteración 4

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0617*** (0.0073)		
α		49.0193*** (0.0326)	30.2255*** (0.0276)
r		3.9921*** (0.0260)	1.8024*** (0.0320)
c			1.2039*** (0.0084)
$\beta_{Dollars}$	0.0001* (0.0001)	0.0000 (0.0001)	0.0000 (0.0002)
$\beta_{QueueLenght}$	-0.0063*** (0.0011)	-0.0056*** (0.0013)	-0.0061*** (0.0016)
$\beta_{OpenQueueCheckCero}$	-0.2666*** (0.0511)	-0.2888*** (0.0588)	-0.3614*** (0.0605)
$\beta_{WaitTime}$	-0.0001* (0.0001)	-0.0002** (0.0001)	-0.0003*** (0.0002)
$\beta_{SuperBowl}$	-0.3039*** (0.1015)	-0.3322** (0.1162)	-0.4078** (0.1184)
$-LL$	158,985	157,928	157,704
AIC	317,982	315,871	315,423
BIC	318,034	315,931	315,429

$n = 41,300$

Tabla A.5: Modelos Iteración 5

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0622*** (0.0076)		
α		48.2650*** (0.0325)	30.7469*** (0.0283)
r		3.9661*** (0.0258)	1.8621*** (0.0328)
c			1.1951*** (0.0084)
$\beta_{Dollars}$	0.0001* (0.0001)	0.0001 (0.0001)	0.0000 (0.0002)
$\beta_{Weekend}$	-0.0211*** (0.0105)	-0.0216*** (0.0129)	-0.0240*** (0.0152)
$\beta_{QueueLenght}$	-0.0063*** (0.0010)	-0.0061*** (0.0012)	-0.0070*** (0.0015)
$\beta_{OpenQueueCheckCero}$	-0.2678*** (0.0506)	-0.2993*** (0.0574)	-0.3729*** (0.0583)
$\beta_{SuperBowl}$	-0.2910** (0.1047)	-0.3153** (0.1209)	-0.3655** (0.1290)
$-LL$	158,979	157,927	157,703
AIC	317,969	315,868	315,422
BIC	318,021	315,929	315,491

$n = 41,300$

Tabla A.6: Modelos Iteración 6

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0606*** (0.0102)		
α		51,1512*** (0.0341)	32.0038*** (0.0304)
r		4.0550*** (0.0263)	1.8784*** (0.0330)
c			1.1949*** (0.0084)
β_{Item}	0.0046*** (0.0011)	0.0048*** (0.0014)	0.0050*** (0.0018)
$\beta_{Dollars}$	-0.0007*** (0.0002)	-0.0008*** (0.0003)	-0.0009*** (0.0004)
$\beta_{Weekend}$	-0.0222*** (0.0105)	-0.0230*** (0.0129)	-0.0273*** (0.0151)
$\beta_{QueueLenght}$	-0.0053*** (0.0012)	-0.0048*** (0.0015)	-0.0053*** (0.0018)
$\beta_{QueueLenghtSelfCheck}$	0.0047*** (0.0032)	0.0049** (0.0039)	0.0052** (0.0046)
$\beta_{OpenQueueCheckCero}$	-0.2649*** (0.0515)	-0.2854*** (0.0593)	-0.3563*** (0.0607)
$\beta_{WaitTime}$	-0.0001* (0.0001)	-0.0002** (0.0001)	-0.0003*** (0.0002)
$\beta_{SuperBowl}$	-0.2971*** (0.1030)	-0.3247** (0.1180)	-0.3940** (0.1209)
$-LL$	158,925	157,889	157,668
AIC	317,867	315,797	315,358
BIC	317,945	315,883	315,452
$n = 41,300$			

Tabla A.7: Modelos Iteración 7

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0611*** (0.0101)		
α		49.4399*** (0.0336)	31.3191*** (0.0302)
r		3.9758*** (0.0258)	1.8566*** (0.0328)
c			1.1969*** (0.0084)
$\beta_{Dollars}$	0.0001** (0.0001)	0.0001 (0.0001)	0.0001 (0.0002)
$\beta_{Weekend}$	-0.0228*** (0.0105)	-0.0232*** (0.0129)	-0.0273*** (0.0152)
$\beta_{QueueLenght}$	-0.0053*** (0.0012)	-0.0047*** (0.0015)	-0.0052*** (0.0018)
$\beta_{QueueLenghtSelfCheck}$	0.0044** (0.0032)	0.0046** (0.0039)	0.0050* (0.0046)
$\beta_{OpenQueueCheckCero}$	-0.2631*** (0.0516)	-0.2868*** (0.0593)	-0.3583*** (0.0605)
$\beta_{WaitTime}$	-0.0001 (0.0001)	-0.0002** (0.0002)	-0.0003* (0.0002)
$\beta_{SuperBowl}$	-0.2961*** (0.1035)	-0.3238** (0.1186)	-0.3907** (0.1221)
$-LL$	158,970	157,918	157,692
AIC	317,956	315,853	315,405
BIC	318,025	315,931	315,491

$n = 41, 300$

Tabla A.8: Modelos Iteración 8

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0674*** (0.0118)		
α		46.7431*** (0.0350)	28.1334*** (0.0314)
r		4.1129*** (0.0266)	1.8648*** (0.0327)
c			1.1974*** (0.0083)
$\beta_{Weekend}$	-0.0196*** (0.0106)	-0.0202*** (0.0129)	-0.0235*** (0.0152)
$\beta_{QueueLenght}$	-0.0050*** (0.0012)	-0.0045*** (0.0015)	-0.0052*** (0.0017)
$\beta_{QueueLenghtSelfCheck}$	0.0038** (0.0031)	0.0043* (0.0039)	0.0046* (0.0046)
$\beta_{OpenQueueCheckCero}$	-0.2520*** (0.0530)	-0.2749*** (0.0605)	-0.3398*** (0.0629)
$\beta_{WaitTime}$	-0.0001* (0.0001)	-0.0002** (0.0001)	-0.0003*** (0.0002)
$\beta_{SuperBowl}$	-0.2797** (0.1073)	-0.3124** (0.1207)	-0.3826** (0.1242)
$\beta_{RatioDollarItem}$	-0.0075*** (0.0014)	-0.0077*** (0.0016)	-0.0093*** (0.0019)
$-LL$	158,887	157,854	157,628
AIC	317,790	315,726	315,275
BIC	317,859	315,804	315,362

$n = 41,300$

Tabla A.9: Modelos Iteración 9

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0616*** (0.0077)		
α		51.3196*** (0.0335)	30.7151*** (0.0284)
r		4.1279*** (0.0267)	1.8396*** (0.0327)
c			1.1999*** (0.0084)
β_{Item}	0.0014*** (0.0005)	0.0013*** (0.0006)	0.0012*** (0.0007)
$\beta_{Weekend}$	-0.0215*** (0.0105)	-0.0227*** (0.0128)	-0.0271*** (0.0152)
$\beta_{QueueLenght}$	-0.0064*** (0.0011)	-0.0059*** (0.0013)	-0.0066*** (0.0016)
$\beta_{OpenQueueCheckCero}$	-0.2730*** (0.0503)	-0.2965*** (0.0575)	-0.3668*** (0.0594)
$\beta_{WaitTime}$	-0.0001** (0.0001)	-0.0002** (0.0001)	-0.0003*** (0.0002)
$\beta_{SuperBowl}$	-0.2899** (0.1050)	-0.3166** (0.1198)	-0.3898** (0.1228)
$-LL$	158,958	157,911	157,689
AIC	317,929	315,837	315,396
BIC	317,990	315,906	315,474

$n = 41,300$

Tabla A.10: Modelos Iteración 11

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0627*** (0.0071)		
α		48.3610*** (0.0325)	30.6003*** (0.0280)
r		4.0031*** (0.0260)	1.8642*** (0.0330)
c			1.1964*** (0.0084)
$\beta_{Weekend}$	-0.0201*** (0.0105)	-0.0206*** (0.0129)	-0.0251*** (0.0155)
$\beta_{QueueLenght}$	-0.0056*** (0.0011)	-0.0053*** (0.0013)	-0.0058*** (0.0015)
$\beta_{OpenQueueCheckCero}$	-0.2659*** (0.0511)	-0.2912*** (0.0585)	-0.3591*** (0.0602)
$\beta_{WaitTime}$	-0.0001* (0.0001)	-0.0002** (0.0001)	-0.0003*** (0.0002)
$\beta_{SuperBowl}$	-0.2958*** (0.1036)	-0.3207** (0.1193)	-0.3932** (0.1214)
$-LL$	158,979	157,923	157,696
AIC	317,970	315,860	315,409
BIC	318,022	315,921	315,478

$n = 41,300$

Tabla A.11: Modelos Iteración 12

Apéndice B

Iteraciones Modelos con Tiempo Censurado

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0582* (0.0296)		
α		29.8962*** (0.0570)	21.3274*** (0.0701)
r		1.4824*** (0.0171)	0.6504*** (0.0119)
c			1.3391*** (0.0026)
β_{Item}	0.0042*** (0.0009)	0.0067*** (0.0017)	0.0078*** (0.0022)
$\beta_{SelfCheckout}$	0.0233 (0.0335)	0.0797** (0.0676)	0.0855** (0.0888)
$\beta_{Dollars}$	-0.0007*** (0.0002)	-0.0009*** (0.0003)	-0.0012*** (0.0005)
$\beta_{Weekend}$	-0.0340*** (0.0130)	-0.0561*** (0.0221)	-0.0701*** (0.0277)
$\beta_{OpenQueue}$	-0.0027 (0.0057)	0.0008 (0.0101)	0.0027 (0.0131)
$\beta_{QueueLenght}$	-0.0037* (0.0037)	-0.0041 (0.0065)	-0.0059 (0.0083)
$\beta_{QueueLenghtSelfCheck}$	0.0001 (0.0032)	0.0033 (0.0058)	0.0061 (0.0075)
$\beta_{QueueLenghtInd}$	-0.0035 (0.0195)	0.0104 (0.0354)	0.0130 (0.0463)
$\beta_{OpenQueueCheckCero}$	-0.2632*** (0.0485)	-0.1302* (0.1070)	-0.1940* (0.1225)
$\beta_{WaitTime}$	-0.0001** (0.0001)	-0.0002** (0.0002)	-0.0003*** (0.0002)
$\beta_{Distancia}$	0.0001*** (0.0000)	0.0001*** (0.0000)	0.0002*** (0.0000)
$\beta_{Anteferiado}$	0.0002 (0.0262)	-0.0482* (0.0439)	-0.0505 (0.0572)
$\beta_{Feriado}$	0.0349* (0.0285)	-0.0137 (0.0479)	-0.0098 (0.0635)
$\beta_{SuperBowl}$	-0.2882*** (0.0762)	-0.3185* (0.1430)	-0.3736* (0.1678)
$\beta_{WeekendCheckout}$	0.0171* (0.0186)	0.0438** (0.0352)	0.0474** (0.0461)
$-LL$	129,158	175,357	155,924
AIC	258,349	350,748	311,884
BIC	258,487	350,894	312,039

$n = 40,975$

Tabla B.1: Modelos Tiempo Censurado Iteración 1

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0570*** (0.0102)		
α		29.8162*** (0.0294)	20.6635*** (0.0296)
r		1.4941*** (0.0172)	0.6556*** (0.0120)
c			1.3356*** (0.0026)
β_{Item}	0.0041*** (0.0009)	0.0068*** (0.0017)	0.0078*** (0.0022)
$\beta_{SelfCheckout}$	0.0302*** (0.0166)	0.0726*** (0.0325)	0.0686*** (0.0418)
$\beta_{Dollars}$	-0.0006*** (0.0002)	-0.0009*** (0.0003)	-0.0012*** (0.0005)
$\beta_{Weekend}$	-0.0337*** (0.0130)	-0.0568*** (0.0220)	-0.0715*** (0.0276)
$\beta_{QueueLenght}$	-0.0048*** (0.0011)	-0.0026** (0.0019)	-0.0040*** (0.0025)
$\beta_{QueueLenghtSelfCheck}$	0.0000 (0.0032)	0.0033 (0.0058)	0.0063 (0.0075)
$\beta_{OpenQueueCheckCero}$	-0.2538*** (0.0470)	-0.1559** (0.0948)	-0.2350** (0.1049)
$\beta_{WaitTime}$	-0.0001** (0.0001)	-0.0002** (0.0002)	-0.0003*** (0.0002)
$\beta_{Distancia}$	0.0001*** (0.0000)	0.0002*** (0.0000)	0.0002*** (0.0000)
$\beta_{SuperBowl}$	-0.2868*** (0.0763)	-0.3081* (0.1459)	-0.3789* (0.1651)
$\beta_{WeekendCheckout}$	0.0131 (0.0183)	0.0436** (0.0350)	0.0483** (0.0459)
$-LL$	129,170	175,360	155,925
AIC	258,364	350,746	311,878
BIC	258,468	350,858	311,999

$n = 40,975$

Tabla B.2: Modelos Tiempo Censurado Iteración 2

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0597*** (0.0100)		
α		25.5849*** (0.0291)	17.4174*** (0.0294)
r		1.4191*** (0.0167)	0.6372*** (0.0119)
c			1.3325*** (0.0027)
$\beta_{SelfCheckout}$	0.0267*** (0.0121)	0.0627*** (0.0239)	0.0594*** (0.0308)
$\beta_{Dollars}$	0.0002*** (0.0001)	0.0004*** (0.0002)	0.0003*** (0.0002)
$\beta_{Weekend}$	-0.0229*** (0.0082)	-0.0266*** (0.0146)	-0.0363*** (0.0185)
$\beta_{QueueLenght}$	-0.0042*** (0.0011)	-0.0019** (0.0019)	-0.0030*** (0.0025)
$\beta_{OpenQueueCheckCero}$	-0.2544*** (0.0470)	-0.1545** (0.0963)	-0.2166** (0.1110)
$\beta_{WaitTime}$	-0.0001** (0.0001)	-0.0002** (0.0002)	-0.0003*** (0.0002)
$\beta_{SuperBowl}$	-0.3031*** (0.0736)	-0.3185* (0.1436)	-0.3963* (0.1604)
$-LL$	129,391	175,874	156,363
AIC	258,797	351,767	312,745
BIC	258,866	351,844	312,832

$n = 40, 975$

Tabla B.3: Modelos Tiempo Censurado Iteración 3

Parámetro	Exponencial	Gamma-Exponencial	Gamma-Weibull
λ	0.0607*** (0.0094)		
α		23.5177*** (0.0208)	16.5120*** (0.0278)
r		1.3859*** (0.0164)	0.6237*** (0.0117)
c			1.3416*** (0.0027)
$\beta_{SelfCheckout}$	0.0224*** (0.0118)	0.0478*** (0.0228)	0.0500*** (0.0297)
$\beta_{Weekend}$	-0.0218*** (0.0082)	-0.0234*** (0.0148)	-0.0334*** (0.0187)
$\beta_{QueueLenght}$	-0.0042*** (0.0011)	-0.0020* (0.0019)	-0.0030** (0.0025)
$\beta_{OpenQueueCheckCero}$	-0.2550*** (0.0469)	-0.1385* (0.1003)	-0.2210** (0.1110)
$\beta_{WaitTime}$	-0.0001* (0.0001)	-0.0002** (0.0002)	-0.0003*** (0.0002)
$\beta_{SuperBowl}$	-0.3049*** (0.0734)	-0.3300* (0.1410)	-0.3957* (0.1620)
$-LL$	129,379	175,892	156,365
AIC	258,772	351,800	312,748
BIC	258,832	351,869	312,825

$n = 40,975$

Tabla B.4: Modelos Tiempo Censurado Iteración 4

