



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

**ESTIMACIÓN DEL IMPACTO DE MEJORAS A LA CALIDAD DE SERVICIO
DE UNA EMPRESA DE DISTRIBUCIÓN LOGÍSTICA EN EL
COMPORTAMIENTO DE SUS CLIENTES**

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTIÓN DE OPERACIONES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

ANGÉLICA DE JESÚS GALINDO VALDÉS

PROFESOR GUÍA:
ANDRÉS IGNACIO MUSALEM SAID

MIEMBROS DE LA COMISIÓN:
DENIS SAURÉ VALENZUELA
CHARLES THRAVES CORTÉS-MONROY

Este trabajo ha sido parcialmente financiado por:
Fondecyt, Proyecto # 1181201

SANTIAGO DE CHILE
2021

RESUMEN DE LA TESIS PARA OPTAR
AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL
CON MAGÍSTER EN GESTIÓN DE OPERACIONES
POR: **ANGÉLICA DE JESÚS GALINDO VALDÉS**
FECHA: 2021
PROF. GUÍA: ANDRÉS IGNACIO MUSALEM SAID

ESTIMACIÓN DEL IMPACTO DE MEJORAS A LA CALIDAD DE SERVICIO DE UNA EMPRESA DE DISTRIBUCIÓN LOGÍSTICA EN EL COMPORTAMIENTO DE SUS CLIENTES

El aumento en la satisfacción de los clientes de una empresa significa mayores niveles de retención de los mismos, sin embargo, el impacto de una mejora en la calidad de servicio es diferente para cada rubro, por lo que es importante identificar qué aspectos de la calidad de servicio tienen un mayor efecto en el comportamiento del consumidor. En particular, en esta tesis se utilizan los datos de una empresa de distribución logística, donde las variables relacionadas a la calidad de servicio son; los tiempos de demora o anticipo que tienen los pedidos en relación a la fecha y hora planificadas de entrega, y los tiempos de despacho de los envíos.

Además, debido a que la calidad de servicio es estocástica, pues el cliente no conoce con certeza la calidad de servicio que va a obtener, es necesario considerar la experiencia de consumo anterior, ya que en base a esta el consumidor actualiza sus creencias de consumo posterior. Para modelar dichas creencias son generados dos modelos de pronóstico con estimación Bayesiana y uno más sencillo utilizado como modelo de control. El primero considera la experiencia de todas las rutas utilizadas por el cliente. El segundo solo utiliza la información de la ruta a pronosticar. El último entrega como pronóstico la observación anterior de la ruta del periodo a estimar. De los tres modelos el que obtiene mejores resultados es el primero, seguido por el segundo modelo que presenta un desempeño similar y un menor tiempo de computo.

Se reconocen las limitaciones de los modelos de pronóstico al no incluir otras variables en su cálculo, como por ejemplo, la distancia del envío, el peso del pedido, entre otras, las cuales podrían generar una creencia del consumidor más cercana a la realidad. Motivo de lo anterior, se proponen otros modelos de pronóstico que otorguen mayor o menor peso a las observaciones de la calidad de servicio dependiendo de su antigüedad.

Finalmente, se estima el impacto que tiene la calidad de servicio en el comportamiento de compra de los clientes, para ello, se realizan diferentes modelos de elección discreta respecto a la probabilidad de que el cliente decida utilizar los servicios como función de sus creencias sobre la calidad de servicio y otras variables independientes. Se concluye que una disminución en la puntualidad de los pedidos y un aumento en los tiempos de despacho tiene un efecto negativo en el comportamiento de compra del consumidor, por lo cual es importante invertir en mejorar estos aspectos para asegurar una mayor retención de los clientes de la empresa.

*A mis padres,
Emelita y Francisco.*

Agradecimientos

Agradezco profundamente a mi madre por acompañarme en cada paso de mi vida con amor y preocupación. Siempre es la primera en alegrarse por mis logros y estoy segura que estará incluso mas contenta que yo por esta etapa que se cierra y da paso a una nueva.

A mi padre que siempre se ha preocupado de que salga adelante y de que sea una persona respetuosa y responsable.

A mi hermana gemela que me ha acompañado cada día y ha expandido mi visión de las cosas para llegar a ser la persona que soy hoy, no podría haber terminado esta etapa sin su apoyo y compañía.

A mis hermanos mayores que siempre supieron darme los consejos que necesité. Me han ayudado con todas las dudas que surgieron durante esta etapa, y sé que estarán ahí para mi paso a la siguiente.

A mis amigos del colegio, Angel, Pancha, Mario y Sabi por todas las risas y momentos de alegría que hemos vivido, me acompañaron en mi etapa escolar y siguieron ahí cuando entré a la U, espero seguir junto a ustedes en todas las etapas de mi vida.

A mis amigos de la U, Agus, Ani, Benja, Caro, Diego, Eybie, José, Marco, Lucas, Nati, Nico, Nicole, Pablo, Pipe y Yerko por hacer que esta etapa de mi vida fuera más que solo estudio, junto a ustedes he explorado muchas cosas nuevas y también he aprendido que hay que soñar en grande, espero sigamos compartiendo muchos más momentos juntos.

A mi profesor guía Andrés Musalem por su gran preocupación y constante apoyo para sacar adelante esta tesis. También por todos los consejos y enseñanzas que me ha otorgado, gracias a ellos mi aprendizaje en la Universidad se vio enormemente enriquecido.

Aún me falta mucha gente por agradecer, mis amigos de bachi con los que aún mantengo contacto, las chiquillas de Angel Wings y la gente que conocí gracias al grupo, mis amigos de Kawai Bakery, los chiquillos de la Telewarida y muchas más personas que pasaron por mi vida y me acompañaron en esta etapa, a todos ustedes, muchas gracias.

Tabla de Contenido

1. Introducción	1
1.1. Objetivos	3
1.1.1. Objetivo general	3
1.1.2. Objetivos específicos	3
2. Marco teórico	4
2.1. Modelo Logit	4
2.2. Inferencia Bayesiana	5
2.3. Simulación con Markov Chain Monte Carlo	7
2.4. Métricas de evaluación	8
2.4.1. Error absoluto promedio	8
2.4.2. Raíz del error cuadrático promedio	8
2.4.3. Error porcentual absoluto promedio	8
2.4.4. Verosimilitud marginal	9
2.5. Modelos de Aprendizaje en Marketing	10
2.5.1. Modelo de aprendizaje Bayesiano	10
3. Contexto Empírico	12
3.1. Nuevas variables	13
3.2. Estadísticos descriptivos	15
4. Modelos	18
4.1. Modelos de aprendizaje de calidad de servicio para múltiples productos	18
4.1.1. Modelo de aprendizaje independiente	18
4.1.2. Modelo de aprendizaje jerárquico	19
4.2. Modelo de pronóstico con el método naïve	20
4.3. Modelo elección de compra	20
4.3.1. Modelo de estimación Bayesiana	21
5. Resultados	24
5.1. Modelos de pronóstico	24
5.1.1. Pronóstico de los tiempos de demora o anticipo	24
5.1.2. Pronóstico del tiempo total de envío	26
5.2. Modelo elección de compra con estimación Bayesiana	28
5.2.1. Método Naïve	29
5.2.2. Aprendizaje independiente	30
5.2.3. Aprendizaje jerárquico	32

6. Contrafactuales	35
6.1. Aumento del tiempo de demora y despacho en todas las rutas	35
6.2. Aumento del tiempo de demora y despacho en una de las rutas	37
7. Conclusiones	39
Bibliografía	41
Anexo A. Código en R	43
A.1. Modelos de aprendizaje para el pronóstico del tiempo de demora o anticipo .	43
A.1.1. Modelo de aprendizaje independiente	43
A.1.2. Modelo de aprendizaje jerárquico	43
A.2. Modelo de aprendizaje para el pronóstico del tiempo total de envío	43
A.2.1. Modelo de aprendizaje independiente	43
A.2.2. Modelo de aprendizaje jerárquico	44
A.3. Modelo elección de compra	44
A.3.1. Modelo Homogéneo en stan	44
A.3.2. Modelo Homogéneo híbrido en stan	44
A.3.3. Modelo Heterogéneo por ruta en stan	44
A.3.4. Estimación Bayesiana de elección de compra	44
A.4. Contrafactuales	44
A.4.1. Aumento del tiempo de demora	44
A.4.2. Aumento del tiempo de despacho	45
A.4.3. Aumento del tiempo de demora y despacho	45
Anexo B. Resultados	46
B.1. Modelo elección de compra con estimación Bayesiana	46
B.1.1. Método Naïve	46
B.1.2. Aprendizaje independiente	47
B.1.3. Aprendizaje jerárquico	50
B.2. Pronósticos del tiempo de demora o anticipo y el tiempo total de envío de los diferentes modelos para un cliente	53

Índice de Tablas

3.1.	N° de clientes por rango de envíos totales.	15
4.1.	Variables independientes.	21
5.1.	Métricas de error para pronóstico de tiempos de demora o anticipo	25
5.2.	Métricas de error para pronóstico de tiempo total de envío	27
5.3.	Resultados para modelos de elección de compra con datos del método naïve . .	29
5.4.	Coefficientes de beta, varianza y desviación estándar para modelo Heterogéneo por ruta con datos del método naïve	30
5.5.	Resultados para modelos de elección de compra con datos del aprendizaje independiente	31
5.6.	Coefficientes de beta, varianza y desviación estándar para modelo Heterogéneo por ruta con datos del aprendizaje independiente	32
5.7.	Resultados para modelos de elección de compra con datos del aprendizaje jerárquico	33
5.8.	Coefficientes de beta, varianza y desviación estándar para modelo Heterogéneo por ruta con datos del aprendizaje jerárquico	34
6.1.	Porcentaje de cambio en la demanda total bajo distintos escenarios y modelos de aprendizaje	36
6.2.	Porcentaje de cambio en la demanda de la ruta 1 bajo distintos escenarios y modelos de aprendizaje	38
6.3.	Porcentaje de cambio en la demanda de la ruta 1 para diferentes escenarios, utilizando los coeficientes del modelo de compra jerárquico.	38
B.1.	Resultados para modelos de elección de compra con datos del método Naïve .	47
B.2.	Coefficientes de beta, varianza y desviación estándar para modelo Homogéneo por ruta con datos del método Naïve	47
B.3.	Resultados para modelos de elección de compra con datos del aprendizaje independiente	49
B.4.	Coefficientes de beta, varianza y desviación estándar para modelo Homogéneo por ruta con datos del aprendizaje independiente	49
B.5.	Resultados para modelos de elección de compra con datos del aprendizaje jerárquico	51
B.6.	Coefficientes de beta, varianza y desviación estándar para modelo Homogéneo por ruta con datos del aprendizaje jerárquico	52

Índice de Ilustraciones

2.1.	Ejemplo de simulación con MCMC de 3 parámetros y 40.000 iteraciones. Fuente: Andrieu y Thoms (2008)	7
3.1.	Histograma del tiempo total de envío en días.	14
3.2.	Distribución de envíos en el año 2013.	16
3.3.	Histograma del tiempo de demora de los envíos en hrs.	16
3.4.	Histograma del tiempo de anticipo de los envíos en hrs.	17
3.5.	Segundo histograma del tiempo de demora de los envíos en hrs.	17
5.1.	Gráficos de pronósticos de los tiempos de demora o anticipo para la ruta 1 de un cliente aleatorio.	26
5.2.	Gráficos de pronósticos del tiempo total de envío para la ruta 1 de un cliente aleatorio.	28

Capítulo 1

Introducción

Según Bolton y Lemon (1999) los clientes que se encuentran más satisfechos con el servicio entregado por la empresa tendrán niveles más altos de uso del servicio en el futuro, lo que los convierte en clientes frecuentes, es decir, un aumento en la tasa de retención de clientes, lo cual tiene un efecto significativo en el crecimiento de la rentabilidad de una empresa (Payne y Rickard 1993). Sin embargo, para cada industria o empresa la calidad de servicio se mide de forma diferente, por lo que una mejora de servicio para una empresa puede no funcionar para otra, por lo que es importante determinar en qué aspectos de un servicio se debe invertir para mejorar la calidad, pues el rendimiento al mejorar la calidad es decreciente, es decir, hay un punto en que el gasto en mejorar la calidad deja de ser rentable para la empresa (Hill 1993).

A menudo en la literatura se utilizan encuestas de satisfacción de clientes para medir la calidad de servicio (por ejemplo, Rust et al. 1995; Bolton y Lemon 1999; Boulding et al. 2012). Sin embargo, estas presentan sesgos de selección ya que no todos los clientes contestan la encuesta y dependiendo de como sean estructuradas pueden guiar la respuesta del cliente o descartar opciones de respuesta por no encontrarse en la encuesta. Además al ser subjetivas resulta difícil pasar de una conclusión a una recomendación de toma de decisión que pueda ser de utilidad para la empresa.

Por otra parte, la calidad de servicio es estocástica debido a que el cliente no conoce con certeza la calidad de servicio que va a obtener y es por lo que varios estudios postulan que el consumidor aprende de su experiencia de consumo y mediante esta actualiza su percepción de la calidad del producto o servicio recibido (Roberts y Urban 1988; Erdem y Keane 1996; Ching et al. 2013). Por ejemplo si se considera una persona que compra zapatillas de escalada y que está comenzando a incursionar en el deporte, posiblemente no tenga un gran conocimiento previo de la categoría por lo que a medida que empiece a practicar el deporte y utilice las zapatillas irá adquiriendo experiencia sobre estas y actualizará la calidad que percibe de las mismas, lo cual a futuro puede significar un cambio o no en su decisión de compra para la misma marca o modelo de zapatillas.

Esta tesis busca estudiar el impacto de mejoras a la calidad de servicio de una empresa de distribución logística (freight forwarder), la cual se encarga de los servicios de transporte y logística de productos de otras empresas para que estos lleguen a su destino, siendo estos, tanto nacionales como internacionales. Para ello, la empresa investiga y planea la mejor ruta,

costos, tiempos de tránsito, transporte a utilizar, etc., de acuerdo a la mercancía a transportar, así como también está a cargo de arreglar los pagos y requisitos para la liberación de la carga y asesorar al cliente en todo momento. En adelante, se considera que el servicio o producto de la empresa, corresponde al envío de la mercancía de sus clientes desde el punto de origen al punto de destino. Un cliente puede tener varios puntos de origen y destino, es decir, diferentes rutas, y cada una es considerada un producto diferente.

Para realizar la investigación se utilizan datos del año 2013, en los cuales se tiene; la fecha y hora de inicio de cada pedido, la fecha y hora de llegada prometida al cliente, y la fecha y hora de llegada efectiva. Estos datos permiten obtener diferentes métricas para medir la calidad de servicio; el tiempo en que un pedido ha llegado tarde según lo planificado (tiempo de demora), el tiempo en que un pedido ha llegado antes de lo planificado (tiempo de anticipo), y el tiempo total que ha tardado el pedido desde que este tuvo inicio, hasta que efectivamente llegó (tiempo de envío).

Se realizan tres modelos; los primeros dos corresponden a modelos de aprendizaje del cliente sobre la calidad de servicio que experimenta cada vez que un envío finaliza. Los resultados de estos primeros modelos son utilizados para estimar el tercer modelo, que corresponde a un modelo de elección discreta respecto de la probabilidad de que el cliente decida utilizar los servicios como función de sus creencias acerca de la puntualidad de la empresa y el tiempo de despacho de los pedidos.

Ambos modelos de aprendizaje permiten tener en cuenta la incertidumbre de los clientes y las creencias que obtienen a medida que utilizan el servicio, lo cual ha sido estudiado anteriormente en la literatura, por ejemplo en el trabajo de Chan y Hamilton (2006), donde se incorpora la actualización del conocimiento con aprendizaje Bayesiano o en el trabajo de Narayanan y Manchada (2009) quienes utilizan actualizadores Bayesianos para actualizar las cualidades de fármacos en base a creencias previas, las cuales son actualizadas al adquirir nueva información gracias a la regla de Bayes que permite obtener creencias posteriores. Sin embargo, a diferencia de estos estudios, para esta tesis se utiliza información no solo de experiencias con el servicio para el cual se quiere inferir su calidad, sino también de otros servicios del mismo proveedor. Además se contrastan y comparan modelos de aprendizaje basados en distintos supuestos de como el cliente utiliza la información de experiencias anteriores para hacer pronósticos sobre la calidad que experimentaría en el futuro. Por último, cabe destacar que en la literatura estudiada el contexto es para servicios B2C, mientras que para el caso de este estudio el servicio es B2B, por lo que la sensibilidad de los clientes a cambiar de servicio es distinta.

De los principales resultados obtenidos se puede observar que el modelo de pronóstico que considera la información de la calidad de servicio de las diferentes rutas que utiliza el cliente para actualizar sus creencias para el periodo y ruta a estimar tienen un mejor desempeño que un modelo que solo tiene en consideración los datos de la ruta que esta siendo pronosticada. Además al analizar distintos escenarios de disminución en la calidad de servicio, como por ejemplo el aumentar en promedio 2 horas los tiempos de demora y despacho de los envíos para todos los consumidores, se tiene que la demanda total disminuye un 21 %.

La tesis comienza con la teoría detrás de los modelos realizados para luego explicar el

contexto en el cual se desenvuelven los datos utilizados, dando paso así a los modelos desarrollados y los resultados correspondientes. El trabajo finaliza con el análisis de escenarios de disminución en la calidad de servicio, se concluye que las creencias pronosticadas para los clientes sobre los tiempos de despacho y demora de los envíos, tienen un efecto negativo en su probabilidad de compra, y se propone como trabajo a futuro la generación de otros modelos de pronóstico que otorguen mayor o menor peso a las observaciones dependiendo de su antigüedad.

1.1. Objetivos

1.1.1. Objetivo general

- Estimar la sensibilidad de los clientes a la calidad de servicio, la cual es determinada por la puntualidad de las entregas realizadas y el tiempo destinado a realizar el pedido.

1.1.2. Objetivos específicos

- Formular y estimar un modelo de pronóstico Bayesiano para estimar la creencia del cliente sobre el tiempo de demora o anticipo en su siguiente pedido.
- Formular y estimar un modelo de pronóstico Bayesiano para estimar la creencia del cliente sobre el tiempo de envío en su siguiente pedido.
- Formular un modelo de elección discreta para la probabilidad de que el cliente decida utilizar los servicios de la empresa en el siguiente periodo.

Capítulo 2

Marco teórico

Para mejorar la calidad de un servicio es importante determinar que aspectos del mismo afectan en mayor grado la percepción del cliente sobre esta, para lo cual se utiliza un modelo logit que permite explicar eventos de tipo binario en base a variables independientes como se detalla más adelante en este capítulo, para luego dar paso a la inferencia Bayesiana, con la cual es posible estimar algunas de las variables de interés para el modelo anterior, como lo es el pronóstico de cada cliente sobre la calidad a futuro del servicio, dada la información de la experiencia que tienen acumulada. Sin embargo, los modelos matemáticos resultantes de la inferencia Bayesiana tienen una alta complejidad, y es por lo cual se utiliza la simulación de Markov Chain Montecarlo explicada en el capítulo, además de diferentes métricas de evaluación para comparar el desempeño de estos. Por último, se habla sobre los modelos de aprendizaje en marketing para comprender la aplicación de la teoría en la estimación del impacto de mejoras a la calidad de servicio en el comportamiento de los clientes.

2.1. Modelo Logit

El modelo binario logit se utiliza para explicar fenómenos en donde la variable a estudiar es binaria, es decir solo puede tomar dos valores, como por ejemplo comprar o no un auto, enviar una carga de productos por medio de una empresa especializada, entre otras. Todas estas decisiones dependen de otras variables que permiten explicar la toma de una u otra decisión respecto del evento estudiado, sea en el caso de enviar una carga de productos; el peso del envío, la distancia a la cual debe enviarse, si es que ha utilizado la empresa con anterioridad; el tiempo de demora o anticipo de envíos anteriores, el tiempo total de envíos anteriores desde que son solicitados hasta que llegan al lugar de destino, etc. Una de las ventajas de utilizar este modelo es que los resultados son altamente interpretables en comparación a otros modelos (por ejemplo, redes neuronales, K-nearest neighbor o support vector machine), por lo cual se puede estimar directamente el peso de cada una de las variables, el que determina la importancia que tiene en la decisión final. En adelante la información contenida en esta sección estará basada en Hosmer y Lemeshow (2000).

Este tipo de modelo es un caso particular de regresiones en donde para este caso se busca explicar la probabilidad de que ocurra un evento o fenómeno binario, representado por y_i donde la variable toma el valor 0 si no ocurre el evento y 1 si es que lo hace, y es representada por la ecuación a continuación:

$$p(y_i = 1) = F(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_k x_{k,i}) = F(B^T x). \quad (2.1)$$

Donde las variables incluidas en el vector x explican el fenómeno a estudiar y β determina el peso que tiene cada variable en la determinación del evento. El producto de ambas se evalúa dentro de la función $F()$, la cual representa la función de distribución de la probabilidad acumulada de la variable dependiente y , donde si esta probabilidad supera el umbral impuesto que va entre 0 y 1, como por ejemplo 0.5, significa que el evento ocurre y de ser menor a este, no lo hace. Por lo anterior la función elegida es la logística como se ve en la ecuación (2.2), ya que entrega valores entre 0 y 1 como es requerido para el modelo.

$$p(y_i = 1) = \frac{\exp(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_k x_{k,i})}{1 + \exp(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_k x_{k,i})}. \quad (2.2)$$

Como se puede observar el modelo no es lineal, por lo que el cambio en el coeficiente β_i no representa un cambio proporcional en la probabilidad de tomar la decisión cuando la variable correspondiente x_i cambia en una unidad, por lo que es necesario calcular el efecto marginal con la derivada de la función respecto a x :

$$\frac{\partial p(y = 1)}{\partial x} = \frac{\exp(\beta^T x)}{1 + \exp(\beta^T x)} \left[1 - \frac{\exp(\beta^T x)}{1 + \exp(\beta^T x)} \right] \beta. \quad (2.3)$$

Además es posible notar que el efecto marginal no es fijo ya que depende de los valores que tomen las variables independientes x , así por ejemplo en el caso de realizar un envío de carga de productos por medio de una empresa especializada, el aumentar el peso del envío cambia la probabilidad de que este sea realizado a través de una empresa especializada según fuese la magnitud del peso antes del cambio, y a su vez según sean los valores del resto de las variables independientes. Es por esto que para tener un efecto marginal representativo se evalúa la ecuación (2.3) en los valores promedio de las variables explicativas, representados por \bar{x} , como se muestra en la ecuación que sigue:

$$\left. \frac{\partial p(y = 1)}{\partial x} \right|_{x=\bar{x}} = \frac{\exp(\beta^T \bar{x})}{1 + \exp(\beta^T \bar{x})} \left[1 - \frac{\exp(\beta^T \bar{x})}{1 + \exp(\beta^T \bar{x})} \right] \beta. \quad (2.4)$$

2.2. Inferencia Bayesiana

La inferencia Bayesiana permite predecir nuevas observaciones de un conjunto de datos, ya sea que estos presenten una distribución conocida o no. Esto mediante el proceso de ajuste de un modelo de probabilidad a los datos ingresados, el cual entrega como resultado una distribución de probabilidad para los parámetros del modelo, con la cual se realiza la predicción. La sección a continuación se basa en los conocimientos entregados por Gelman et al. (2020).

Componentes

Para entender la inferencia Bayesiana primero es necesario entender su notación y los componentes que la conforman; los parámetros, los datos y la predicción.

- **Parámetros:** son denotados por letras griegas como θ y representan un vector de cantidades no observables o parámetros de interés de la población (para el caso de este estudio es la probabilidad de realizar un envío).
- **Datos:** son todos los datos observables del modelo, donde se denota y al vector que se desea predecir (como por ejemplo la realización de un pedido por parte de un cliente), y x a los vectores de variables explicatorias (por ejemplo, el peso de un paquete o la distancia recorrida).
- **Predicción:** son los datos desconocidos, pero potencialmente observables y se denotan por \hat{y}

Regla de Bayes

Una vez que se tiene conocimiento de los componentes es necesario conocer como funciona la probabilidad conjunta, la cual sigue la regla de Bayes a continuación:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)}. \quad (2.5)$$

Omitiendo el factor $p(y)$ en la ecuación (2.5) ya que este no depende de θ se obtiene la ecuación equivalente (2.6) conocida como la densidad posterior no normalizada, la cual es la base de la inferencia Bayesiana, ya que interesa saber el valor de $p(\theta|y)$, donde \propto denota proporcionalidad.

$$p(\theta|y) \propto p(\theta)p(y|\theta). \quad (2.6)$$

La ecuación (2.6) se divide en 3 partes:

- $p(\theta)$ es denominado **prior** y corresponde a la creencia que se tienen sobre la distribución del evento que gatilla el suceso estudiado sin haber observado ningún dato, por ejemplo, la distribución de probabilidad de los clientes respecto a realizar un envío.
- $p(y|\theta)$ es denominado **verosimilitud** y es como se cree que los datos están distribuidos dada la distribución de θ . Si fuera posible conocer la probabilidad de un cliente de realizar un pedido, esto entregaría el número de veces que el cliente realiza un envío.
- $p(\theta|y)$ es denominado **posterior** y es la creencia posterior de los parámetros luego de observar la evidencia (y). Para el caso de esta tesis es la probabilidad de que un cliente realice un envío dado que ha realizado un número determinado de ellos.

Luego para definir correctamente el modelo es necesario representar la función de distribución de **verosimilitud** $p(y|\theta)$ y **prior** $p(\theta)$ ya que el producto de ambas da como resultado la distribución de la creencia posterior o **posterior** $p(\theta|y)$, como se muestra en la ecuación (2.7), donde $F(y|\theta)$ es la función de distribución para la verosimilitud y $G(\theta)$ es la función de distribución a priori.

$$p(\theta|y) \propto F(y|\theta) \cdot G(\theta). \quad (2.7)$$

Además en el caso de que la distribución a priori sea una familia conjugada para la distribución de la verosimilitud, se cumple que la distribución posterior tendrá la misma forma paramétrica del prior, lo cual es matemáticamente conveniente ya que la distribución posterior tiene una forma conocida.

2.3. Simulación con Markov Chain Monte Carlo

Si bien la inferencia Bayesiana permite llegar a formulas para problemas estadísticos complejos de resolver, también el cálculo de estos puede tomar la forma de múltiples integrales o sumatorias difíciles o imposibles de resolver, y ahí es cuando entra en juego Markov Chain Montecarlo que permite llegar a una solución simulando varios ejemplos de una probabilidad de distribución dada. Esta información y la de la sección completa esta basada en los conocimientos entregados por Gelman et al. (2020).

El método Markov Chain Monte Carlo (en adelante MCMC) consiste en generar valores para θ a partir de una distribución dada e ir corrigiendo este valor para obtener una mejor aproximación de la distribución posterior objetivo $p(\theta|y)$. Esto se logra al crear una o más secuencias de valores para θ basada en el último valor generado, de esta forma se obtiene una cadena de Markov ¹ por cada secuencia simulada.

Los valores de θ se generan de forma aleatoria a partir de la distribución entregada, donde en un principio el primer valor de θ puede ser generado de forma aleatoria, para luego “saltar” o moverse a otro valor según el criterio definido por el algoritmo utilizado en la simulación, donde en particular para esta tesis se utilizará el algoritmo de NUTS (Hoffman y Gelman 2011).

La clave de la simulación de cadenas de Markov es crear un proceso de Markov cuya distribución estacionaria sea la de $p(\theta|y)$ y que la simulación sea lo suficientemente larga (el número de iteraciones de θ es el largo de la secuencia simulada) para que esta converja a la distribución estacionaria buscada como se muestra en la Figura 2.1, donde a la izquierda se puede ver como la simulación no converge y luego de 30.000 iteraciones si lo hace.

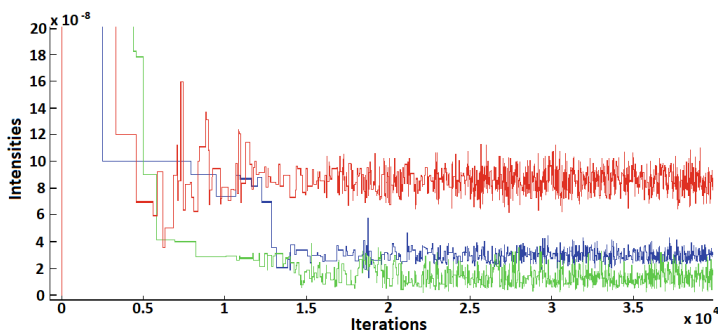


Figura 2.1: Ejemplo de simulación con MCMC de 3 parámetros y 40.000 iteraciones. Fuente: Andrieu y Thoms (2008)

¹ Una cadena de Markov es una secuencia aleatoria de variables $\theta^1, \theta^2, \dots$, para la cual para cualquier t , la distribución de θ^t dado todos los θ previos depende solamente del valor más reciente, es decir, θ^{t-1}

2.4. Métricas de evaluación

Una vez que se tienen los resultados de un modelo, es importante calcular métricas de evaluación, para de esta forma comparar su desempeño con otro u otros modelos que se pudiesen realizar para la misma métrica a estimar, y así determinar cual de los modelos tiene un mejor ajuste o es más afín a los propósitos del estudio dependiendo de las métricas utilizadas para su evaluación. Algunas métricas comúnmente utilizadas y que serán descritas a continuación son; el error absoluto promedio (MAE por sus siglas en inglés), la raíz del error cuadrático promedio (RSME) y el error porcentual absoluto promedio (MAPE) que serán utilizadas para comparar el desempeño de los pronósticos sobre la calidad de servicio, y por otra parte la media armónica que será utilizada para comparar el desempeño de los modelos de elección de compra del cliente.

2.4.1. Error absoluto promedio

El error absoluto promedio (MAE) mide la magnitud promedio de los errores en un set de predictores, sin considerar su dirección. Es el promedio de la diferencia absoluta entre el valor real y el pronóstico realizado, donde todas las diferencias individuales tienen el mismo peso como se muestra en la ecuación 2.8, donde y_j es el valor real, \hat{y}_j el pronóstico y n el número de observaciones predichas.

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|. \quad (2.8)$$

Esta métrica solo puede ser aplicada sobre modelos de pronóstico del mismo set de datos, ya que esta expresada en la escala de los datos predichos, además puede tomar valores de 0 a infinito, donde un valor más cercano a 0 significa un mejor ajuste del modelo.

2.4.2. Raíz del error cuadrático promedio

La raíz del error cuadrático promedio (RMSE) mide la magnitud promedio de los errores de un set de predictores. Es la raíz del promedio de la diferencia al cuadrado entre el valor real y el pronóstico realizado, como se muestra en la ecuación 2.9, donde y_j es el valor real, \hat{y}_j el pronóstico y n el número de observaciones predichas.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}. \quad (2.9)$$

Al igual que el MAE esta métrica solo puede ser aplicada sobre modelos de pronóstico del mismo set de datos, ya que esta expresada en la escala de los datos predichos, además puede tomar valores de 0 a infinito, donde un valor más cercano a 0 significa un mejor ajuste del modelo.

2.4.3. Error porcentual absoluto promedio

El error porcentual absoluto promedio (MAPE) mide la magnitud promedio de los errores de un set de predictores, pero en términos porcentuales, por lo que no es necesario saber la escala del set de datos para interpretar esta métrica. Es el promedio de la diferencia absoluta

entre el valor real y el pronóstico realizado, dividida por el valor real, como se muestra en la ecuación 2.10, donde y_j es el valor real, \hat{y}_j el pronóstico y n el número de observaciones predichas.

$$\text{MAPE} = \frac{100}{n} \sum_{j=1}^n \left| \frac{y_j - \hat{y}_j}{y_j} \right|. \quad (2.10)$$

Esta métrica solo puede ser aplicada a pronósticos donde el valor real no contenga un 0, pues se indefine la ecuación de ser el caso, además toma valores de 0% a 100%, donde un valor más cercano a 0 significa un mejor ajuste del modelo.

2.4.4. Verosimilitud marginal

La verosimilitud marginal es denominada la probabilidad predictiva de los datos condicional en un modelo, es decir, la probabilidad de ver los datos que realmente se observaron bajo los supuestos del modelo, calculada antes de que los datos estuvieran disponibles (Kass y Raftery 1995). Una forma de estimar esta métrica es mediante el método de la media armónica (HM por sus siglas en ingles), de los valores de la verosimilitud (L) obtenidos en cada una de las j iteraciones de la estimación Bayesiana por simulación con cadenas de Markov.

$$\text{HM} = \frac{1}{\frac{1}{n} \sum_{j=1}^n \frac{1}{L_j}}, \quad (2.11)$$

donde n es el número de iteraciones de la cadena de Markov. Para esta tesis se utiliza el logaritmo de este valor para facilitar la interpretación de esta métrica. Además para evitar problemas numéricos de underflow y overflow, por lo cual se realiza la siguiente transformación:

$$\begin{aligned} \text{HM} &= \frac{1}{\frac{1}{n} \sum_{j=1}^n \frac{1}{L_j}} \\ &= \frac{1}{\frac{1}{n} \sum_{j=1}^n \frac{1}{\exp(\log(L_j))}} \\ &= \frac{1}{\frac{1}{n} \sum_{j=1}^n \exp(-\log(L_j))} \\ &= \frac{1}{\frac{1}{n} \sum_{j=1}^n \exp(-\log(L_j) + \min(L) - \min(L))} \\ &= \frac{1}{\frac{1}{n} \sum_{j=1}^n \exp(-\min(L)) \exp(-(\log(L_j) - \min(L)))} \\ \log(\text{HM}) &= \log(1) - \log\left(\frac{1}{n} \sum_{j=1}^n \exp(-\min(L)) \exp(-(\log(L_j) - \min(L)))\right) \\ \log(\text{HM}) &= \min(L) - \log\left(\frac{1}{n} \sum_{j=1}^n \exp(\min(L) - \log(L_j))\right). \end{aligned} \quad (2.12)$$

Donde L_j corresponde al vector de la verosimilitud marginal para cada una de las j

iteraciones, es decir, la probabilidad del dato observado dado el modelo.

2.5. Modelos de Aprendizaje en Marketing

En la mayoría de los modelos de economía o marketing se asume que el consumidor elige lo que le otorga mayor utilidad, sin embargo, esto implica un gran supuesto de por medio, y es que el consumidor conoce a cabalidad los atributos del producto, lo cual por lo general no sucede, por ejemplo, si es lanzado un producto nuevo al mercado el consumidor no tendrá conocimiento de sus atributos hasta utilizar el producto. Dado que existe este elemento de incertidumbre se calcula una utilidad promedio esperada, la cual es actualizada con cada nueva experiencia del cliente y de esta forma mejorar el promedio (Erdem y Keane 1996).

2.5.1. Modelo de aprendizaje Bayesiano

Este modelo utiliza la experiencia del consumidor para actualizar la creencia que este tiene sobre el producto y fue desarrollado por Erdem y Keane (1996). La clave del modelo es la incerteza acerca de los atributos del producto, por lo cual antes de recibir cualquier información por experiencia de uso, la calidad del producto Q_j para la marca j es modelada como una distribución normal con media Q_{j1} y varianza σ_{j1}^2 como se muestra en (2.13), donde los valores de Q_{j1} y σ_{j1}^2 pueden ser influenciados por muchos factores, como la reputación de la empresa, recomendaciones de conocidos, etc.

$$Q_j \sim N(Q_{j1}, \sigma_{j1}^2), \quad j = 1, \dots, J. \quad (2.13)$$

Además la experiencia por uso puede variar dependiendo del consumidor y el periodo en que se haya realizado, por lo que dicha experiencia no revela completamente la calidad del producto, es decir, la calidad percibida por el consumidor i para la marca j en la compra en el periodo t , denotada por Q_{ijt}^E , estará dada por la verdadera calidad Q_j y la variabilidad de experiencia denotada por ϵ_{ijt} , quedando así la siguiente ecuación:

$$Q_{ijt}^E = Q_j + \epsilon_{ijt}, \quad \text{donde} \quad \epsilon_{ijt} \sim N(0, \sigma_\epsilon^2) \quad (2.14)$$

Luego para representar la calidad percibida en el periodo $t = 2$ es necesario realizar la actualización en la creencia de la calidad que tiene el consumidor basado en la experiencia de uso del producto en el periodo $t = 1$, la cual está representada por las ecuaciones (2.15) y (2.16), en donde es posible ver que mientras más acertada sea la señal, es decir, mientras más pequeño sea σ_ϵ^2 , mejor será la actualización obtenida.

$$Q_{ij2}^E = \frac{\sigma_{j1}^2}{\sigma_{j1}^2 + \sigma_\epsilon^2} Q_{ij1}^E + \frac{\sigma_\epsilon^2}{\sigma_{j1}^2 + \sigma_\epsilon^2} Q_{j1}. \quad (2.15)$$

$$\sigma_{ij2}^2 = \frac{1}{\left(\frac{1}{\sigma_{j1}^2}\right) + \left(\frac{1}{\sigma_\epsilon^2}\right)}. \quad (2.16)$$

Definiendo d_{ijt} como la marca j que es consumida en el periodo t por la persona i es posible extender las ecuaciones (2.15) y (2.16) para N número de periodos de la siguiente forma:

$$Q_{ijt}^E = \frac{\sigma_{j1}^2}{N_{ij}(t)\sigma_{j1}^2 + \sigma_\epsilon^2} \sum_{s=1}^t Q_{ijs}^E d_{ijs} + \frac{\sigma_\epsilon^2}{N_{ij}(t)\sigma_{j1}^2 + \sigma_\epsilon^2} Q_{j1}. \quad (2.17)$$

$$\sigma_{ij2}^2 = \frac{1}{\left(\frac{1}{\sigma_{j1}^2}\right) + N_{ij}(t)\left(\frac{1}{\sigma_\epsilon^2}\right)}. \quad (2.18)$$

Por último se asume una forma funcional para la utilidad en la ecuación (2.19), donde P_{ijt} es el precio, w_P es la utilidad marginal del ingreso y e_{ijt} es un error idiosincrático.

$$U_i(Q_{jt}^E, P_{jt}) = f(Q_{ijt}^E) - w_P P_{ijt} + e_{ijt}. \quad (2.19)$$

Si se asume que $f(Q_{ijt}^E)$ toma la forma de aversión absoluta al riesgo constante², y se denota I_{it} como el conjunto de información del consumidor i en el periodo t , entonces la utilidad esperada es dada por:

$$E[U_i(Q_{jt}^E, P_{jt})|I_{it}] = -exp(-r(Q_{ijt} - \frac{r}{2}(\sigma_{ijt}^2 + \sigma_\epsilon^2))) - w_P P_{ijt} + e_{ijt}. \quad (2.20)$$

De la ecuación (2.20) se puede ver que si σ_{ijt}^2 es alta, es decir, la incerteza respecto a la calidad del producto es alta, entonces la utilidad esperada del consumidor disminuye, pero a su vez el utilizar una marca que no es familiar aumenta los conocimientos de la misma, lo cual genera nueva información para a futuro disminuir el valor de σ_{ijt}^2 .

² La aversión absoluta al riesgo constante o CARA por sus siglas en ingles (Constant absolute risk-aversion) es una forma funcional de la utilidad, siendo esta una exponencial negativa, donde $r > 0$ significa una aversión al riesgo respecto a la variación en la calidad del producto, para este caso [5].

Capítulo 3

Contexto Empírico

La tesis es realizada sobre los datos de una empresa de distribución logística (freight forwarder), la cual se especializa en proporcionar una variedad de funciones para facilitar el movimiento tanto de envíos internacionales como nacionales. Las funciones que proporciona incluyen: servicios de entrega y distribución; manejo de arreglos de pago; asistencia en el papeleo necesario para las entregas; servicios de información; seguro; servicios de aduana; entre otros.

La empresa tiene años de experiencia en este rubro, sin embargo, existe un límite en cuanto al manejo que puede realizar de los tiempos de llegada de la carga, debido a que solo se encarga de la contratación del transporte mas no de transportar la carga, por lo cual es de relevancia para la empresa saber cómo afecta la puntualidad de las entregas a la calidad percibida por el cliente y si esto afecta en tal medida que el cliente no siga realizando envíos a través de la empresa.

Los datos facilitados para el estudio corresponden a datos de envío de un total de 725 clientes para el año 2013 los cuales suman un total de 26.037 observaciones, en las cuales cada una presenta las siguientes variables:

- Start time: Fecha y hora de inicio
- Final time: Fecha y hora de termino
- Carrier: empresa por la cual se realiza el envío.
- Packages: n° de paquetes enviados.
- Weight Kg: peso en kg del envío.
- Volume CBM: volumen en metros cúbicos.
- From country: país de origen.
- To country: país de destino.
- Delay: tiempo de atraso en horas.
- Early: tiempo de adelanto en horas.

- Difference: diferencia entre early y delay.
- Shipments: número de envíos realizados.
- Distance: Distancia del envío en metros.

Al analizar cada una de las variables es posible observar que existen 1.945 observaciones que no registran el peso del envío, por lo cual se opta por ingresar el peso promedio de los envíos del cliente, sin embargo dos clientes no tienen ningún dato sobre el peso de sus envíos por lo que son eliminados de la base (en total suman 72 observaciones).

3.1. Nuevas variables

Se crean variables adicionales a las ya existentes en la base de datos y a partir de estas se realiza una limpieza de la misma.

Period

Para analizar el comportamiento de los clientes de forma secuencial se divide el año en periodos de mitad de semana para facilitar la interpretación posterior de los datos sin disminuir significativamente la granularidad de los mismos. De esta forma de lunes a miércoles se tiene un periodo y de jueves a domingo otro, por lo que el periodo 1 corresponde al 03 de enero del año 2013 hasta el 06 de enero del año 2013, luego el periodo 2 corresponde al 07 de enero hasta el 09 de enero, y así sucesivamente se van creando el resto de los periodos del año.

Domestic

Como una forma de caracterizar los envíos como nacionales o internacionales se crea esta variable en base a la distancia del envío, donde toma el valor 1 si la ruta es considerada doméstica, es decir, presenta una distancia menor o igual a 3.682 metros (esto ya que es la máxima distancia que presenta una ruta con el mismo país de origen y destino en la base de datos) y un 0 si no.

Time

Como otra forma de medir la calidad de servicio se crea la variable para el tiempo total que demora el envío desde que es solicitado hasta que es recibido por el cliente, en base a las variables “start time” y “final time”, y para la cual se realiza el siguiente histograma:

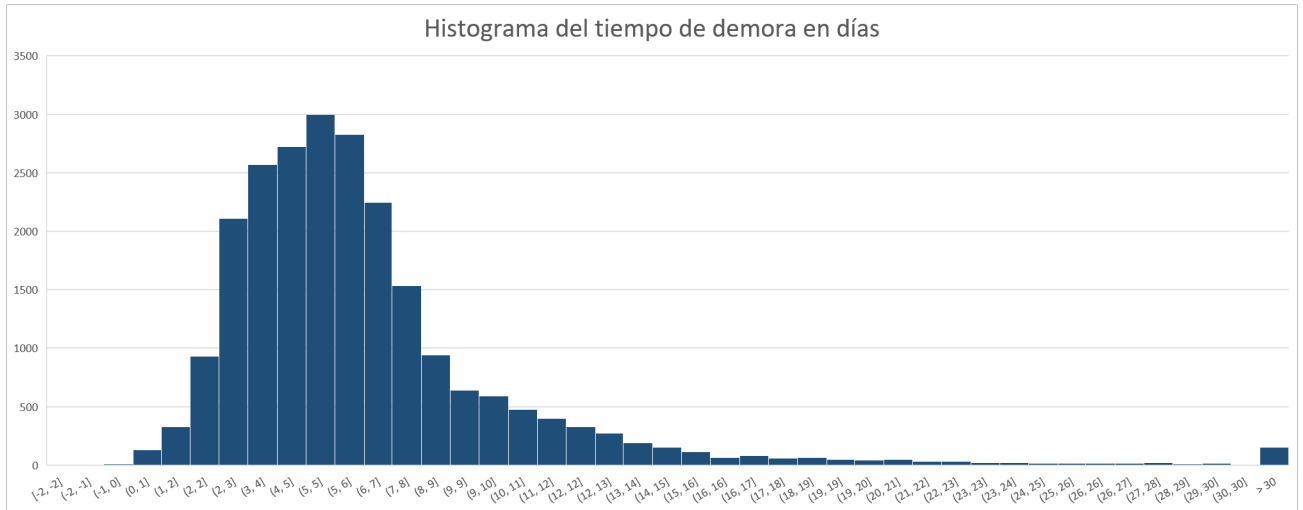


Figura 3.1: Histograma del tiempo total de envío en días.

De la Figura 3.1 se observa que existen envíos con tiempos de demora negativos y que la mayoría de los envíos se acumulan en un tiempo de demora menor a 20 días. Por otra parte, se crea una variable auxiliar de la velocidad del envío calculada como la distancia dividida por el tiempo de envío en horas. De esta forma aquellas observaciones que presentan velocidades mayores a 600 km/hr, o que presentan un tiempo de envío negativo o mayor a 20 días son marcadas como anómalas, y por consiguiente eliminadas de la base. Además aquellos clientes que presentan 3 o más observaciones anómalas son eliminados de la base, por lo que en total se eliminan 1.658 observaciones, quedando así un total de 677 clientes.

Route

Para diferenciar los envíos de un cliente en diferentes productos se crea la variable “route” en base al país de origen y destino que sigue el envío de cada cliente de la siguiente forma:

1. Enumerar las rutas del cliente donde 1 es la ruta con la mayor cantidad de envíos y la n-ésima es la que menos tiene.
2. Agrupar las rutas que tienen al menos 2 envíos desde el periodo 1 al 25 y al menos 4 envíos desde el periodo 26 en adelante.
3. Las rutas que no cumplen el criterio anterior son agrupadas en dos rutas, donde un grupo serán las rutas consideradas domésticas y el otro grupo las rutas internacionales.

Cabe destacar que luego de realizar la codificación de las rutas se vuelve a implementar el criterio del segundo punto y aquellas rutas que no lo cumplen son eliminadas de la base de datos, lo cual se descartan 2.678 observaciones.

Order Frequency

Para cada cliente se calcula el promedio de la secuencia de envíos que tuvo en los primeros 25 periodos para cada una de las rutas que utiliza, por ejemplo, si algún cliente realizó envíos

para la ruta 1 en los periodos 2,9,11,15 y 19, entonces el valor de la variable “order frequency” será el resultado de 5 dividido en 25, es decir 0.2. Esto pretende capturar la frecuencia con que cada cliente realiza pedidos, donde son utilizados solo los primeros 25 periodos ya que estos no se incluyen en el modelo de elección de compra como se explicará más adelante, de esta forma se tiene una variable que refleja la heterogeneidad de los clientes en cuanto a la frecuencia de ordenar pedidos.

Demand

Se calcula la demanda del periodo como la suma de los envíos realizados en ese periodo por el resto de los clientes para tener una variable de control estacionario en el modelo.

3.2. Estadísticos descriptivos

Se realiza un análisis descriptivo de las variables más relevantes, como lo son, los tiempos de demora y anticipo, el tiempo total de envío, cantidad de envíos por cliente, etc.

Envíos

En total se utiliza el registro de 21.629 envíos, de los cuales en promedio cada cliente realiza 32 envíos a lo largo de un año, donde la máxima cantidad de envíos es de 77 y el mínimo es de 11, lo cual supone una gran variabilidad en la cantidad de envíos que realiza cada cliente, pero al observar la Tabla 3.1 es fácil ver que solo un 2% de los clientes se encuentra en el rango más alto del número de envíos, mientras que la mayoría, un 81% de los clientes, realiza de 21 a 50 envíos.

Tabla 3.1: N° de clientes por rango de envíos totales.

N° de envíos por cliente	N° de clientes	Porcentaje del total
[11-20]	80	11,8 %
[21-30]	272	40,2 %
[31-40]	191	28,2 %
[41-50]	88	13,0 %
[51-60]	31	4,6 %
[61-70]	12	1,8 %
[71-80]	3	0,4 %
TOTAL	677	100,00 %

Por otra parte, al analizar como se distribuyen los envíos a lo largo del año, en la Figura 3.2 se puede observar que si bien en promedio se realizan cerca de 2.000 pedidos por mes, para el caso del mes de diciembre este número es cercano a la mitad de los envíos promedio que hay para el resto de los meses, por lo cual puede significar un efecto estacionario o puede

deberse a una tendencia a la baja en la cantidad de envíos a medida que avanza el año.

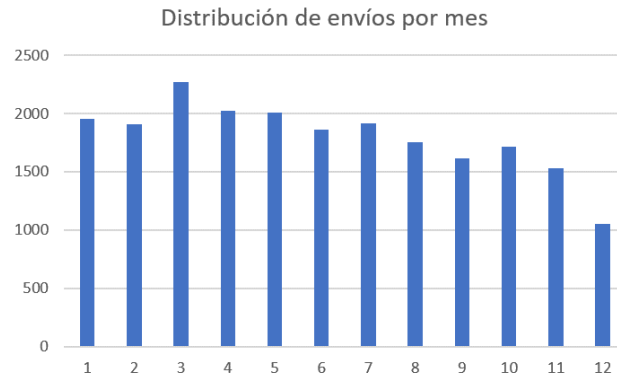


Figura 3.2: Distribución de envíos en el año 2013.

Delay

En la Figura 3.3 se puede observar que el 90 % de las observaciones que presentan algún tiempo de demora se encuentran en un rango de 1 a 6 horas, con un promedio de 2.9 horas. Además del total de las observaciones, solo un 17.5 % tiene un tiempo de demora mayor a una hora. Por último, se observa que del total de envíos, al menos 7 tienen una demora mayor a un día, siendo la máxima de 37 horas.

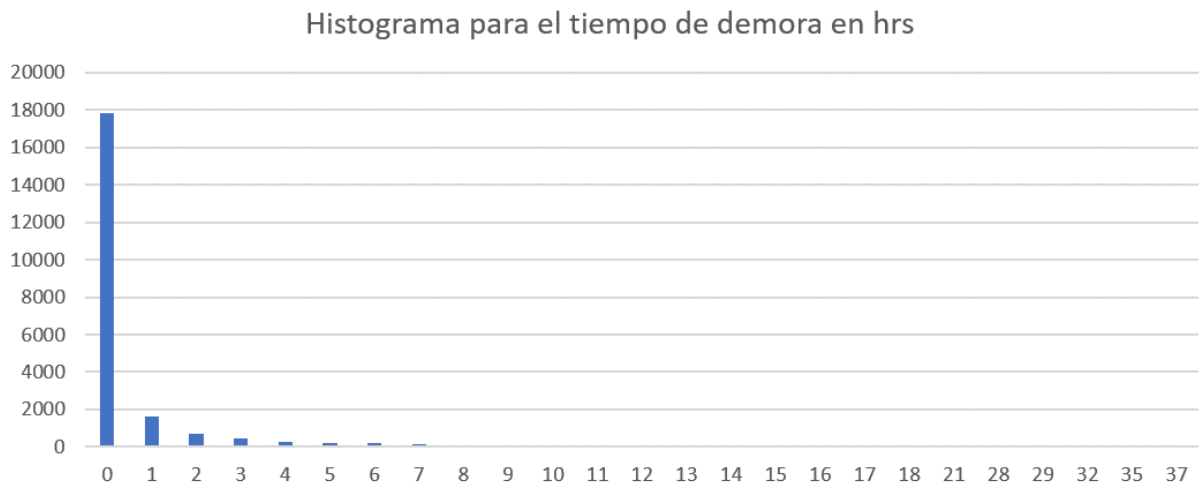


Figura 3.3: Histograma del tiempo de demora de los envíos en hrs.

Early

En la Figura 3.4 se puede observar que el 92 % de las observaciones que presentan algún tiempo de anticipo se encuentran en un rango de 1 a 5 horas, con un promedio de 2.6 horas. Además del total de las observaciones, un 54.9 % tiene un tiempo de anticipo mayor a una hora, esto es más de la mitad de las observaciones, por lo que podría significar una sobreestimación en la planificación del tiempo de llegada, lo cual explicaría a su vez el bajo porcentaje de los tiempos de demora. Por último, se observa que del total de envíos, al menos 4 tienen anticipo mayor a un día, siendo la máxima de 31 horas.



Figura 3.4: Histograma del tiempo de anticipo de los envíos en hrs.

Time

Luego de la eliminación de los datos anómalos se realiza nuevamente el histograma para el tiempo total de envío como se muestra en la Figura 3.5, en la cual se puede observar que el 92 % de las observaciones tiene un rango entre 1 a 10 días para el tiempo total que tarda en ser realizado el envío, donde el promedio se encuentre entre 5 a 6 días.

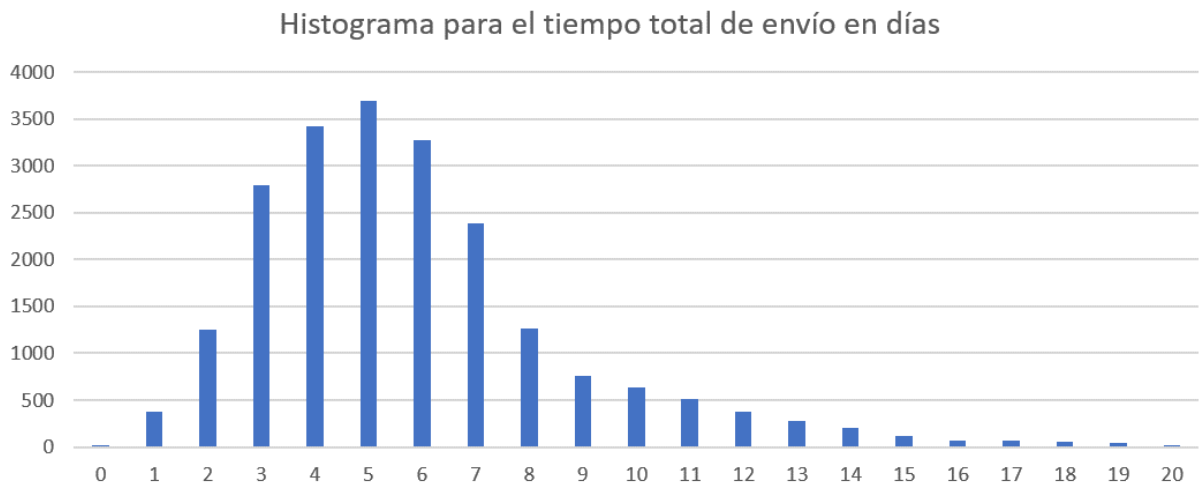


Figura 3.5: Segundo histograma del tiempo de demora de los envíos en hrs.

Capítulo 4

Modelos

En este capítulo se detallan los modelos a utilizar en el desarrollo de la tesis, donde en primer lugar se describen los modelos de aprendizaje del cliente sobre la calidad del servicio que experimenta con cada envío para múltiples productos, luego un modelo de pronóstico simple con el método naïve que permita comparar los resultados de los modelos más complejos, para luego dar paso a un modelo de elección discreta respecto de la probabilidad de que el cliente decida utilizar los servicios, como función de sus creencias sobre la calidad del servicios y otras variables utilizadas en el modelo.

4.1. Modelos de aprendizaje de calidad de servicio para múltiples productos

Cada cliente de la empresa utiliza diferentes rutas de envío para hacer llegar sus encomiendas, y dependiendo de una u otra, los tiempos de llegada del paquete pueden encontrarse adelantados o con retraso respecto de la fecha y hora estimada para la llegada, así como también pueden variar el tiempo necesario para enviar cada pedido, lo cual afecta la calidad de servicio percibida por los clientes. Por esto se busca estimar dos modelos de pronóstico; uno sobre el tiempo de demora o anticipo que un cliente piensa que tendrá su pedido en base a los tiempos de demora o anticipo observados con anterioridad y para cada ruta, y otro análogo sobre el tiempo total de envío. Para ello se definen 2 modelos de pronóstico, cada uno con menor o mayor grado de complejidad, y para cada uno se realiza una simulación con Markov Chain Montecarlo (MCMC) mediante el software de programación R, para estimar los valores de cada uno de los parámetros y de esta forma obtener el valor del pronostico para cada cliente (ver código en Anexos).

4.1.1. Modelo de aprendizaje independiente

Al utilizar un modelo de aprendizaje independiente el tiempo de demora o anticipo de cada pedido no depende de la distribución que tengan los pedidos del cliente para todas las rutas, sino que se estima en base a la observación y , en el periodo t , para cada ruta j por separado.

$$y_{ijt} = \begin{cases} 1 & \text{si el cliente } i \text{ realiza un pedido para la ruta } j \text{ en el periodo } t \\ 0 & \text{si no} \end{cases}$$

Este modelo será utilizado para realizar tanto un pronóstico para los tiempos de demora y anticipo, como para el tiempo total de envío. De esta forma se tienen las siguientes ecuaciones:

$$N(y_{ijT}|\theta_{ijt}, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left(-\frac{1}{2} \left(\frac{y_{ijT} - \theta_{ijt}}{\sigma}\right)^2\right). \quad (4.1)$$

$$\theta_{ijt} \sim N(0, 30). \quad (4.2)$$

$$\sigma^2 \sim IG(1.05, 10). \quad (4.3)$$

Donde T corresponde al conjunto de periodos donde el cliente ha recibido un envío, y θ_{ijt} es el pronóstico realizado por un cliente i de la ruta j para el periodo t , y sigue una distribución normal como se muestra en la ecuación (4.2). Se le asigna el valor 0 a la media y 30 a la varianza en base a la distribución observada de la variable “difference” (tiempo de demora o anticipo de cada pedido), mientras que σ^2 es modelada con una distribución inverse gamma ya que son necesarios valores positivos. Es importante notar que la demora esperada para la ruta j (θ_{ij}) no influye en la distribución de las demoras de las otras rutas.

En forma análoga para el caso de los tiempos de envío, estos pronósticos son modelados con la variable τ_{ij} la cual sigue una distribución $N(0, 300)$ debido a que la distribución del tiempo total de un envío (variable “time”), tiene mayor magnitud que la variable utilizada para los tiempos de demora o anticipo, mientras que la distribución de la varianza se mantiene igual.

4.1.2. Modelo de aprendizaje jerárquico

El modelo de aprendizaje jerárquico utiliza la información de todos los pedidos del cliente en el pronóstico por ruta de los tiempos de demora o anticipo que tendrá el pedido. Para ello se define la variable θ_{ijt} como el pronostico realizado por un cliente i en el periodo t del tiempo de demora o anticipo que tendrá su pedido para la ruta j , en base a los tiempos de anticipo y demora que ha observado en periodos anteriores.

$$\theta_{ijt} \sim N(\theta_{it}, \xi). \quad (4.4)$$

Como se observa en la ecuación (4.4) se define que θ_{ijt} sigue una distribución normal con media θ_{it} y varianza ξ^2 . Donde θ_{it} es el pronóstico realizado por un cliente i para el periodo t del tiempo de demora o anticipo que tendrá su pedido, sin importar la ruta, y sigue una distribución normal como se muestra en la ecuación (4.5), donde se le asigna el valor 0 a la media y 30 a la varianza en base a la distribución observada de la variable “difference”, mientras que ξ es modelada con una distribución inverse gamma ya que son necesarios valores positivos.

$$\theta_{it} \sim N(0, 30). \quad (4.5)$$

$$\xi^2 \sim IG(1.05, 3). \quad (4.6)$$

Luego es necesario estimar el pronóstico realizado por los clientes en base a la información que han obtenido por observación de los periodos anteriores, lo cual es realizado mediante la ecuación (4.7), donde T corresponde al conjunto de periodos donde el cliente ha recibido un envío.

$$N(y_{ijT}|\theta_{ijt}, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left(-\frac{1}{2} \left(\frac{y_{ijT} - \theta_{ijt}}{\sigma}\right)^2\right). \quad (4.7)$$

$$\sigma^2 \sim IG(1.05, 10). \quad (4.8)$$

En forma análoga para el caso de los tiempos de envío, estos pronósticos son modelados con la variable τ_{it} la cual sigue una distribución $N(0, 300)$ debido a que la distribución del tiempo total de un envío (variable “time”), tiene mayor magnitud que la variable utilizada para los tiempos de demora o anticipo, mientras que la distribución de la varianza se mantiene igual.

4.2. Modelo de pronóstico con el método naïve

El método naïve consiste en utilizar como pronóstico la última observación en los datos, es decir, el tiempo pronosticado de demora o anticipo del pedido será lo registrado para el último pedido que hubiese recibido el cliente. Este modelo será utilizado para realizar tanto un pronóstico para los tiempos de demora y anticipo, como para el tiempo total de envío. Para ello se define la variable θ_{ijt} como el pronóstico realizado por un cliente i en el periodo t del tiempo de demora o anticipo que tendrá su pedido para la ruta j , el cual está dado por y_{ijt^-} , donde t^- corresponde al periodo anterior o igual a t donde hubo una observación para el cliente. En forma análoga para τ con los tiempos de envío w . De esta forma se tienen las siguientes ecuaciones:

$$\theta_{ijt} = y_{ijt^-}. \quad (4.9)$$

$$\tau_{ijt} = w_{ijt^-}. \quad (4.10)$$

4.3. Modelo elección de compra

Para el modelo de elección de compra se utiliza un modelo logit, el cual tiene como variable dependiente a y_{ijt} , definida a continuación:

$$y_{ijt} = \begin{cases} 1 & \text{si el cliente } i \text{ realiza un pedido para la ruta } j \text{ en el periodo } t \\ 0 & \text{si no} \end{cases}$$

La decisión de compra de un cliente puede verse afectada por diferentes variables y por ello se prueba con diferentes conjuntos de variables independientes para estimar el comportamiento que tendrá el cliente, donde las variables a incluir en los diferentes modelos se

encuentran en la Tabla 4.1.

Tabla 4.1: Variables independientes.

Variable	Definición (unidad)
weight	peso del envío (kg)
distance	distancia del envío (m)
first_half_week	1 si el envío es realizado en la primera mitad de la semana (binaria)
order_frequency	porcentaje de envíos en los primeros 25 periodos del cliente (número)
route_num	número de rutas que tiene el cliente (número)
domestic	1 si la ruta es doméstica (binaria)
θ	pronóstico para el tiempo de demora o anticipo del envío (número)
σ_θ	varianza de θ dentro de la ruta (número)
ξ_θ	varianza de θ entre rutas (número)
τ	pronóstico para el tiempo de envío total (número)
σ	varianza de τ dentro de la ruta (número)
ξ	varianza de τ entre rutas (número)
demand	cantidad de envíos para el resto de clientes en un periodo (número)
θ_{dist}	θ dividido por la distancia (número)

De esta forma el modelo de elección de compra esta dado por la siguiente ecuación:

$$p(y_{ijt} = 1) = \frac{\exp(\beta_0 + \beta_1 x_{1,ijt} + \beta_2 x_{2,ijt} + \dots + \beta_k x_{k,ijt})}{1 + \exp(\beta_0 + \beta_1 x_{1,ijt} + \beta_2 x_{2,ijt} + \dots + \beta_k x_{k,ijt})}. \quad (4.11)$$

Donde β_0 corresponde a la probabilidad base de un cliente de comprar, es decir, si β_0 es negativo, el cliente presenta una disposición negativa a comprar antes de tener en cuenta el resto de las variables que afectan la decisión de compra, las cuales están representadas por $x_{1\dots k,ijt}$ y son las presentadas en la tabla anterior.

Cada una de estas variables presenta un peso diferente en la decisión de compra del cliente, los cuales están representados por $\beta_{1\dots k,i}$ y son estimadas mediante simulación MCMC con un modelo de estimación Bayesiana.

4.3.1. Modelo de estimación Bayesiana

Este modelo estima la probabilidad de que un cliente i en el periodo t decida realizar un envío por la ruta j . Para esto se tienen dos posibles escenarios; el cliente realiza un pedido por alguna de las rutas o no lo hace, donde el máximo número de rutas que puede tener algún cliente es 4.

1. El cliente realiza un envío:

$$p(y_{ijt} = 1) = \frac{\exp(\beta_0 + \beta_1 x_{1,ijt} + \beta_2 x_{2,ijt} + \dots + \beta_k x_{k,ijt})}{1 + \exp(\beta_0 + \beta_1 x_{1,ijt} + \beta_2 x_{2,ijt} + \dots + \beta_k x_{k,ijt})}. \quad (4.12)$$

2. El cliente no realiza envío:

$$p(y_{ijt} = 0) = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_{1,ijt} + \beta_2 x_{2,ijt} + \dots + \beta_k x_{k,ijt})}. \quad (4.13)$$

Ahora que se tiene la probabilidad de ambos escenarios se definen dos conjuntos para t de tal forma de separar aquellos en que el cliente realiza un pedido y aquellos en que no lo hace, quedando lo siguiente; $T_{i1} \equiv \{t : \max_j(y_{ijt}) = 1\}$ y $T_{i0} \equiv \{t : \max_j(y_{ijt}) = 0\}$. Además, se define r_{it} como la ruta elegida cuando $y_{it} \equiv \sum_{j=1}^J y_{ijt} = 1$, luego $r_{it} = \sum_{j=1}^J j y_{ijt}$. De esta forma la verosimilitud de todos los datos de un cliente es la siguiente:

$$\mathcal{L}(y_i) = \prod_{t \in T_{i1}} \frac{\exp(\beta_0 + \beta'_k x_{k,ir_{it}})}{1 + \exp(\beta_0 + \beta'_k x_{k,ir_{it}})} \prod_{t \in T_{i0}} \left(\sum_{j=1}^J \frac{1}{1 + \exp(\beta_0 + \beta'_k x_{k,ijt})} \right). \quad (4.14)$$

Luego se realizan 3 variaciones de este modelo otorgando mayor o menor flexibilidad a la estimación de beta, los cuales son descritos a continuación:

Modelo Homogéneo

El modelo de elección de compra homogéneo estima un β único para cada variable independiente y es el mismo para todos los clientes, descrito por la ecuación (4.14), donde la distribución a priori para β es una $N(0, 10)$. (Ver código en Anexos)

Modelo Homogéneo Híbrido

Al igual que el modelo homogéneo, estima un β único para cada variable independiente y es el mismo para todos los clientes, sin embargo, β_0 es estimado por cliente y para cada una de las rutas, por lo que el modelo sigue la ecuación (4.15), donde β_0 ahora tiene una media y una varianza, que siguen la distribución a priori presentada a continuación:

$$\mathcal{L}(y_i) = \prod_{t \in T_{i1}} \frac{\exp(\beta_{0,ir} + \beta'_k x_{k,ir_{it}})}{1 + \exp(\beta_{0,ir} + \beta'_k x_{k,ir_{it}})} \prod_{t \in T_{i0}} \left(\sum_{j=1}^J \frac{1}{1 + \exp(\beta_{0,ir} + \beta'_k x_{k,ijt})} \right). \quad (4.15)$$

$$\mu_{\beta_0} \sim N(0, 10). \quad (4.16)$$

$$\sigma_{\beta_0}^2 \sim IG(3, 1). \quad (4.17)$$

$$\beta_{0,i,r} \sim N(\mu_{\beta_0}, \sigma_{\beta_0}). \quad (4.18)$$

Modelo Heterogéneo por ruta

Este modelo de elección de compra estima un β por cliente para cada variable independiente k , donde para los coeficientes de las variables asociadas a la calidad de servicio, se utiliza la distribución *Lognormal* y luego se multiplica por (-1), para asegurar que un peor desempeño disminuya la probabilidad de compra de los clientes. Además, β_0 es estimado por cliente, pero también es estimado para cada una de las rutas. Esto se formaliza mediante el modelo y las distribuciones a priori siguientes:

$$\mathcal{L}(y_i) = \prod_{t \in T_{i1}} \frac{\exp(\beta_{0,ir} + \beta'_{k,i} x_{k,iritt})}{1 + \exp(\beta_{0,ir} + \beta'_{k,i} x_{k,iritt})} \prod_{t \in T_{i0}} \left(\sum_{j=1}^J \frac{1}{1 + \exp(\beta_{0,ir} + \beta'_{k,i} x_{k,ijt})} \right). \quad (4.19)$$

$$\mu_{\beta_0}, \mu_{\beta_k} \sim N(0, 10). \quad (4.20)$$

$$\sigma_{\beta_0}^2, \sigma_{\beta_k}^2 \sim IG(3, 1). \quad (4.21)$$

$$\beta_{0,i,r} \sim N(\mu_{\beta_0}, \sigma_{\beta_0}). \quad (4.22)$$

$$\beta_{1,k,i} \sim N(\mu_{\beta_{1\dots k-2}}, \sigma_{\beta_{1\dots k-2}}). \quad (4.23)$$

$$\beta_{2,k,i} \sim \text{Lognormal}(\mu_{\beta_{k-1\dots k}}, \sigma_{\beta_{k-1\dots k}}). \quad (4.24)$$

Donde β_2 es el vector de coeficientes para las variables independientes asociadas a la calidad de servicio y β_1 corresponde al vector del resto de las variables.

Capítulo 5

Resultados

Se presentan los resultados obtenidos de los modelos expuestos en el capítulo anterior, donde los modelos de pronóstico son evaluados con diferentes métricas para ver cual de ellos tiene un mejor desempeño. Además se observa que variables tienen una influencia significativa en el comportamiento de compra de los clientes a partir de los estimadores obtenidos de los modelos realizados con los diferentes pronósticos de tiempo.

5.1. Modelos de pronóstico

Para cada uno de los clientes de la base de datos se realizó el pronóstico del tiempo de demora o anticipo y el tiempo total de envío de su pedido de 3 formas diferentes; la primera y mas simple, con el método Naïve y las otras dos con modelos de aprendizaje; independiente y jerárquico, los cuales son calibrados con datos de $1...T$ para estimar el pronóstico, donde T es el periodo a pronosticar y puede tomar valores $T = 2, \dots, 104$.

5.1.1. Pronóstico de los tiempos de demora o anticipo

La variable pronosticada se denomina “Demora”, esta toma un valor negativo si el pedido del cliente llega de forma anticipada y un valor positivo si llega con retraso según lo planificado, donde el mayor tiempo de demora es de 37 horas aproximadamente y el mayor tiempo de anticipo es de 31 horas, es decir, la variable tiene un rango de $[-31,37]$. Además para cada cliente se pronostica para un largo de tiempo de 104 periodos, sin embargo, debido a que los modelos de aprendizaje utilizan las observaciones acumuladas al momento del pronóstico y a que el periodo de la primera observación varia para cada cliente, es que la evaluación de cada modelo es realizada para los pronósticos desde el periodo 25 en adelante, ya que de esta forma se utilizan al menos el 70 % de los datos, sin incluir aquellos que aún se necesitan para calibrar el modelo.

Para evaluar los modelos se calculan las siguientes métricas; error absoluto máximo, error absoluto promedio y raíz del error cuadrático promedio, las cuales se presentan en la tabla 5.1 a continuación:

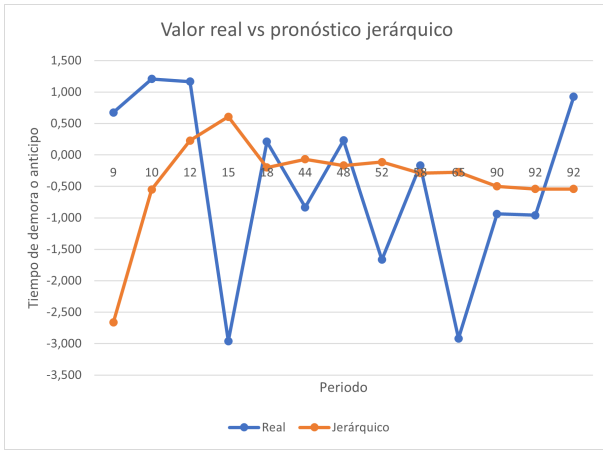
Tabla 5.1: Métricas de error para pronóstico de tiempos de demora o anticipo

Modelo o método	Métricas		
	Error absoluto máximo	Error absoluto promedio	Raíz del error cuadrático promedio
Jerárquico	42,83	1,63	2,56
Independiente	43,80	1,66	2,59
Naïve	45,81	2,07	3,26

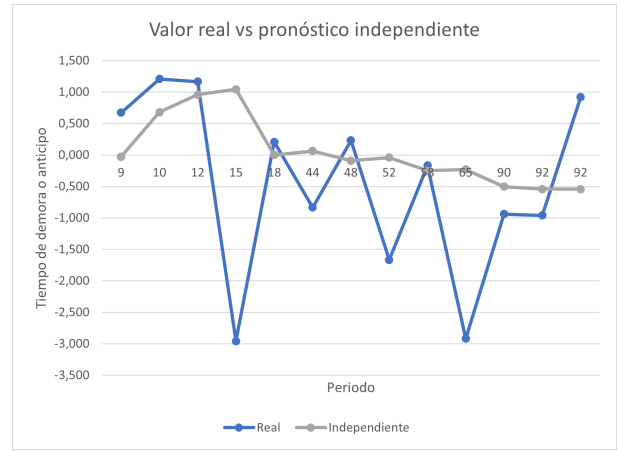
De la tabla se puede observar que para todas las métricas, el modelo que tiene mejor desempeño es el de aprendizaje jerárquico, seguido por el modelo de aprendizaje independiente y por último se encuentra el método Naïve, lo cual va en línea con la complejidad de cada uno de los modelos. Sin embargo, la diferencia en las métricas de error absoluto promedio y la raíz del error cuadrático promedio para el modelo de aprendizaje jerárquico y el independiente es de solo 0.03 unidades, por lo que, al considerar el pronóstico de los tiempos de demora o anticipo en los pedidos estos modelos tienen un desempeño similar. Por otra parte, este es un modelo que busca estimar las predicciones que tendrán los consumidores, y estos pueden o no utilizar el modelo con mejor ajuste para ello.

En cuanto al tiempo requerido para calibrar estos modelos, para el modelo de aprendizaje jerárquico fueron necesarias 51 horas y 30 minutos, para el modelo de aprendizaje independiente se necesitaron 45 horas y para el método Naïve solo fueron necesarios 5 minutos.

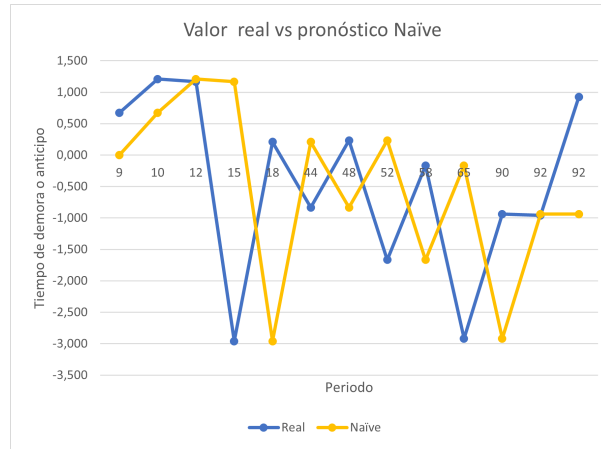
Además, se realiza un análisis para un cliente elegido de forma aleatoria, para el cual se observan todos los pedidos realizados, sin descartar los de los primeros 25 periodos, ya que interesa ver el comportamiento del pronóstico a medida que se adquieren más observaciones (ver pronóstico para los diferentes modelos en Anexos). En particular se analiza la ruta uno para los diferentes pronósticos como muestra la Figura 5.1, en la cual se puede observar que los pedidos del cliente comienzan en el periodo 9, es decir, para este periodo el valor del pronóstico debiese ser el valor de los parámetros iniciales, y esto sucede para el modelo Independiente y el método Naïve, pues toman el valor 0, pero para el modelo Jerárquico es distinto ya que tiene en cuenta la información de la ruta a pronosticar, como también la del resto de las rutas, y ya que en el periodo 7 hay un pedido para la ruta 2 que llega 2,8 horas antes, entonces el pronóstico de todas las rutas del cliente es actualizado, por lo que el pronóstico de la ruta 1 para el periodo 9 da como resultado un anticipo de 2,7 horas.



(a) Pronóstico Jerárquico.



(b) Pronóstico Independiente.



(c) Pronóstico Naïve.

Figura 5.1: Gráficos de pronósticos de los tiempos de demora o anticipo para la ruta 1 de un cliente aleatorio.

Por otra parte, tanto en el gráfico del modelo jerárquico como del modelo independiente, es posible ver que luego de 4 observaciones el pronóstico se estabiliza aproximadamente en el promedio, es decir, a partir del periodo 18 el pronóstico para ambos modelos presenta una baja variabilidad, lo cual podría dificultar el encontrar efecto de estas variables en el comportamiento de compra de este cliente, sin embargo, para otros clientes puede haber una mayor variabilidad que permita determinar el efecto de las demoras o anticipos en la preferencia de compra.

5.1.2. Pronóstico del tiempo total de envío

La variable pronosticada se denomina Time, siendo esta el tiempo total que demora el envío desde que es solicitado hasta que es recibido por el cliente, donde el menor tiempo de demora es de 3 horas aproximadamente y el mayor tiempo es de 478 horas, es decir, la variable tiene un rango de $[3,478]$. Además y al igual que con la variable para el tiempo de demora o anticipo, para cada cliente se pronostica para un largo de tiempo de 104 periodos, sin embargo, debido a que los modelos de aprendizaje utilizan las observaciones acumuladas al momento del pronóstico y a que el periodo de la primera observación varía para cada cliente,

es que la evaluación de cada modelo es realizada para los pronósticos desde el periodo 25 en adelante.

Para evaluar los modelos se calculan las siguientes métricas; error absoluto máximo, error absoluto promedio, raíz del error cuadrático promedio y error porcentual absoluto promedio, las cuales se presentan en la tabla 5.2 a continuación:

Tabla 5.2: Métricas de error para pronóstico de tiempo total de envío

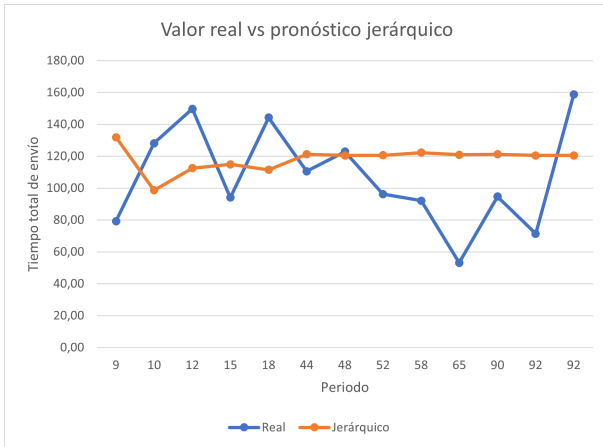
Modelo o método	Error absoluto máximo	Error absoluto promedio	Métricas	
			Raíz del error cuadrático promedio	Error porcentual absoluto promedio
Jerárquico	379,24	45,19	63,96	39,89 %
Independiente	372,24	45,01	64,07	38,45 %
Naïve	395,27	55,37	80,46	45,50 %

De la tabla se puede observar que los modelos de aprendizaje tienen un desempeño similar al igual que para el pronóstico de los tiempos de demora o anticipo, y nuevamente el método Naïve queda en último lugar. Además, a pesar de que el error absoluto promedio es de 45 aproximadamente para los modelos de aprendizaje, lo cual significa un error de casi 2 días para el pronóstico de los tiempos de despacho, esta variable tiene una varianza significativa debido al amplio rango en el que se encuentran sus valores, por lo que se considera un error aceptable para los pronósticos.

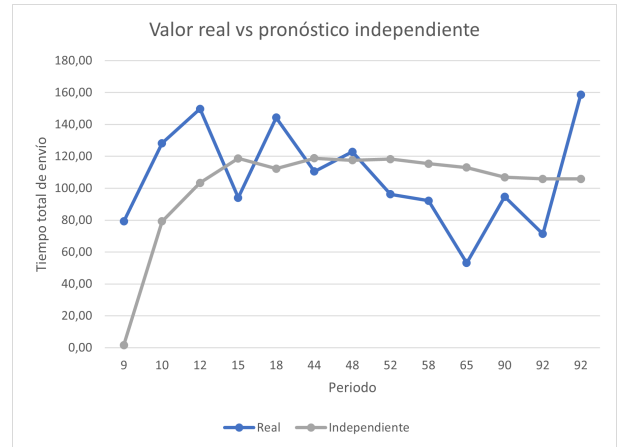
Por otra parte, se tiene el tiempo que tarda en correr cada uno de estos modelos o método, donde para el modelo de aprendizaje jerárquico fueron necesarias 71 horas, para el modelo de aprendizaje independiente se necesitaron 39 horas y para el método Naïve solo fueron necesarios 5 minutos, es decir, el modelo jerárquico a pesar de tener un desempeño similar al modelo independiente, demora casi el doble de tiempo y se tarda significativamente más si es comparado con el método Naïve. Por esto, en términos de ajuste de pronóstico y tiempo utilizado el modelo recomendado es el de aprendizaje independiente y dependiendo del tamaño de la muestra que se desee pronosticar, puede ser posible ceder en la precisión del pronóstico en post de obtener resultados en un menor tiempo utilizando el método Naïve, pues al igual que para el tiempo de demoras o anticipos, este pronóstico busca estimar la predicción que tendrá el cliente sobre el tiempo total de envío de su pedido, por lo cual podría o no utilizar el modelo con mejor ajuste para realizar su predicción.

Además, al igual que para el pronóstico de demoras o anticipos se realiza un análisis para el mismo cliente elegido de forma aleatoria, para el cual se observan todos los pedidos realizados, sin descartar los de los primeros 25 periodos, ya que interesa ver el comportamiento del pronóstico a medida que se adquieren más observaciones (ver pronóstico para los diferentes modelos en Anexos). En particular se analiza la ruta uno para los diferentes pronósticos como muestra la Figura 5.2, donde es posible ver que para el modelo jerárquico luego de 2 observaciones, es decir, desde el periodo 12 en adelante, el pronóstico se estabiliza aproximadamente en el promedio, mientras que para el modelo independiente sucede luego de 3 observaciones, es decir en el periodo 15, por lo que al igual que para los tiempos de demora o anticipo podría significar una dificultad para encontrar significancia de esta variable en el modelo de elección de compra de este cliente, ya que los valores de pronóstico de los periodos

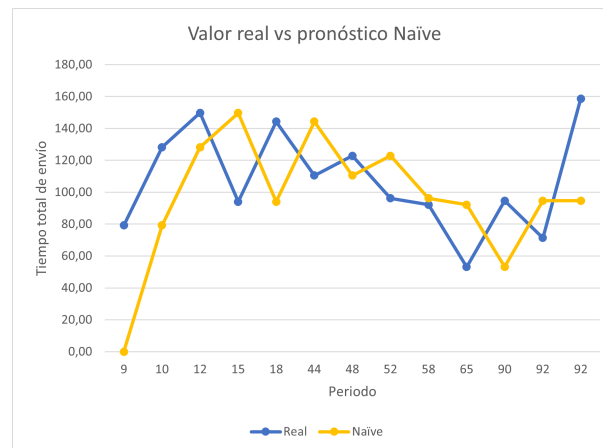
utilizados no varían mucho entre sí, sin embargo, para otros clientes puede haber una mayor variabilidad que permita determinar el efecto del tiempo de despacho en la preferencia de compra de los clientes.



(a) Pronóstico Jerárquico.



(b) Pronóstico Independiente.



(c) Pronóstico Naïve.

Figura 5.2: Gráficos de pronósticos del tiempo total de envío para la ruta 1 de un cliente aleatorio.

5.2. Modelo elección de compra con estimación Bayesiana

Para cada uno de los clientes se calibraron los modelos de elección de compra con estimación Bayesiana expuestos en el Capítulo 4 ¹, donde para cada uno fueron utilizados los datos de pronósticos de los diferentes métodos de aprendizaje. La métrica utilizada para medir el ajuste de los modelos es la verosimilitud marginal estimada mediante el método de media armónica, donde el aprendizaje jerárquico obtuvo los mejores resultados de ajuste, seguido por el aprendizaje independiente y por último el método Naïve, lo cual va en línea con la

¹ También se realizó un modelo heterogéneo y un modelo heterogéneo por ruta sin restringir el signo de los coeficientes para las variables de calidad de servicio, donde sus resultados se encuentran en el Anexo B.1.

complejidad de los modelos y las métricas de error observadas en la sección anterior.

5.2.1. Método Naïve

Para el método Naïve se tienen los datos de pronóstico tanto de la calidad de servicio sobre los tiempos de anticipo y demora, como también del tiempo total de despacho del pedido, los cuales son incluidos en el modelo de elección de compra, junto a las variables; weight, distance, order_frequency, route_num, domestic, demand, first_half_week y el intercepto. Los coeficientes del estimador de cada una de las variables independientes se encuentran en la Tabla 5.3, donde para el modelo Homogéneo híbrido y el Heterogéneo por ruta se puede observar la omisión de las variables; distance, order_frequency, route_num y domestic, ya que el efecto de estas es capturado por el intercepto, el cual es diferenciado para cada ruta del cliente. También, se incluye θ normalizado por distancia en lugar de θ por si solo, ya que al comparar los modelos con y sin esta variable, el que presentó mejor ajuste para ambos casos fue el que incluía a θ normalizado (θ_{dist}).

Tabla 5.3: Resultados para modelos de elección de compra con datos del método naïve

Variable	Homogéneo		Homogéneo híbrido		Heterogéneo por ruta	
weight	0,0018		0,0004		0,0051	
distance	-0,0062		-		-	
order_frequency	1,7258	**	-		-	
route_num	-0,2258	**	-		-	
domestic	-0,0240		-		-	
θ	-0,0069	**	-		-	
τ	-0,0132	**	-0,0060		-0,0223	**
demand	0,8531	**	0,8819	**	0,8868	**
first_half_week	0,0262		0,0264		0,0129	
θ_{dist}	-		-0,0528	**	-0,0447	**
intercepto	-3,2023	**	-3,6289	**	-3,5224	**
Verosimilitud marginal	-42.231		-40.812		-40.557	

De los coeficientes de las variables independientes se puede observar que tanto θ como θ_{dist} son negativas para los modelos homogéneos, que son los que tienen libertad de signo para esta variable y significativas al 5% para todos los modelos, lo cual significa que a mayor tiempo de atraso en los pedidos, menor será la probabilidad de que el cliente siga realizando pedidos, y lo mismo ocurre con τ , que es negativa para todos los modelos y significativa al 5% para 2 de los 3 modelos, de esta forma mientras mayor sea el tiempo de despacho de los envíos, menor será la probabilidad de los clientes de seguir comprando a través de la empresa. Además y aún cuando no está relacionado a la calidad de servicio, es importante destacar que el coeficiente de la demanda es positivo y significativo para todos los modelos, es decir, mientras mayor sea la demanda del resto de los clientes en un periodo, mayor es la

probabilidad de que el cliente realice un pedido en ese periodo.

Por otra parte, en cuanto al ajuste de los modelos, el cual se puede ver en la última fila con la verosimilitud marginal de cada modelo, se tiene que el modelo con peor ajuste es el Homogéneo y el mejor es el Heterogéneo por ruta, y es por lo cual el análisis se centrará en este modelo, donde la varianza y desviación estándar de cada una de sus variables se encuentra en la Tabla 5.4, en la cual se observa que los coeficientes que acompañan a las variables relacionadas a la calidad de servicio son las que tienen menor varianza, es decir, las preferencias de los clientes en cuanto a estas métricas no varían mucho entre sí, por lo que el efecto del cliente promedio es representativo para el resto de los clientes en cuanto a cambios que puedan surgir en la calidad de servicio.

Tabla 5.4: Coeficientes de beta, varianza y desviación estándar para modelo Heterogéneo por ruta con datos del método naïve

Variable	Media		Varianza	Desviación estándar
weight	0,0051		0,017	0,130
τ	-0,0223	**	0,000	0,002
demand	0,8868	**	0,023	0,153
first_half_week	0,0129		0,154	0,392
θ_{dist}	-0,0447	**	0,000	0,009
intercepto	-3,5224	**	0,273	0,523

Nota: la media y varianza de τ y θ_{dist} son calculadas utilizando las formulas para estos momentos correspondientes a una distribución Lognormal.

5.2.2. Aprendizaje independiente

Para el aprendizaje independiente además de tener los datos de pronóstico tanto de la calidad de servicio sobre los tiempos de anticipo y demora, como también del tiempo total de despacho del pedido, se tiene la varianza de estos pronósticos dentro de cada ruta, las cuales son incluidas en el modelo de elección de compra, junto al resto de las variables independientes. Los coeficientes del estimador de cada una de las variables independientes se encuentran en la Tabla 5.5, donde al igual que para los datos del método Naïve se omiten variables para el modelo Homogéneo híbrido y el Heterogéneo por ruta y se incluye θ normalizado por distancia en lugar de θ por si solo, ya que al comparar los modelos con y sin esta variable, el que presentó mejor ajuste para ambos casos fue el que incluía a θ normalizado.

De los coeficientes de las variables independientes se puede observar que τ es negativa para los modelos homogéneos, que son los que tienen libertad de signo para esta variable y significativas al 5% para todos los modelos, lo cual quiere decir que es posible ver un efecto negativo de los tiempos de despacho en la probabilidad de compra del cliente, es decir, que al aumentar los tiempos de despacho disminuye la probabilidad de que los clientes prefieran la empresa. Por otra parte el coeficiente de σ es positivo y significativo para los modelos homogéneos, así mientras mayor sea la varianza del tiempo de despacho para una ruta, mayor será la probabilidad de los clientes de seguir realizando pedidos, sin embargo, esto no hace sentido con una mejor calidad de servicio, y como se ve en el modelo heterogéneo

esta variable deja de ser significativa, por lo que se atribuye su significancia en los modelos homogéneos a la falta de flexibilidad para las preferencias de los clientes.

Tabla 5.5: Resultados para modelos de elección de compra con datos del aprendizaje independiente

Variable	Homogéneo		Homogéneo híbrido		Heterogéneo por ruta	
weight	0,0022		0,0006		0,0050	
distance	-0,0074	**	-		-	
order_frequency	1,7275	**	-		-	
route_num	-0,2312	**	-		-	
domestic	-0,0341		-		-	
θ	-0,0131	**	-		-	
σ_θ	-0,0524	**	-0,0325	**	-0,0748	**
τ	-0,0234	**	-0,0235	**	-0,0631	**
σ	0,0025	**	0,0023	**	0,0028	
demand	0,8621	**	0,8913	**	0,9406	**
first_half_week	0,0244		0,0268		0,0263	
θ_{dist}	-		-0,1170	**	-2,2186	**
intercepto	-3,1505	**	-3,5989	**	-3,6632	**
Verosimilitud marginal	-42.218		-40.810		-40.268	

Además, se tiene que tanto θ como θ_{dist} son negativas para los modelos homogéneos, que son los que tienen libertad de signo para esta variable y significativas al 5% para todos los modelos, por lo cual a mayor tiempo de atraso en los pedidos, menor será la probabilidad de que el cliente siga realizando pedidos y lo mismo ocurre con σ_θ , que es negativa y significativa para los 3 modelos, de esta forma mientras mayor sea la varianza del tiempo de demora o anticipo dentro de la ruta, menor será la probabilidad de los clientes de seguir comprando a través de la empresa, lo cual quiere decir que los clientes prefieren que sus tiempos de atraso o anticipo en los pedidos se mantengan en un rango pequeño y no varíen mucho entre sí. Además, al igual que en los modelos con los datos del método Naïve el coeficiente de la demanda es positivo y significativo, es decir, mientras mayor sea la demanda del resto de los clientes en un periodo, mayor es la probabilidad de que el cliente realice un pedido en ese periodo.

Tabla 5.6: Coeficientes de beta, varianza y desviación estándar para modelo Heterogéneo por ruta con datos del aprendizaje independiente

Variable	Media		Varianza	Desviación estándar
weight	0,0050		0,018	0,134
σ_θ	-0,0748	**	0,056	0,236
τ	0,0631	**	0,014	0,118
σ	0,0028		0,003	0,059
demand	0,9406	**	0,038	0,195
first_half_week	0,0263		0,219	0,468
θ_{dist}	-2,2186	**	229,873	15,162
intercepto	-3,6632	**	3,113	1,764

Nota: la media y varianza de τ y θ_{dist} son calculadas utilizando las formulas para estos momentos correspondientes a una distribución Lognormal.

Luego, en cuanto al ajuste de los modelos, el cual se puede ver en la última fila con la verosimilitud marginal de cada modelo, se tiene que el modelo con peor ajuste es el Homogéneo y el mejor es el Heterogéneo por ruta, y es por lo cual el análisis se centrará en este modelo, donde la varianza y desviación estándar de cada una de sus variables se encuentra en la Tabla 5.6, en la cual se observa que a pesar de que θ_{dist} es negativo y significativo, su varianza y desviación estándar son significativamente altas, por lo que hay clientes para los cuales el tiempo de demoras y anticipos normalizado por la distancia tiene un efecto positivo, es decir, clientes para los cuales al aumentar la demora de un pedido se esta aumentando la probabilidad de que sigan pidiendo, sin embargo, para el cliente promedio el efecto sería el contrario. Por otro lado, la demanda y el intercepto son las variables con menor heterogeneidad entre clientes, por lo cual el coeficiente del cliente promedio es una buena representación para las preferencias de los clientes respecto de estas variables.

5.2.3. Aprendizaje jerárquico

Para el aprendizaje jerárquico se tienen los datos de la varianza de los pronósticos tanto dentro de cada ruta, como entre rutas, las cuales son añadidas al modelo de elección de compra junto al resto de las variables independientes. Los coeficientes del estimador de cada una de las variables independientes se encuentran en la Tabla 5.7, donde al igual que para los datos del método Naïve y el aprendizaje independiente se omiten variables para el modelo Homogéneo híbrido y el Heterogéneo por ruta y se incluye θ normalizado por distancia en lugar de θ por si solo, ya que al comparar los modelos con y sin esta variable, el que presentó mejor ajuste para ambos casos fue el que incluía a θ normalizado.

De los coeficientes de las variables independientes se puede observar que τ es negativa para los modelos homogéneos, que son los que tienen libertad de signo para esta variable y significativa al 5% para todos los modelos, por lo que se puede decir que un aumento en los tiempos de despacho disminuye la probabilidad de estos de seguir pidiendo. Además, al considerar que las preferencias de los clientes son diferentes entre sí en el modelo heterogéneo, se tiene que ξ es negativa y significativa, por lo que a mayor varianza entre los tiempos de

despacho de las rutas menor será la probabilidad de que el cliente realice un envío, pero al considerar las mismas preferencias para todos los clientes en el modelo homogéneo el efecto es el contrario, sin embargo, como el modelo heterogéneo presenta un mejor ajuste que el modelo homogéneo, se tendrán en cuenta los resultados del modelo con mejor ajuste.

Tabla 5.7: Resultados para modelos de elección de compra con datos del aprendizaje jerárquico

Variable	Homogéneo		Homogéneo híbrido		Heterogéneo por ruta	
weight	0,0024		0,0006		0,0038	
distance	-0,0068	*	-		-	
order_frequency	1,7238	**	-		-	
route_num	-0,2487	**	-		-	
domestic	-0,0267		-		-	
θ	-0,0163	**	-		-	
σ_θ	-0,0417	**	-0,0489	**	-0,0641	
ξ_θ	-0,0343		0,3556	**	0,1110	
τ	-0,0258	**	-0,0244	**	-0,0446	**
σ	0,0020	**	0,0010		0,0024	
ξ	0,0014	*	-0,0014		-0,0298	**
demand	0,8634	**	0,8792	**	0,9359	**
first_half_week	0,0262		0,0266		0,0298	
θ_{dist}	-		-0,1159	**	-2,8217	**
intercepto	-3,0500	**	-4,2135	**	-3,7488	**
Verosimilitud marginal	-42.227		-40.808		-40.230	

Por otra parte, y al igual que para los modelos realizados con los datos del método Naïve y el aprendizaje independiente, tanto θ como θ_{dist} son negativas para los modelos homogéneos, que son los que tienen libertad de signo para esta variable y significativas al 5% para todos los modelos, lo cual significa que a mayores tiempo de atraso en los pedidos, menor será la probabilidad de que el cliente siga realizando pedidos y lo mismo ocurre con σ_θ , que es negativa y significativa para los modelos homogéneos, por lo cual a mayor varianza del tiempo de demora o anticipo dentro de la ruta, menor será la probabilidad de los clientes de seguir comprando a través de la empresa.

Además, al observar el modelo homogéneo sin diferenciar el intercepto por ruta, se tiene que el coeficiente que acompaña a la variable de la distancia es negativo y significativo, por lo cual a mayor distancia del pedido menor será la probabilidad de que el cliente prefiera la empresa, así como también a mayor cantidad de rutas que utilice el cliente es menos probable que realice un envío. También se puede notar que mientras mayor sea la cantidad de pedidos realizados por el cliente en los primeros 25 periodos, mayor será la probabilidad de que el cliente vuelva a pedir.

Por último, se puede ver que el intercepto es negativo y significativo para todos los modelos, es decir, los clientes tienen una predisposición negativa a realizar pedidos, es decir, que si todas las variables independientes se fijaran en 0, el cliente no realizaría un pedido en la empresa, por lo cual es necesario aumentar el valor de las variables positivas y significativas, y disminuir el de las negativas para suplir este efecto, sin embargo, para el modelo Heterogéneo por ruta, el cual es el modelo de interés al tener el mejor ajuste respecto al resto de los modelos, las únicas variables significativas son; el pronóstico del tiempo de despacho, la varianza del tiempo de despacho entre rutas, la demanda y el pronóstico del tiempo de demora o anticipo normalizado por la distancia, donde no es posible aumentar la demanda y la varianza del tiempo de despacho entre rutas esta sujeta a la heterogeneidad de las rutas de cada cliente, por lo cual es necesario disminuir los tiempos de despacho y los tiempos de atraso o anticipo en los pedidos, lo cual puede ser realizado al mejorar el pronóstico de la fecha y hora de llegada del envío.

Tabla 5.8: Coeficientes de beta, varianza y desviación estándar para modelo Heterogéneo por ruta con datos del aprendizaje jerárquico

Variable	Media		Varianza	Desviación estándar
weight	0,0038		0,018	0,134
σ_θ	-0,0641		0,063	0,250
ξ_θ	0,1110		0,068	0,260
τ	-0,0446	**	0,000	0,008
σ	0,0024		0,004	0,060
ξ	-0,0298	**	0,021	0,145
demand	0,9359	**	0,040	0,200
first_half_week	0,0298		0,226	0,475
θ_{dist}	-2,8217	**	431,851	20,781
intercepto	-3,7488	**	0,308	0,555

Nota: la media y varianza de τ y θ_{dist} son calculadas utilizando las formulas para estos momentos correspondientes a una distribución Lognormal.

Luego, se tiene la varianza y desviación estándar de cada una de las variables independientes del modelo Heterogéneo por ruta en la Tabla 5.8, en la cual se observa que θ_{dist} es la variable con la varianza más alta al igual que para los datos del aprendizaje independiente, siendo aún mayor que para el caso anterior, ya que la desviación estándar es 7 veces el valor del coeficiente, es decir, tiene una gran heterogeneidad entre clientes, por lo cual a pesar de que su coeficiente es negativo y significativo para el promedio de los clientes, aún existe un número de clientes para los cuales el efecto de las demoras aumenta la probabilidad de realizar un envío. Por otra parte, la variable con menor heterogeneidad entre clientes es la demanda, lo cual tiene sentido ya que esta variable es la suma de los pedidos realizados por el resto de clientes para un periodo, por lo que entre clientes la demanda no tiene una diferencia significativa.

Capítulo 6

Contrafactuales

Interesa ver de que forma afecta en el comportamiento de compra del consumidor una disminución en la calidad de servicio, es decir, ver cuanto disminuye la probabilidad de compra de los clientes si aumenta el tiempo que se demora su pedido luego de la fecha y hora planificada, o al aumentar el tiempo de despacho total del envío, y es por lo que se analizan estos escenarios. De esta forma se contrasta la probabilidad de pedir de los clientes en el escenario original y en uno con peor calidad de servicio. Esto es realizado con los coeficientes estimados con el modelo heterogéneo por ruta y para cada uno de los métodos de aprendizaje, de los cuales el que presenta mayor efecto en la probabilidad de compra es el aprendizaje jerárquico.

6.1. Aumento del tiempo de demora y despacho en todas las rutas

Se aumentan los tiempos de demora y despacho para cada ruta en una desviación estándar del tiempo de demora de cada ruta, por ejemplo, si un cliente tiene 2 rutas, se calcula la desviación estándar de los tiempos de demora de la ruta 1 y se suma este tiempo a todas las observaciones de los tiempos de demora de la ruta 1 para ese cliente, así como también se suma a los tiempos de despacho, y se repite el proceso con la ruta 2. Luego de aumentar los tiempos de demora y despacho observados por el cliente, se vuelve a calcular el pronóstico del último periodo estudiado, recalibrando el modelo de aprendizaje del cliente considerando todas las observaciones anteriores.

Una vez que se tiene el nuevo pronóstico de cada cliente para el tiempo de demora y despacho, se calcula la probabilidad de compra antes del cambio y después de este con los coeficientes obtenidos del modelo heterogéneo por ruta. De esta forma se obtiene la suma de la probabilidad de compra de todos los clientes con los tiempos de demora y despacho originales y luego de aumentar estos tiempos, donde en la Tabla 6.1 se puede ver la diferencia porcentual de ambas sumas de probabilidad para diferentes escenarios:

- Aumento del tiempo de demora: se aumentan los tiempos de demora de todas las rutas en una desviación estándar de los tiempos de demora de cada ruta, es decir, un escenario donde se prometió al cliente una fecha y hora de llegada anterior a la original.
- Aumento del tiempo de despacho: se aumentan los tiempos de despacho de todas las

rutas en una desviación estándar de los tiempos de demora de cada ruta, es decir, un escenario donde se prometió al cliente una fecha y hora de llegada posterior a la original, pero el pedido tuvo un tiempo de despacho mayor por lo que el cliente observa el mismo tiempo de demora o anticipo en su envío.

- Aumento del tiempo de demora y despacho: se aumenta el tiempo de demora y despacho de todas las rutas en una desviación estándar de los tiempos de demora de cada ruta, es decir, un escenario donde se mantiene la fecha y hora de llegada prometida al cliente, pero el pedido tiene un tiempo de despacho mayor por lo que el cliente observa un aumento en el tiempo de demora y despacho de su pedido.

Tabla 6.1: Porcentaje de cambio en la demanda total bajo distintos escenarios y modelos de aprendizaje

Método de aprendizaje	Aumento tiempo demora	Aumento tiempo despacho	Aumento tiempo demora y despacho
Jerárquico	-20,79 %	-0,23 %	-21,05 %
Independiente	-18,78 %	-0,26 %	-18,98 %
Naïve	-1,83 %	-0,19 %	-2,02 %

De la Tabla 6.1 es posible ver que para todos los escenarios la demanda total disminuye, lo cual va en línea con el signo de los coeficientes estimados para el modelo heterogéneo por ruta y con el supuesto de que al empeorar la calidad de servicio, el cliente disminuye su probabilidad de compra. Además se observa que la mayor caída en la demanda es al calcular el pronóstico con el modelo de aprendizaje jerárquico en el escenario en donde tanto los tiempos de demora como los tiempos de despacho aumentan, disminuyendo así un 21 % la demanda total al aumentar en promedio 2 horas el tiempo de demora del pedido y el tiempo total del envío.

Además, se tiene que la disminución de la demanda para los casos en que solo se aumenta el tiempo de despacho es menor al 0.5 % para todos los métodos de aprendizaje, lo cual puede estar dado por dos razones; una es que aumentar 2 horas el tiempo de despacho no causa un efecto significativo en la calidad de servicio, pues en la práctica estos tiempos de despacho se encuentran entre 5 a 6 días en promedio, y la otra es que el coeficiente estimado para τ , que es la variable para el pronóstico de los tiempos de despacho, varía entre -0.06 a -0.02 para los 3 métodos de aprendizaje, por lo cual un cambio en ella no produce un efecto significativo en la probabilidad de compra.

Por último, para dos de los tres escenarios el método de aprendizaje jerárquico es el que tiene mayor efecto en la demanda total de los clientes al disminuir la calidad de servicio, sin embargo el método independiente es bastante similar, por lo cual es bueno ver como se comportan ambos al cambiar la calidad de servicio en solo una de las rutas.

6.2. Aumento del tiempo de demora y despacho en una de las rutas

Se aumentan los tiempos de demora y despacho para una de las rutas, en este caso la ruta 1. El aumento corresponde a una desviación estándar del tiempo de demora de esa ruta, por ejemplo, si un cliente tiene 2 rutas, se calcula la desviación estándar de los tiempos de demora de la ruta 1 y se suma este tiempo a todas las observaciones de los tiempos de demora de la ruta 1 para ese cliente, mientras que para la ruta 2 se mantienen los tiempos originales. Luego de aumentar los tiempos de demora y despacho observados por el cliente, se vuelve a calcular el pronóstico del último periodo estudiado, recalibrando el modelo de aprendizaje del cliente considerando todas las observaciones anteriores.

Una vez que se tiene el nuevo pronóstico de cada cliente para el tiempo de demora y despacho, se calcula la probabilidad de compra antes del cambio y después de este con los coeficientes obtenidos del modelo heterogéneo por ruta. De esta forma se obtiene la suma de la probabilidad de compra de todos los clientes que utilizaron la ruta 1 con los tiempos de demora y despacho originales y luego de aumentar estos tiempos, donde en la Tabla 6.2 se puede ver la diferencia porcentual de ambas sumas de probabilidad para diferentes escenarios:

- Aumento del tiempo de demora: se aumentan los tiempos de demora de la ruta 1 en una desviación estándar de los tiempos de demora de esa ruta, es decir, un escenario donde se prometió al cliente una fecha y hora de llegada anterior a la original.
- Aumento del tiempo de despacho: se aumentan los tiempos de despacho de la ruta 1 en una desviación estándar de los tiempos de demora de esa ruta, es decir, un escenario donde se prometió al cliente una fecha y hora de llegada posterior a la original, pero el pedido tuvo un tiempo de despacho mayor por lo que el cliente observa el mismo tiempo de demora o anticipo en su envío.
- Aumento del tiempo de demora y despacho: se aumenta el tiempo de demora y despacho de la ruta 1 en una desviación estándar de los tiempos de demora de esa ruta, es decir, un escenario donde se mantiene la fecha y hora de llegada prometida al cliente, pero el pedido tiene un tiempo de despacho mayor por lo que el cliente observa un aumento en el tiempo de demora y despacho de su pedido.

Para cada uno de estos escenarios debería observarse un efecto negativo en la demanda de los clientes para la ruta 1, y se esperaría que este efecto fuera mayor para el modelo de aprendizaje independiente en comparación al modelo de aprendizaje jerárquico. Esto debido a que el modelo jerárquico considera las observaciones de las otras rutas al generar el pronóstico del periodo siguiente. De esta forma, debido a que las rutas 2, 3 y 4 no fueron alteradas en su calidad de servicio, esto podría disminuir el impacto en la baja de la calidad de servicio esperada para la ruta 1, y por lo cual se esperaría que el efecto en la demanda para este modelo fuese menor que para el modelo independiente, que solo tiene en consideración las observaciones de la ruta 1 para realizar el pronóstico. La siguiente tabla muestra que esta lógica no se cumple.

Tabla 6.2: Porcentaje de cambio en la demanda de la ruta 1 bajo distintos escenarios y modelos de aprendizaje

Método de aprendizaje	Aumento tiempo demora	Aumento tiempo despacho	Aumento tiempo demora y despacho
Jerárquico	-20,95 %	-0,39 %	-21,26 %
Independiente	-20,02 %	-0,26 %	-20,22 %
Naïve	-2,00 %	-0,19 %	-2,18 %

De la Tabla 6.2 es posible ver que para todos los escenarios la demanda total de la ruta 1 efectivamente disminuye, sin embargo el modelo jerárquico es el que presenta la mayor disminución. Esto se debe a que el coeficiente asociado a las demoras (θ_{dist}) tiene una mayor magnitud en el modelo de compra estimado utilizando un modelo jerárquico de aprendizaje (-2,8217) en comparación al obtenido bajo un modelo de aprendizaje independiente (-2,2186).

Para ilustrar este fenómeno, se realizan nuevamente los cálculos del efecto en la demanda para el modelo de aprendizaje independiente, pero esta vez se utilizan los coeficientes del modelo de conversión estimados con el modelo de aprendizaje jerárquico.

Tabla 6.3: Porcentaje de cambio en la demanda de la ruta 1 para diferentes escenarios, utilizando los coeficientes del modelo de compra jerárquico.

Método de aprendizaje	Aumento tiempo demora	Aumento tiempo despacho	Aumento tiempo demora y despacho
Jerárquico	-20,95 %	-0,39 %	-21,26 %
Independiente	-22,55 %	-0,41 %	-22,85 %

La Tabla 6.3 evidencia que después de controlar por la magnitud de los coeficientes del modelo de conversión, el modelo de aprendizaje independiente genera una mayor caída en la demanda de la ruta 1. Además se observa que la mayor caída en la demanda se obtiene bajo el escenario en donde tanto los tiempos de demora como los tiempos de despacho aumentan, disminuyendo así cerca de un 23 % la demanda de la ruta 1 al aumentar en promedio 2 horas el tiempo de demora de los pedidos y el tiempo total de los envíos de la ruta 1.

Capítulo 7

Conclusiones

En esta tesis se analizó la sensibilidad de clientes respecto a la calidad de servicio de una empresa de distribución logística (freight forwarder), la cual fue medida a través de dos variables; el tiempo de despacho de cada pedido, y el tiempo de demora o anticipo del envío. Para ello se propuso dos modelos de aprendizaje; uno jerárquico, el cual utiliza la información de la calidad de servicio de las diferentes rutas que utiliza el cliente para actualizar las creencias del cliente para el periodo y ruta a estimar, y uno independiente, el cual solo tiene en consideración los datos de la ruta que esta siendo pronosticada. Además se generó un tercer modelo utilizando el método Naïve, el cual entrega la información del último periodo observado como pronóstico para el siguiente pedido.

Una vez que se tenían las creencias de los clientes para ambas variables sobre la calidad de servicio, estas se utilizaron en conjunto con otras variables independientes, para calibrar 3 modelos de elección de compra con estimación Bayesiana para cada uno de los modelos o método de pronóstico realizados. El primero, un modelo homogéneo que no considera diferencias en las preferencias de los clientes, por lo cual estima un β único para cada variable independiente. El segundo, un modelo homogéneo híbrido que a pesar de no considerar diferencias en las preferencias de los clientes, si considera el intercepto diferenciado por ruta y por cliente. El tercero y último, un modelo heterogéneo por ruta, el cual estima el intercepto diferenciado por ruta y por cliente, y un β por cliente, considerando de esta forma las diferencias en las preferencias de los mismos, donde para los coeficientes de las variables asociadas a la calidad de servicio, fue utilizada una distribución *Lognormal*, para de esta forma al luego ser multiplicada por (-1) asegurar la obtención de un efecto negativo de estas variables sobre la probabilidad de compra del cliente.

De los 3 modelos de elección de compra, y cada uno con las creencias sobre la calidad de servicio estimadas con los diferentes modelos y método de pronóstico, se obtuvo que los tiempos de demora de los envíos, así como los tiempos de despacho, tienen un efecto negativo en la probabilidad de compra del cliente sin importar si las preferencias de estos eran consideradas como iguales o diferentes entre ellos, es decir, el que un pedido llegue tarde o tenga un mayor tiempo de despacho disminuye la probabilidad de que el cliente siga prefiriendo la empresa.

Lo anterior fue visto claramente al realizar los diferentes escenarios sobre una peor calidad de servicio para el cliente, donde para el modelo de aprendizaje jerárquico es donde hubo

un mayor impacto en la caída de la demanda total de los clientes, ya que al aumentar en promedio 2 horas los tiempos que llegaban tarde los pedidos y los tiempos de despacho para todas las rutas de los clientes, se vio una disminución de un 21 % en la demanda total de los clientes, mientras que al empeorar el servicio solo de la ruta 1, se observó una disminución de un 23 % para el modelo de aprendizaje independiente.

Por otra parte, en relación al desempeño de los modelos sobre la estimación de las creencias de los consumidores, si bien el modelo de aprendizaje jerárquico tuvo los mejores resultados para las métricas de error y para el efecto en la elección de compra del cliente al disminuir la calidad de servicio ofrecida por la empresa, este modelo tiene un tiempo de cómputo significativo, ya que para estimar las creencias de 677 clientes con 104 periodos cada uno, fueron necesarias 51 horas y 30 minutos para los tiempos de demora o anticipo, y 71 horas para los tiempos de despacho, mientras que para el modelo independiente, el cual tiene un desempeño similar en cuanto a los resultados entregados, estos números se reducen a 45 horas y 39 horas respectivamente. Por esto y debido a que estos modelos buscan aproximar las creencias de consumidores, los cuales no necesariamente utilizarán el modelo con mejor ajuste para predecir la calidad de servicio de su siguiente pedido, es que se recomienda el uso del modelo de aprendizaje independiente.

Una de las limitaciones sobre los modelos de pronóstico sobre la calidad de servicio de los consumidores es que solo consideran la información de la calidad de servicio anterior observada por el cliente, pero no tienen en consideración otras variables que podrían afectar su creencia sobre la calidad de servicio del siguiente pedido, como podría ser el peso del paquete a enviar, la distancia que debe recorrer, el volumen del envío, entre otras variables, ya que por ejemplo, un consumidor podría dar menor peso a la demora de un envío que debe recorrer una mayor distancia, que a otro con el mismo tiempo de demora, pero que recorre una distancia menor.

Para trabajo a futuro se proponen otros modelos de pronóstico que por limitaciones de tiempo no pudieron ser implementados, como por ejemplo, un modelo que otorgue mayor peso a información de periodos más cercanos a la fecha a estimar y menor peso a información de periodos más antiguos, o un modelo que solo tome la información de los últimos 20 o 25 periodos del cliente para estimar la creencia sobre el siguiente periodo, de esta forma, si el cliente tuvo una mala calidad de servicio en sus primeros pedidos, pero más adelante esta mejoró, se tendrá en cuenta la información más reciente al momento de estimar las creencias del siguiente periodo, ya que información con más antigüedad podría haber sido olvidada por el consumidor o tener menor importancia.

Finalmente, se concluye, que la disminución en la calidad de servicio relacionada a los tiempos de despacho y demora de los envíos de una empresa de distribución logística tienen un impacto negativo en la probabilidad de compra de sus clientes, por lo cual es importante invertir en la mejora de estos aspectos para asegurar una mayor retención de los consumidores que posee la empresa.

Bibliografía

- [1] *Introduction to the Logistic Regression Model*, chapter 1, pages 1–30. John Wiley Sons, Ltd, 2000. ISBN 9780471722144. doi: <https://doi.org/10.1002/0471722146.ch1>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/0471722146.ch1>.
- [2] Christophe Andrieu and Johannes Thoms. A tutorial on adaptive mcmc. *Statistics and computing*, 18(4):343–373, 2008.
- [3] Ruth N. Bolton and Katherine N. Lemon. A dynamic model of customers’ usage of services: Usage as an antecedent and consequence of satisfaction. *Journal of Marketing Research*, 36(2):171–186, 1999. ISSN 00222437. doi: 10.2307/3152091.
- [4] Tat Y Chan and Barton H Hamilton. Learning , Private Information , and the Economic Evaluation of Randomized Experiments. 114(6), 2006.
- [5] Etelvina Chavez, Gastón Milanesi, and Gabriela Pesce. Funciones de utilidad y estimación de la aversión al riesgo. *Escritos Contables y de Administración*, 7(2):97–118, 2016. ISSN 1853-2055.
- [6] Andrew T. Ching, Tülin Erdem, and Michael P. Keane. Learning models: An assessment of progress, challenges, and new developments. *Marketing Science*, 32(6):913–938, 2013. ISSN 1526548X. doi: 10.1287/mksc.2013.0805.
- [7] Tülin Erdem and Michael P. Keane. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science*, 15(1):1–20, 1996. ISSN 07322399. doi: 10.1287/mksc.15.1.1.
- [8] Andrew Gelman, John B. Carlin, Hal S. Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. Bayesian Data Analysis Third edition (with errors fixed as of 13 February 2020). (February):677, 2020. URL [http://files/25/Gelmanm.fl.-BayesianDataAnalysisThirdedition\(witherrors\).pdf](http://files/25/Gelmanm.fl.-BayesianDataAnalysisThirdedition(witherrors).pdf).
- [9] Robert C. Hill. When the going gets rough: A baldrige award winner on the line. *The Academy of Management Executive (1993-2005)*, 7(3):75–79, 1993. ISSN 10795545. URL <http://www.jstor.org/stable/4165137>.
- [10] Matthew D. Hoffman and Andrew Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo, 2011.
- [11] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.
- [12] Sridhar Narayanan and Puneet Manchanda. Heterogeneous Learning and the Targeting of Marketing Communication for New Products Heterogeneous Learning and the Targeting of Marketing Communication for New Products. (August 2017), 2009. doi:

10.1287/mksc.1080.0410.

- [13] Adrian Payne and John Rickard. Relationship marketing, customer retention and service firm profitability. 1993.
- [14] John H. Roberts and Glen L. Urban. Modeling Multiattribute Utility, Risk, and Belief Dynamics for New Consumer Durable Brand Choice. *Management Science*, 34(2):167–185, 1988. ISSN 0025-1909. doi: 10.1287/mnsc.34.2.167.
- [15] Roland T. Rust, Anthony J. Zahorik, and Timothy L. Keiningham. Return on Quality (ROQ): Making Service Quality Financially Accountable. *Journal of Marketing*, 59(2): 58, 1995. ISSN 00222429. doi: 10.2307/1252073.
- [16] Richard Staelin William Boulding, Ajay Kalra and Valarie A. Zeithaml. A Dynamic Process Model of Service Quality : From Expectations to Behavioral Intentions . *Journal of Marketing Research*, 30(1):7–27, 2012.

Anexo A

Código en R

A.1. Modelos de aprendizaje para el pronóstico del tiempo de demora o anticipo

A.1.1. Modelo de aprendizaje independiente

1. Primero es necesario crear una base de datos con el formato necesario para correr el modelo, lo cual se realiza con un código en R ([ver link](#)) y es guardada en un archivo “.RData”.
2. Luego se corre el modelo de stan ([ver link](#)).
3. Para correr el modelo de stan se utiliza otro código en R ([ver link](#)), el cual guarda los datos del modelo en el archivo “.RData” anterior.

A.1.2. Modelo de aprendizaje jerárquico

1. Primero es necesario crear una base de datos con el formato necesario para correr el modelo, lo cual se realiza con un código en R ([ver link](#)) y es guardada en un archivo “.RData”.
2. Luego se corre el modelo de stan ([ver link](#)).
3. Para correr el modelo de stan se utiliza otro código en R ([ver link](#)), el cual guarda los datos del modelo en el archivo “.RData” anterior.

A.2. Modelo de aprendizaje para el pronóstico del tiempo total de envío

A.2.1. Modelo de aprendizaje independiente

1. Primero es necesario crear una base de datos con el formato necesario para correr el modelo, lo cual se realiza con un código en R ([ver link](#)) y es guardada en un archivo “.RData”.
2. Luego se corre el modelo de stan ([ver link](#)).

3. Para correr el modelo de stan se utiliza otro código en R ([ver link](#)), el cual guarda los datos del modelo en el archivo “.RData” anterior.

A.2.2. Modelo de aprendizaje jerárquico

1. Primero es necesario crear una base de datos con el formato necesario para correr el modelo, lo cual se realiza con un código en R ([ver link](#)) y es guardada en un archivo “.RData”.
2. Luego se corre el modelo de stan ([ver link](#)).
3. Para correr el modelo de stan se utiliza otro código en R ([ver link](#)), el cual guarda los datos del modelo en el archivo “.RData” anterior.

A.3. Modelo elección de compra

A.3.1. Modelo Homogéneo en stan

Modelo homogéneo de estimación Bayesiana en stan ([ver link](#)).

A.3.2. Modelo Homogéneo híbrido en stan

Modelo homogéneo híbrido de estimación Bayesiana en stan ([ver link](#)).

A.3.3. Modelo Heterogéneo por ruta en stan

Modelo heterogéneo por ruta de estimación Bayesiana en stan ([ver link](#)).

A.3.4. Estimación Bayesiana de elección de compra

1. Primero es necesario crear una base de datos con el formato necesario para correr el modelo, lo cual se realiza con un código en R ([ver link](#)) y es guardada en un archivo tipo “my_base_modelo_X.RData”.
2. Luego se corre alguno de los modelos de stan, ya sea el homogéneo, homogéneo híbrido o el heterogéneo por ruta.
3. Para correr los modelos de stan se utiliza otro código en R ([ver link](#)), el cual guarda los datos del modelo en un archivo tipo “modelo_X.RData”.

A.4. Contrafactuales

A.4.1. Aumento del tiempo de demora

Cálculo del efecto en la probabilidad de aumentar los tiempos de demora ([ver link](#)).

A.4.2. Aumento del tiempo de despacho

Cálculo del efecto en la probabilidad de aumentar los tiempos de despacho ([ver link](#)).

A.4.3. Aumento del tiempo de demora y despacho

Cálculo del efecto en la probabilidad de aumentar los tiempos de demora y despacho ([ver link](#)).

Anexo B

Resultados

B.1. Modelo elección de compra con estimación Bayesiana

Resultados de los modelos de elección de compra con libertad de signo para las variables τ y θ_{dist} con los datos de pronóstico de los diferentes métodos de aprendizaje.

B.1.1. Método Naïve

Para el método Naïve se tienen los datos de pronóstico tanto de la calidad de servicio sobre los tiempos de anticipo y demora, como también del tiempo total de despacho del pedido, los cuales son incluidos en el modelo de elección de compra, junto a las variables; weight, distance, order_frequency, route_num, domestic, demand, first_half_week y el intercepto. Los coeficientes del estimador de cada una de las variables independientes se encuentran en la Tabla B.1, donde para el modelo Homogéneo híbrido y el Heterogéneo por ruta se puede observar la omisión de las variables; distance, order_frequency, route_num y domestic, ya que el efecto de estas es capturado por el intercepto, el cual es diferenciado para cada ruta del cliente, y también, se incluye θ normalizado por distancia en lugar de θ por si solo, ya que al comparar los modelos con y sin esta variable, el que presentó mejor ajuste para ambos casos fue el que incluía a θ normalizado (θ_{dist}).

De los coeficientes de las variables independientes se puede observar que tanto θ como θ_{dist} son negativas y significativas al 5% para la mayoría de los modelos, a excepción del modelo Heterogéneo, donde no es significativa, lo cual significa que a mayores tiempo de atraso en los pedidos, menor será la probabilidad de que el cliente siga realizando pedidos y lo mismo ocurre con τ , que es negativa y significativa para 3 de los 4 modelos, de esta forma mientras mayor sea el tiempo de despacho de los envíos, menor será la probabilidad de los clientes de seguir comprando a través de la empresa. Además y aún cuando no esta relacionado a la calidad de servicio, es importante destacar que el coeficiente de la demanda es positivo y significativo, es decir, mientras mayor sea la demanda del resto de los clientes en un periodo, mayor es la probabilidad de que el cliente realice un pedido en ese periodo.

Tabla B.1: Resultados para modelos de elección de compra con datos del método Naïve

Variable	Homogéneo		Homogéneo híbrido		Heterogéneo		Heterogéneo por ruta	
weight	0,0018		0,0004		0,0032		0,0048	
distance	-0,0062		-		-0,0464 **		-	
order_frequency	1,7258 **		-		2,0124 **		-	
route_num	-0,2258 **		-		-0,2648 **		-	
domestic	-0,0240		-		-0,1917 **		-	
θ	-0,0069 **		-		-0,0049		-	
τ	-0,0132 **		-0,0060		-0,0144 **		-0,0102 *	
demand	0,8531 **		0,8819 **		0,9018 **		0,8948 **	
first_half_week	0,0262		0,0264		0,0215		0,0170	
θ_{dist}	-		-0,0528 **		-		-0,0809 **	
intercepto	-3,2023 **		-3,6289 **		-3,0307 **		-3,6150 **	
Probabilidad marginal	-42.231		-40.812		-40.575		-40.413	

** significativo al 5%, * significativo al 10 %

Por otra parte, en cuanto al ajuste de los modelos, el cual se puede ver en la última fila con la métrica de la probabilidad marginal de cada modelo, se tiene que el modelo con peor ajuste es el Homogéneo y el mejor es el Heterogéneo por ruta, y es por lo cual el análisis se centrará en este modelo, donde la varianza y desviación estándar de cada una de sus variables se encuentra en la Tabla B.2, en la cual se observa que a pesar de que θ_{dist} es negativo y significativo, su varianza y desviación estándar son significativamente altas, por lo que hay clientes para los cuales el tiempo de demoras y anticipos normalizado por la distancia tiene un efecto positivo, es decir, clientes para los cuales al aumentar la demora de un pedido se esta aumentando la probabilidad de que sigan pidiendo, sin embargo, para el cliente promedio el efecto sería el contrario.

Tabla B.2: Coeficientes de beta, varianza y desviación estándar para modelo Homogéneo por ruta con datos del método Naïve

Variable	Beta		Varianza		Desviación estándar	
weight	0,005		0,017		0,131	
τ	-0,01 *		0,008		0,092	
demand	0,895 **		0,025		0,159	
first_half_week	0,017		0,169		0,411	
θ_{dist}	-0,081 **		0,097		0,312	
intercepto	-3,615 **		0,272		0,521	

B.1.2. Aprendizaje independiente

Para el aprendizaje independiente además de tener los datos de pronóstico tanto de la calidad de servicio sobre los tiempos de anticipo y demora, como también del tiempo total de

despacho del pedido, se tiene la varianza de estos pronósticos dentro de cada ruta, las cuales son incluidas en el modelo de elección de compra, junto al resto de las variables independientes. Los coeficientes del estimador de cada una de las variables independientes se encuentran en la Tabla B.3, donde al igual que para los datos del método Naïve se omiten variables para el modelo Homogéneo híbrido y el Heterogéneo por ruta y se incluye θ normalizado por distancia en lugar de θ por si solo, ya que al comparar los modelos con y sin esta variable, el que presentó mejor ajuste para ambos casos fue el que incluía a θ normalizado.

De los coeficientes de las variables independientes se puede observar que τ es negativa y significativa para los modelos homogéneos presentados en la tabla, lo cual quiere decir que cuando se considera que los clientes tienen las mismas preferencias de servicio es posible ver un efecto negativo de los tiempos de despacho en la probabilidad de compra del cliente, es decir, que al aumentar los tiempos de despacho disminuye la probabilidad de que los clientes prefieran la empresa, sin embargo, no es posible realizar la misma afirmación cuando se considera que los clientes son heterogéneos en sus preferencias, ya que si bien el coeficiente que acompaña a τ es negativo, la variable no es significativa. Por otra parte el coeficiente de σ es positivo y significativo para los modelos homogéneos, así mientras mayor sea la varianza del tiempo de despacho para una ruta, mayor será la probabilidad de los clientes de seguir realizando pedidos, sin embargo, esto no hace sentido con una mejor calidad de servicio, y como se ve en los modelos heterogéneos esta variable deja de ser significativa, por lo que se atribuye su significancia en los modelos homogéneos a la falta de flexibilidad para las preferencias de los clientes.

Además, se tiene que tanto θ como θ_{dist} son negativas y significativas al 5% para la mayoría de los modelos, a excepción del modelo Heterogéneo, donde no es significativa, por lo cual a mayores tiempo de atraso en los pedidos, menor será la probabilidad de que el cliente siga realizando pedidos y lo mismo ocurre con σ_θ , que es negativa y significativa para los 4 modelos, de esta forma mientras mayor sea la varianza del tiempo de demora o anticipo dentro de la ruta, menor será la probabilidad de los clientes de seguir comprando a través de la empresa, lo cual quiere decir que los clientes prefieren que sus tiempos de atraso o anticipo en los pedidos se mantengan en un rango pequeño y no varíen mucho entre sí. Además al igual que en lo modelos con los datos del método Naïve el coeficiente de la demanda es positivo y significativo, es decir, mientras mayor sea la demanda del resto de los clientes en un periodo, mayor es la probabilidad de que el cliente realice un pedido en ese periodo.

Tabla B.3: Resultados para modelos de elección de compra con datos del aprendizaje independiente

Variable	Homogéneo		Homogéneo híbrido		Heterogéneo		Heterogéneo por ruta	
weight	0,0022		0,0006		0,0041		0,0048	
distance	-0,0074	**	-		-0,0488	**	-	
order_frequency	1,7275	**	-		2,1536	**	-	
route_num	-0,2312	**	-		-0,4194	**	-	
domestic	-0,0341		-		-0,1440		-	
θ	-0,0131	**	-		-0,0009		-	
σ_θ	-0,0524	**	-0,0325	**	-0,0988	**	-0,1052	**
τ	-0,0234	**	-0,0235	**	-0,0177		-0,0077	
σ	0,0025	**	0,0023	**	0,0009		0,0019	
demand	0,8621	**	0,8913	**	0,9394	**	0,9394	**
first_half_week	0,0244		0,0268		0,0281		0,0269	
θ_{dist}	-		-0,1170	**	-		-0,4312	**
intercepto	-3,1505	**	-3,5989	**	-2,6479	**	-3,6414	**
Probabilidad marginal	-42.218		-40.810		-40.377		-40.244	

Luego, en cuanto al ajuste de los modelos, el cual se puede ver en la última fila con la métrica de la probabilidad marginal de cada modelo, se tiene que el modelo con peor ajuste es el Homogéneo y el mejor es el Heterogéneo por ruta, y es por lo cual el análisis se centrará en este modelo, donde la varianza y desviación estándar de cada una de sus variables se encuentra en la Tabla B.4, en la cual se observa que θ_{dist} es la variable con la varianza más alta, es decir, tiene la mayor heterogeneidad entre clientes, por lo cual a pesar de que su coeficiente es negativo y significativo para el promedio de los clientes, aún existe un número de clientes para los cuales el efecto de las demoras aumenta la probabilidad de realizar un envío, y por otro lado la demanda y el intercepto son las variables con menor heterogeneidad entre clientes, por lo cual el coeficiente del cliente promedio es una buena representación para las preferencias de los clientes respecto de estas variables.

Tabla B.4: Coeficientes de beta, varianza y desviación estándar para modelo Homogéneo por ruta con datos del aprendizaje independiente

Variable	Beta		Varianza		Desviación estándar	
weight	0,0048		0,0179		0,1338	
σ_θ	-0,1052	**	0,0585		0,2418	
τ	-0,0077		0,0249		0,1576	
σ	0,0019		0,0036		0,0602	
demand	0,9394	**	0,0386		0,1965	
first_half_week	0,0269		0,2216		0,4707	
θ_{dist}	-0,4312	**	1,9265		1,3880	
intercepto	-3,6414	**	0,2731		0,5226	

B.1.3. Aprendizaje jerárquico

Para el aprendizaje jerárquico se tienen los datos de la varianza de los pronósticos tanto dentro de cada ruta, como entre rutas, las cuales son añadidas al modelo de elección de compra junto al resto de las variables independientes. Los coeficientes del estimador de cada una de las variables independientes se encuentran en la Tabla B.5, donde al igual que para los datos del método Naïve y el aprendizaje independiente se omiten variables para el modelo Homogéneo híbrido y el Heterogéneo por ruta y se incluye θ normalizado por distancia en lugar de θ por si solo, ya que al comparar los modelos con y sin esta variable, el que presentó mejor ajuste para ambos casos fue el que incluía a θ normalizado.

De los coeficientes de las variables independientes se puede observar que τ es negativa para los 4 modelos presentados en la tabla, pero solo es significativa para el modelo Homogéneo, por lo que al considerar que todos los clientes tienen las mismas preferencias se podría decir que un aumento en los tiempos de despacho disminuye la probabilidad de estos de seguir pidiendo, sin embargo, al añadir flexibilidad al modelo y asumir que los clientes son diferentes entre sí, no se puede hablar de un efecto en la elección de compra de los clientes para la calidad de servicio medida por el tiempo de despacho de los pedidos. Por otra parte, y al igual que para los modelos realizados con los datos del método Naïve y el aprendizaje independiente, tanto θ como θ_{dist} son negativas y significativas al 5% para la mayoría de los modelos, a excepción del modelo Heterogéneo, donde no es significativa, lo cual significa que a mayores tiempo de atraso en los pedidos, menor será la probabilidad de que el cliente siga realizando pedidos y lo mismo ocurre con σ_θ , que es negativa y significativa para estos modelos, por lo cual a mayor varianza del tiempo de demora o anticipo dentro de la ruta, menor será la probabilidad de los clientes de seguir comprando a través de la empresa.

Por otra parte, al ver el modelo homogéneo y heterogéneo sin diferenciar el intercepto por ruta, se tiene que el coeficiente que acompaña a la variable de la distancia es negativo y significativo, por lo cual a mayor distancia del pedido menor será la probabilidad de que el cliente prefiera la empresa, así como también a mayor cantidad de rutas que utilice el cliente es menos probable que realice un envío. También se puede notar que mientras mayor sea la cantidad de pedidos realizados por el cliente en los primeros 25 periodos, mayor será la probabilidad de que el cliente vuelva a pedir.

Tabla B.5: Resultados para modelos de elección de compra con datos del aprendizaje jerárquico

Variable	Homogéneo		Homogéneo híbrido		Heterogéneo		Heterogéneo por ruta	
weight	0,0024		0,0006		0,0042		0,0040	
distance	-0,0068	*	-		-0,0470	**	-	
order_frequency	1,7238	**	-		2,2258	**	-	
route_num	-0,2487	**	-		-0,5205	**	-	
domestic	-0,0267		-		-0,1458		-	
θ	-0,0163	**	-		-0,0024		-	
σ_θ	-0,0417	**	-0,0489	**	-0,0597		-0,0835	**
ξ_θ	-0,0343		0,3556	**	-0,2135	**	-0,0305	
τ	-0,0258	**	-0,0244	**	-0,0375		-0,0222	
σ	0,0020	**	0,0010		0,0007		0,0017	
ξ	0,0014	*	-0,0014		-0,0297	**	-0,0299	**
demand	0,8634	**	0,8792	**	0,9416	**	0,9434	**
first_half_week	0,0262		0,0266		0,0295		0,0280	
θ	-		-0,1159	**	-		-0,5143	**
intercepto	-3,0500	**	-4,2135	**	-1,9065	**	-3,4862	**
Probabilidad marginal		-42.227		-40.808		-40.333		-40.191

Por último se puede ver que el intercepto es negativo y significativo para todos los modelos, es decir, los clientes tienen una predisposición negativa a realizar pedidos, es decir, que si todas las variables independientes se fijaran en 0, el cliente no realizaría un pedido en la empresa, por lo cual es necesario aumentar el valor de las variables positivas y significativas, y disminuir el de las negativas para suplir este efecto, sin embargo, para el modelo Heterogéneo por ruta, el cual es el modelo de interés al tener el mejor ajuste respecto al resto de los modelos, las únicas variables significativas son; la varianza del tiempo de demora o anticipo por ruta, la varianza del tiempo de despacho entre rutas, la demanda y el pronóstico del tiempo de demora o anticipo normalizado por la distancia, donde no es posible aumentar la demanda y la varianza del tiempo de despacho entre rutas esta sujeta a la heterogeneidad de las rutas de cada cliente, por lo cual es necesario disminuir la varianza en los tiempos de demora o anticipo para cada ruta y los tiempos de atraso o anticipo en los pedidos, lo cual puede ser realizado al mejorar el pronóstico de la fecha y hora de llegada del envío.

Luego, se tiene la varianza y desviación estándar de cada una de las variables independientes del modelo Heterogéneo por ruta en la Tabla B.6, en la cual se observa que θ_{dist} es la variable con la varianza más alta al igual que para los datos del aprendizaje independiente, siendo aún mayor que para el caso anterior, ya que la desviación estándar es casi 4 veces el valor del coeficiente, es decir, tiene una gran heterogeneidad entre clientes, por lo cual a pesar de que su coeficiente es negativo y significativo para el promedio de los clientes, aún existe un número de clientes para los cuales el efecto de las demoras aumenta la probabilidad de realizar un envío. Por otra parte, la variable con menor heterogeneidad entre clientes es la demanda, lo cual tiene sentido ya que esta variable es la suma de los pedidos realizados por el resto de clientes para un periodo, por lo que entre clientes la demanda no tiene una diferencia significativa.

Tabla B.6: Coeficientes de beta, varianza y desviación estándar para modelo Homogéneo por ruta con datos del aprendizaje jerárquico

Variable	Beta		Varianza	Desviación estándar
weight	0,0040		0,0177	0,1331
σ_θ	-0,0835	**	0,0638	0,2526
ξ_θ	-0,0305		0,0705	0,2655
τ	-0,0222		0,0309	0,1757
σ	0,0017		0,0038	0,0616
ξ	-0,0299	**	0,0215	0,1468
demand	0,9434	**	0,0400	0,1999
first_half_week	0,0280		0,2271	0,4765
θ_{dist}	-0,5143	**	3,2593	1,8053
intercepto	-3,4862	**	0,2827	0,5317

B.2. Pronósticos del tiempo de demora o anticipo y el tiempo total de envío de los diferentes modelos para un cliente

Periodo	Ruta	Pronóstico del tiempo de demora o anticipo				Pronóstico del tiempo total de envío			
		Real	Jerárquico	Independiente	Naïve	Real	Jerárquico	Independiente	Naïve
7,00	2,00	-2,80	0,00	0,00	0,00	132,06	0,00	0,00	0,00
9,00	1,00	0,67	-2,66	-0,03	0,00	79,21	131,95	1,65	0,00
10,00	1,00	1,21	-0,55	0,68	0,67	128,20	98,67	79,20	79,21
12,00	1,00	1,17	0,23	0,96	1,21	149,71	112,49	103,40	128,20
14,00	2,00	-0,26	-1,39	-2,77	-2,80	85,84	122,45	130,30	132,06
15,00	1,00	-2,96	0,61	1,04	1,17	94,11	114,93	118,70	149,71
18,00	3,00	-1,95	-0,62	-0,06	0,00	114,45	111,45	-4,03	0,00
18,00	1,00	0,21	-0,20	0,00	-2,96	144,24	111,53	112,28	94,11
22,00	2,00	0,79	-1,19	-1,53	-0,26	158,95	115,96	108,36	85,84
24,00	3,00	-1,09	-1,12	-1,94	-1,95	99,69	120,77	113,38	114,45
26,00	3,00	-2,50	-1,10	-1,51	-1,09	164,51	117,99	106,71	99,69
28,00	3,00	-0,94	-1,45	-1,86	-2,50	90,74	123,20	125,74	164,51
28,00	3,00	-0,84	-1,45	-1,86	-2,50	122,62	123,20	125,74	164,51
32,00	2,00	-0,13	-0,76	-0,75	0,79	113,21	120,76	125,33	158,95
34,00	3,00	-1,19	-1,25	-1,47	-0,84	102,83	119,87	118,21	122,62
37,00	3,00	0,04	-1,26	-1,41	-1,19	203,60	118,65	115,49	102,83
40,00	2,00	0,72	-0,61	-0,60	-0,13	109,11	124,00	121,84	113,21
42,00	3,00	-1,06	-1,09	-1,21	0,04	96,48	123,37	128,31	203,60
44,00	1,00	-0,83	-0,07	0,06	0,21	110,51	121,23	118,78	144,24
48,00	3,00	-0,96	-1,09	-1,18	-1,06	138,05	121,00	124,21	96,48
48,00	1,00	0,23	-0,17	-0,09	-0,83	122,89	120,56	117,54	110,51
50,00	3,00	-1,01	-1,08	-1,17	-0,96	97,98	122,28	125,81	138,05
52,00	1,00	-1,67	-0,11	-0,04	0,23	96,20	120,57	118,28	122,89
52,00	3,00	-0,67	-1,07	-1,14	-1,01	186,33	120,93	123,09	97,98
54,00	3,00	-1,04	-1,05	-1,11	-0,67	95,70	122,74	128,75	186,33
56,00	3,00	-1,67	-1,04	-1,10	-1,04	161,34	121,63	125,95	95,70
58,00	1,00	-0,17	-0,29	-0,25	-1,67	92,17	122,31	115,41	96,20
58,00	3,00	-0,90	-1,10	-1,14	-1,67	120,71	123,26	128,72	161,34
62,00	3,00	-1,04	-1,08	-1,12	-0,90	118,34	122,03	127,98	120,71
65,00	1,00	-2,92	-0,27	-0,23	-0,17	53,19	120,95	112,98	92,17
68,00	3,00	0,00	-1,09	-1,12	-1,04	95,76	120,07	127,35	118,34
72,00	3,00	-1,12	-1,02	-1,05	0,00	116,22	119,28	125,67	95,76
72,00	2,00	0,88	-0,40	-0,33	0,72	113,05	118,57	119,29	109,11
74,00	3,00	-0,25	-1,03	-1,05	-1,12	135,92	119,00	124,83	116,22
81,00	2,00	2,87	-0,20	-0,13	0,88	211,42	118,51	118,57	113,05
82,00	3,00	-2,03	-0,98	-1,01	-0,25	136,98	121,85	125,60	135,92
82,00	3,00	-0,30	-0,98	-1,01	-0,25	107,11	121,85	125,60	135,92
84,00	2,00	0,92	0,20	0,30	2,87	115,46	121,91	131,58	211,42
88,00	2,00	-0,04	0,29	0,38	0,92	115,41	121,78	129,66	115,46
89,00	3,00	0,00	-0,99	-1,02	-0,30	178,44	121,62	125,20	107,11
90,00	1,00	-0,94	-0,50	-0,50	-2,92	94,71	121,21	106,95	53,19
92,00	1,00	-0,96	-0,54	-0,54	-0,94	71,39	120,46	105,91	94,71
92,00	1,00	0,92	-0,54	-0,54	-0,94	158,72	120,46	105,91	94,71
96,00	3,00	-0,82	-0,95	-0,97	0,00	73,85	122,38	127,61	178,44
96,00	3,00	1,87	-0,95	-0,97	0,00	187,99	122,38	127,61	178,44
97,00	3,00	-4,08	-0,82	-0,85	1,87	42,27	122,69	128,05	187,99
98,00	3,00	-0,79	-0,95	-0,98	-4,08	85,06	120,55	124,37	42,27
102,00	3,00	-0,92	-0,95	-0,97	-0,79	117,48	119,88	122,75	85,06