

# Estimación de la migración de usuarios de transporte público utilizando datos pasivos

## ESTIMATION OF THE MIGRATION OF PUBLIC TRANSPORT USERS WITH PASIVE DATA

Camilo Leng, Marcela A. Munizaga\*

División Ingeniería de Transporte, Departamento de Ingeniería Civil, Universidad de Chile

\* Autor para correspondencia:  
mamuniza@ing.uchile.cl

### RESUMEN

Esta investigación usa un Modelo Logit Ordinal para modelar la frecuencia de uso o posible abandono de usuarios del transporte público. Los datos disponibles son generados de forma pasiva a partir del mismo uso del sistema. La frecuencia de uso se mide según una clusterización de usuarios en tres categorías y el abandono cuando una tarjeta observada en un periodo anterior está ausente en un segundo periodo. Esta investigación utiliza técnicas de imputación de información para caracterizar de mejor forma la muestra de tarjetas. Los resultados muestran que los usuarios con tendencia a hacer un uso más frecuente del sistema son aquellos que combinan todos los modos (bus, metro, zonas paga, servicios expresos) y aquellos que viajan en horarios punta durante días laborales. La segmentación por ingreso según comuna de residencia permitió detectar algunos efectos para segmentos específicos, como la velocidad de viaje para el quintil de ingreso más alto.

Palabras clave: modelo logit ordinal, migración, datos de tarjetas inteligentes.

### ABSTRACT

*This investigation uses an Ordinal Logit Model to model the frequency of use or possible migration from users of a public transport system. Passive data collected from the system itself is available for the model. The frequency of use is measured using a clusterization of the users in three groups, the user migration happens when a smartcard used in a previous period is absent in a subsequent period. In this investigation datamining technics are used to input socioeconomic information to the users. The results show that users who combine modes and services (bus, metro, express services, off-bus payment), and those who travel during peak hours in working days tend to be more frequent/loyal. The income segmentation by zone of residence allowed to find some specific effects, such as the relevance of average speed of the trip for users of the higher income quintile.*

*Keywords: ordinal logistic model, migration, smart card data.*

## 1. INTRODUCCIÓN

Santiago, la capital de Chile, posee alrededor de 6 millones de habitantes, los que en un día laboral realizan cerca de 18 millones de viajes para llevar a cabo sus labores y satisfacer sus necesidades. La movilidad urbana en una ciudad de estas características no es un tema sencillo y en Santiago existe una tendencia preocupante. Las últimas dos encuestas Origen-Destino de Santiago han dejado en evidencia que existe una constante transferencia de usuarios desde el transporte público al

privado. La partición modal del transporte público ha disminuido de un 30,1% en 2001 a un 26,1% en 2012. Mientras tanto la del transporte privado ha aumentado de un 21,0% a un 26,1% en los mismos años (Muñoz et al., 2016). De continuar así, el sistema de transporte público se verá gravemente deteriorado a causa de la falta de demanda y la movilidad urbana en la ciudad será cada vez más caótica debido al aumento en el número de autos.

En Santiago constantemente se llevan a cabo proyectos con el fin de mejorar la calidad de servicio del transporte público, sin embargo, la tendencia a la baja de la partición modal sigue igual. El problema surge de la forma en que se invierten los recursos. Una inversión de recursos global en el sistema genera mejoras pequeñas en la calidad de servicio de cada individuo. Si se conocieran los usuarios con una predisposición a abandonar el sistema, la inversión podría realizarse de forma focalizada. Así, se mejora el servicio de las personas más próximas a abandonar el sistema y se atiende el problema de la disminución de la demanda. El objetivo de esta investigación es identificar los factores que afectan la decisión de los usuarios de disminuir su uso del sistema o abandonarlo.

El sistema de transporte público de Santiago, Transantiago, es un sistema integrado de buses y metro. La flota de buses es proporcionada por siete operadores particulares que en total cuentan con sobre 6.600 buses incorporados con tecnología GPS. Estos sirven alrededor de 378 servicios troncales y alimentadores. En términos de infraestructura, el sistema cuenta con 72 km de vías segregadas, 31 km de vías exclusivas, 200 km de pistas solo bus, más de 11.000 paraderos y 125 estaciones con sistema de pre-pago. El metro está compuesto por 108 estaciones distribuidas en cinco líneas. El sistema cuenta con 186 trenes que cubren una red de 104 km. El único método de pago admitido en buses es la tarjeta inteligente “bip!”, lo que ha permitido que su tasa de penetración sea de un 97% (DTPM, 2017).

En la literatura se pueden encontrar muchas investigaciones atendiendo el problema de cómo atraer nuevos usuarios al sistema. No obstante, la búsqueda de medidas para mantener a los usuarios en el sistema no es un tema tan considerado. Los trabajos encontrados suelen confirmar que tanto las variables de calidad de servicio como el tiempo de viaje, el tiempo de espera, el tiempo de caminata y la cantidad de transbordos tienen un impacto significativo al verse empeoradas. También, que en comunas de altos ingresos el abandono es mayor. Por último, que realizar cambios en la tarifa posee un impacto más moderado en la decisión de migrar de los usuarios que los cambios en las variables de calidad de servicio (Imaz et al., 2015; Bass et al., 2011; Munizaga y Calvo, 2015).

Previas investigaciones con los datos pasivos del transporte público de Santiago han permitido generar las bases para esta investigación. La metodología descrita por Munizaga y Palma (2012) estima la parada de bajada de sobre un 80% de las validaciones minimizando una función del tiempo generalizado de viaje que considera el tiempo de viaje en el vehículo y el de caminata al siguiente paradero donde ocurra una validación. Además, Devillaine et al. (2012) utiliza reglas simples para determinar la ubicación de las actividades de los viajes diferenciando paradas de transferencia de paradas de actividades, permitiendo así construir la estructura de viajes de cada usuario con sus respectivas etapas en cada viaje. Como la metodología no es capaz de estimar la bajada de todos los viajes es necesario incorporar factores de expansión para poder construir una matriz Origen-Destino compatible con los viajes en transporte público reportados por la Encuesta Origen-Destino 2001 de Santiago. Munizaga et al. (2014) continua el trabajo validando la matriz Origen-Destino con tres fuentes: los mismos datos pasivos de forma endógena, una encuesta de viaje realizada por Metro de Santiago y un pequeño grupo de voluntarios que accedieron a entregar su información de viajes. Esta investigación aprovechó de mejorar la metodología y corregir errores. Las mejoras más relevantes incluyen la introducción de flexibilidad en la distancia de caminata máxima, la redefinición del corte entre un día y el siguiente según el momento de menor actividad en el sistema, la posibilidad de estimar la bajada de viajes únicos en un día a partir de la información del día siguiente y nuevos criterios para identificar paradas de transferencia.

Las tarjetas utilizadas en el transporte público de Santiago son anónimas. Esto impide conocer información demográfica o socio económica sobre sus usuarios. Con el fin de resolver esta falta de información Amaya y Munizaga (2017) proponen una metodología capaz de estimar la zona de residencia de los usuarios. Esta investigación se basa en el hecho de que los usuarios suelen pasar las noches en sus hogares, por lo tanto, la primera validación en la mañana debe ser cerca de su zona de residencia. La metodología trabaja sólo con usuarios frecuentes del sistema, es decir, aquellos que hayan viajado cuatro o más días a la semana. La estimación se hace observando la primera validación de cada día laboral que ocurra entre las 4:00 AM y las 12:00 PM. Si estas validaciones se encuentran dentro de un radio razonable para caminar, se asume que el usuario vive en esa zona. De esta forma, el propósito de esta investigación es identificar los factores que afectan la decisión de los usuarios de disminuir su uso del sistema o abandonarlo, utilizando principalmente los datos pre procesados del sistema de transporte público de Santiago.

## 2. DATOS

### 2.1 Descripción de datos

La principal fuente de información de esta investigación son los datos generados de forma pasiva por el Transantiago que luego son procesados según las metodologías descritas por distintos autores para enriquecer su contenido. El resultado de este proceso son dos bases de datos agregadas por etapas y viajes con los principales campos necesarios para describir el uso de un sistema de transporte público:

- Identificador de tarjeta
- Tipo de tarjeta
- Fecha y hora del origen y destino
- Coordenadas X e Y del origen y destino
- Paradero asignado al origen y destino
- Servicio utilizado
- Dirección del servicio
- Estimación de tiempos de viaje, espera y caminata

Donde todas las variables asociadas al destino de un viaje/etapa están sujetas a su correcta estimación. Los tipos de tarjetas son básicamente las tarjetas de adulto regular y las tarjetas de estudiante de educación básica, media y superior. Además, varias otras variables derivadas de estas son incluidas en la base de datos como el modo de transporte, el día de la semana, la comuna de origen y destino, el periodo del día, la distancia euclidiana y en ruta entre el origen y destino, etc. La estimación de residencia es imputada en los casos donde las tarjetas son frecuentes.

Para analizar la migración de usuarios es necesario observar dos cortes temporales. Esta investigación observa la experiencia de viaje de los usuarios en un primer periodo y luego analiza la frecuencia de uso y posible migración del sistema en un periodo posterior. El primer periodo comienza el 18 de abril de 2015 y termina el 25 del mismo mes y año. Debido al tamaño y tiempo de procesamiento necesario para confeccionar las bases de datos no es posible trabajar con un periodo más extenso. El segundo periodo es el mes de noviembre de 2015. En este caso es posible trabajar con un periodo más largo, ya que para observar la frecuencia de uso de las tarjetas no se requiere que los datos sean procesados. Los periodos fueron cuidadosamente escogidos tal que el comportamiento de los usuarios en la ciudad sea lo más regular posible y que en ningún caso sean periodos de vacaciones. Los periodos se encuentran separados por una diferencia de seis meses y dentro del mismo año escolar con el objetivo de minimizar la cantidad de usuarios que cambie de hogares y/o de lugares de trabajo o estudio, dado que de existir cambios de esta índole los modos alternativos de viaje cambian la utilidad que le reportan a los usuarios y la migración puede deberse a este factor.

### 2.2. Limpieza de datos

Como la mayoría de las bases de datos, esta no está exenta de errores y falta de información. Las causas más comunes de estos errores son el mal funcionamiento de los instrumentos utilizados como los pulsos GPS o los validadores de tarjetas y los errores humanos como la mal asignación de un servicio a su validador. Además, existen comportamientos especiales de los usuarios que las metodologías no son capaces de explicar, como el caso de una persona validando dos veces en un mismo validador con la misma tarjeta, con el fin de pagar por el viaje de dos personas. Debido a todas estas fuentes de errores es que es importante realizar una limpieza de los datos.

En la base de datos existen tarjetas con una cantidad de viajes y etapas sobre lo esperado. Se cree que estas tarjetas son utilizadas con propósitos especiales en el transporte público y no representan el comportamiento regular de ningún usuario. Para estos casos se creó una metodología de detección de observaciones atípicas. En un principio, no basta con contabilizar los viajes de cada tarjeta y eliminar aquellas consideradas outliers por algún criterio, ya que es necesario considerar el tiempo

que le tomó al usuario realizar los viajes. Por lo tanto, las tarjetas son agrupadas por la cantidad de días viajados y luego al interior de cada grupo se eliminan aquellas tarjetas que posean más viajes/etapas que la media más tres veces la desviación estándar del conjunto. La Tabla 1 muestra la cantidad máxima de viajes y etapas permitida por días viajados.

Tabla 1  
*Máxima cantidad de viajes y etapas por días viajados*

Días viajados	1	2	3	4	5	6	7	8
Viajes máximos	4,5	7,0	10,5	13,0	18,3	21,4	24,6	26,5
Etapas máximas	5,4	10,4	16,2	20,7	30,3	39,2	46,4	50,2

En el caso de las transacciones en buses las coordenadas de estas son estimadas según la interpolación de la ubicación y la hora de los dos pulsos GPS más cercanos a la hora de la transacción. Cuando los pulsos GPS entregan información errónea, la ubicación de las transacciones no es correctamente estimada y tampoco su asignación a un paradero. En los datos se han observado casos donde las coordenadas de las transacciones ocurren fuera de los límites de la Región Metropolitana. Para evitar estos casos se ha delimitado un rectángulo que contiene a todos los paraderos establecidos por Transantiago en su interior con un margen de error de 100 metros. Cualquier validación con coordenadas fuera de este rectángulo es eliminada de la base de datos.

Los tiempos de viaje, espera y caminata estimados por las metodologías previamente mencionadas también pueden contener errores. Estos pueden ocurrir por errores en la estimación del paradero de bajada o por comportamientos no esperados como correr o usar modos de transporte alternativos tales como la bicicleta durante una transferencia. Cuando esto ocurre, los tiempos de espera compensan a los sobrestimados tiempos de caminata, pudiendo llegar a tener valores negativos. Las etapas con tiempos de espera negativos fueron reemplazadas por “sin observación” para no perder información relevante sobre el resto del viaje. Además, existe un 20% de las etapas que no poseen estimación de bajada. En estos casos no es posible calcular los tiempos de viaje, espera y caminata. Para atender estos problemas, después de agrupar la base de datos por tarjeta (para su posterior modelación), aquellas que tuvieran algún campo sin información son eliminadas del análisis.

### 2.3 Clasificación de usuarios

Basándose en la clasificación de usuarios realizada por Langlois et al. (2016), la muestra de tarjetas es segmentada en tres categorías. Utilizando el algoritmo de clusterización k-medias las tarjetas son agrupadas según sus días viajados y la cantidad de transacciones realizadas. De esta forma, las tarjetas pueden ser clasificadas en: Frecuente, Semi Frecuente y No Frecuente. En la Figura 1 se presentan los resultados de la clasificación de usuarios para el mes de noviembre del 2015. El tamaño de los círculos muestra la cantidad de tarjetas que tienen la misma cantidad de transacciones y días viajados, el color la pertenencia a la tres categorías y las cruces son los centroides de cada conjunto. Al observar la cantidad de tarjetas por grupo y el total de transacciones que realizan, se concluye que las tarjetas frecuentes pese a ser solo un 13,9% del total, realizan un 48,6% de las transacciones. En cambio, las tarjetas no frecuentes son un 63,7% y solo realizan un 18,8% de las transacciones. Estos resultados enseñan que existe un conjunto de usuarios del cual depende el Transantiago, y sus recursos deberían enfocarse en disminuir la migración de estos. Por último, las tarjetas semi frecuentes representa un 22,4% de la muestra y son responsables de un 32,5% de las transacciones.

### 2.4 Encuesta CASEN 2015

La Encuesta de Caracterización Socioeconómica Nacional (CASEN) 2015, realizada por el Ministerio de Desarrollo Social, contiene un conjunto de atributos capaces de explicar la situación de los hogares y la población a lo largo de todo el país. El tamaño muestral de la encuesta es de 83.887 hogares, con un total de 266.968 personas encuestadas. Del conjunto de variables disponibles en la encuesta, esta investigación utiliza el ingreso, la cantidad de vehículos y el grado de pobreza por hogar para calcular el ingreso medio, la tasa de motorización y el porcentaje de pobreza de las comunas donde opera el

Transantiago. Estos atributos son posteriormente imputados a las tarjetas según su estimación de residencia. Dada la fuerte desigualdad de ingreso por comuna, observable en la Figura 2, las comunas son agrupadas por los quintiles de ingreso. El propósito de esta segmentación es poder identificar comportamientos propios de cada estrato socio económico.

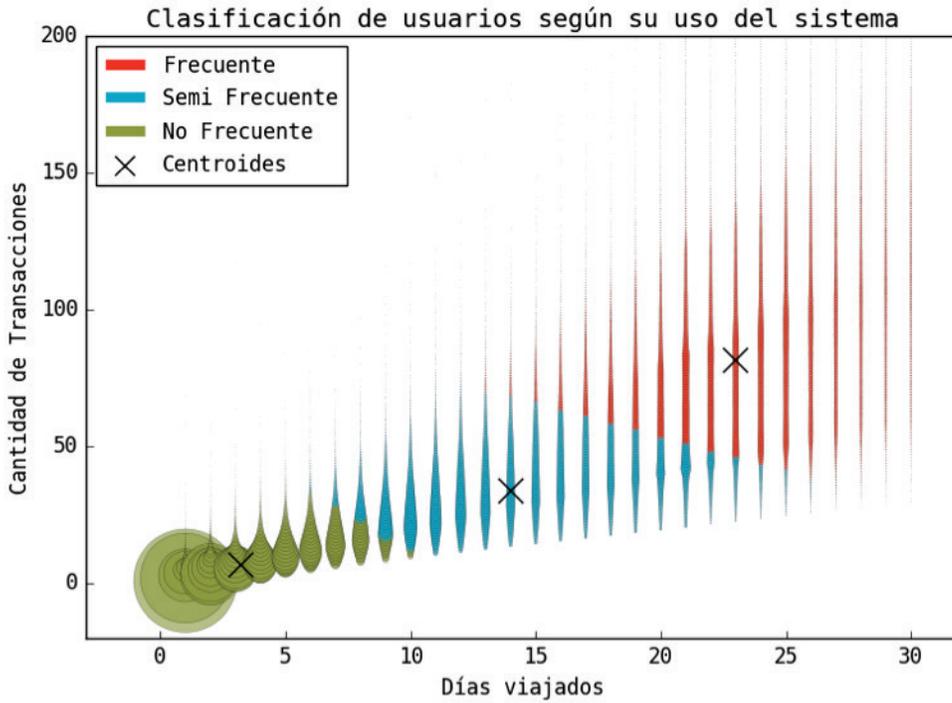


Figura 1. Clasificación de usuarios según su grado de uso del sistema

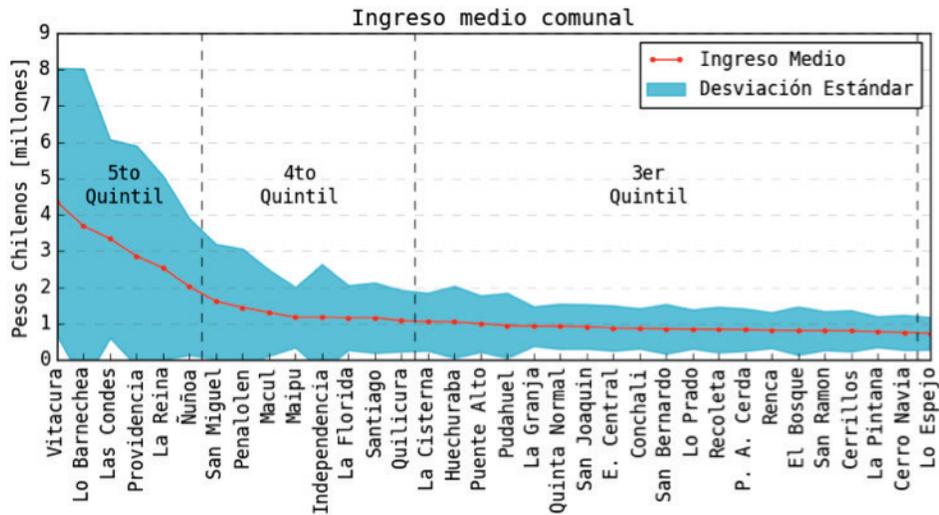


Figura 2. Ingreso medio y desviación estándar por comunas en Santiago

### 3. MODELO

Para plantear el modelo, los usuarios son clasificados en cuatro categorías, las primeras tres resultantes de la clasificación de usuarios según su grado de uso del sistema (No Frecuente, Semi Frecuente, Frecuente) y la última (Ausente) ocurre cuando la tarjeta no está presente en el segundo periodo del análisis. Esta definición permite determinar qué factores hacen que los usuarios disminuyan su uso del sistema sin necesariamente haber migrado o cuales son las variables que los usuarios frecuentes más presentan. El modelo se plantea como un modelo con rezago, donde las variables explicativas se observan en un periodo de una semana, y la variable explicada (frecuencia de uso) en un periodo posterior de un mes. Dado que la variable a explicar no es binaria, sino que posee cuatro categorías con un orden natural, se utiliza un modelo logit ordinal ya que este es el que se adecua mejor a la naturaleza de la variable dependiente. A diferencia del modelo logit multinomial, el modelo logit ordinal no calcula las probabilidades de cada evento por separado, sino que considera la probabilidad de ocurrencia de un evento y de todos los ordenados anteriormente. De esta forma, el modelo logit ordinal se expresa como en la Ecuación 1.:

$$\ln(\theta_j) = \alpha_j - \beta X_i \quad (1)$$

Donde  $j$  toma los valores de 1 hasta el número de categorías menos 1 e  $i$  va desde 1 hasta el número de variables. El vector  $\beta$  posee los parámetros asociados a la utilidad marginal de cada variable y es independiente de las categorías. El signo negativo frente al vector de parámetros, permite una interpretación intuitiva de los resultados. Un parámetro de valor positivo, implica una utilidad marginal positiva, por lo cual un aumento en la variable asociada produce una disminución de los puntos de corte  $\theta_j$  disminuyendo la probabilidad de caer en las categorías iniciales (migrar o no frecuente) y aumentando la probabilidad de caer en las últimas (semi frecuente o frecuente).

En resumen, las variables asociadas a coeficiente positivos aumentan la probabilidad de ser catalogado como frecuente, y las variables asociadas a coeficientes negativos explican la migración y la disminución del uso del sistema por parte de los usuarios.

El modelo asume que cada tarjeta posee un dueño único y que cada usuario utiliza solo una tarjeta. Otro supuesto es que la ausencia de una tarjeta en el segundo periodo es a causa de la migración y no de otros motivos como pueden ser enfermedades, accidentes o viajes de trabajo. Por otra parte, se asume que los modos alternativos no sufren grandes cambios entre los dos periodos analizados, por lo tanto, la decisión de migrar está fundamentada solamente en la experiencia de viaje en el transporte público.

#### Variables explicativas

Para la construcción de las variables explicativas del modelo, se han agregados las bases de datos de etapas y viajes por tarjetas. De esta forma se obtiene una observación única por usuario que explica la experiencia que percibe este del sistema en el primer periodo. En este proceso se han calculado nuevas variables asociadas a los modos de transporte utilizados, a la calidad de servicio experimentada, a los atributos socio económicos de las comunas y a los patrones temporales de los usuarios. A continuación, se presentan las variables asociadas a la calidad de servicio.

- Variabilidad en el tiempo de viaje [min]
- Tiempo de viaje corto
- Tiempo de viaje largo
- Tiempo de espera [min]
- Tiempo de caminata [min]
- Velocidad [m/s]

Las variables tiempo de viaje corto y largo son variables binarias que toman el valor 1 cuando el tiempo medio de viaje de una tarjeta es menor al percentil 25 o mayor al percentil 75 del tiempo medio de viaje de toda la muestra respectivamente. Los tiempos de espera y caminata, en minutos, son calculados directamente como el promedio de estos tiempos en los viajes de

cada tarjeta. De la misma forma, la velocidad de viaje, medida en metros por segundo, es el promedio de los cocientes entre la distancia euclidiana y el tiempo total de los viajes de cada tarjeta. La variabilidad en el tiempo de viaje, en minutos, es la desviación estándar de los tiempos de viajes de cada tarjeta.

Observando los modos utilizados por cada tarjeta y los patrones temporales de sus validaciones se crean variables que revelen sus preferencias y comportamientos. Las variables y su descripción se muestran en la Tabla 2.

Tabla 2

*Variables creadas según los modos de transporte y patrones temporales utilizados.*

Variable	Descripción
Usa bus	Usuario que realizó al menos un viaje en bus
Usa metro	Usuario que realizó al menos un viaje en metro
Usa buses expresos	Usuario que realizó al menos un viaje en servicios de buses expresos
Usa zona paga	Usuario que realizó al menos un viaje desde un paradero con sistema de pre pago
Solo buses	Usuario que realizó todos sus viajes en bus
% TV en metro	Porcentaje del tiempo de viaje total realizado en el modo metro
Solo horarios punta	Usuario que viaja solo en periodos punta (mañana o tarde)
Solo días laborales	Usuario que viaja solo entre lunes y viernes

Por último, a las tarjetas se les imputa la tasa de motorización y porcentaje de pobreza comunal según su estimación de residencia. El ingreso es utilizado solamente para segmentar a la muestra en tres conjuntos. Existen dos variables binarias que indican la clasificación de los usuarios en el primer periodo.

#### 4. RESULTADOS

Santiago es una ciudad con una distribución del ingreso muy dispereja; esto genera zonas en la ciudad con altos estándares de vida segregados del resto. Se cree que cada uno de estos estratos socio económicos posee distintas preferencias con respecto al transporte público. Para incluir este fenómeno en la modelación, la base de datos ha sido dividida en tres conjuntos según el ingreso medio comunal imputado a cada usuario por su estimación de residencia. En la Figura 2 se aprecia el ingreso medio comunal por hogar, la variabilidad de este al interior de cada comuna y la distribución de las comunas por quintiles. Los quintiles fueron calculados considerando los hogares de toda la Región Metropolitana. Cabe destacar que no existe ninguna comuna en el primer quintil y que solo la comuna de Lo Espejo pertenece al segundo quintil. Este último grupo no será modelado debido a la falta de variabilidad de sus observaciones. Resultando en tres modelos para los tres quintiles más altos.

En la Tabla 3 se presentan los coeficientes y t-valores de los tres modelos considerando una significancia del 95%. Lo primero que se puede observar es que en los tres modelos los coeficientes asociados al tiempo de viaje corto son negativos y que los coeficientes asociados al tiempo de viaje largo son positivos. Esto se explica por la fuerte correlación que existe entre la distancia y el tiempo de un viaje. Así, lo que estas variables buscan explicar es que los viajes cortos en transporte público pueden ser reemplazados por otros modos como la caminata, la bicicleta o el taxi fácilmente sin incurrir en mayores costos, aumentando la probabilidad de migrar de estos usuarios. En cambio, como los viajes largos no pueden ser reemplazados por modos no-motorizados y reemplazarlos por un modo motorizado, como el auto privado o el taxi, significa un costo importante, estos usuarios tienden a ser extremadamente fieles y frecuentes.

El tiempo de espera y el tiempo de caminata resultaron significativos en los modelos del cuarto y tercer quintil respectivamente. Sus signos negativos dicen que a medida que aumenta el tiempo promedio que un usuario gasta esperando o caminando respectivamente durante sus viajes, mayor será la probabilidad de abandonar el sistema. Se destaca la significancia de la variabilidad en el tiempo de viaje en el modelo del cuarto quintil, este resultado puede deberse a los problemas de regularidad y frecuencia de los servicios de superficie en el Transantiago. Que la variable velocidad de viaje sea significativa para los usuarios de las comunas del quinto quintil se entiende como que estos consideran que el tiempo que le toma al transporte

público recorrer ciertas distancias es relevante a la hora de determinar su uso del sistema. Esto puede deberse a que estos usuarios disponen de más alternativas de transporte por su mayor ingreso.

Si se observan las variables que describen los modos de preferencia de los usuarios se deduce que los paraderos con sistema de pre-pago (zonas pagas) y los servicios de buses expresos son considerados como medidas positivas independiente del nivel de ingreso de los usuarios. Los coeficientes que acompañan a estas variables son positivos en los tres modelos, lo que significa que los usuarios que usan estos servicios tienen una mayor probabilidad de ser frecuentes. Al comparar el valor de los coeficientes de estas variables se concluye que los usuarios del tercer y cuarto quintil se ven más influenciados por los zonas pagas que por los servicios expresos. Lo contrario sucede con los usuarios del quinto quintil.

Tabla 3

*Variables creadas según los modos de transporte y patrones temporales utilizados.*

Modelo	III Quintil		IV Quintil		V Quintil	
	Valor	(t-valor)	Valor	(t-valor)	Valor	(t-valor)
Variabilidad en el TV	-	-	-0,006	(-8,6)	-	-
Tiempo de viaje corto	-0,262	(-14,8)	-0,307	(-16,9)	-0,272	(-14,2)
Tiempo de viaje largo	0,253	(21,0)	0,321	(22,3)	0,272	(9,1)
Tiempo de espera	-	-	-0,005	(-3,9)	-	-
Tiempo de caminata	-0,010					
(-2,3)	-	-	-	-		
Velocidad	-	-	-	-	0,003	(3,1)
Usa bus	0,255	(4,9)	0,243	(4,8)	0,188	(8,5)
Usa zonas pagas	0,146	(12,7)	0,128	(10,1)	0,090	
(3,7)						
Usa buses expresos	0,049	(3,8)	0,072	(5,2)	0,222	(7,3)
Solo buses	-0,061	(-3,5)	-0,076	(-3,9)	-0,180	
(-7,2)						
% del TV en metro	0,250					
(11,7)	0,068	(2,9)	-	-		
Solo días laborales	0,121	(8,4)	0,122	(7,6)	-	-
Solo horarios punta	0,190					
(7,5)	0,222	(7,7)	0,238	(6,3)		
Tasa motorización	-0,039	(-2,8)	-0,039	(-3,1)	-	-
% de pobreza	-	-	-1,529	(-4,2)	-1,684	(-2,4)
Semi frecuente	-0,752	(-54,1)	-0,753	(-48,0)	-0,773	(-42,7)
No frecuente	-1,757	(-23,5)	-1,502	(-9,3)	-1,521	(-14,7)
<i>Interceptos</i>						
Ausente   No frecuente	-10,477	(-13,6)	-11,037	(-14,5)	-11,036	(-11,5)
No Frec.   Semi Frec.	-0,89					
(-15,7)	-1,237	(-20,3)	-1,355	(-46,6)		
Semi Frec.   Frecuente	0,368	(6,5)	0,067	(1,1)	0,641	(22,5)
Desviación Residual	253847,59	210781,75	103818,47			
AIC	253879,59	210817,75	103846,47			

El porcentaje del tiempo de viaje en metro es significativo en los modelos del tercer y cuarto quintil. En ambos casos su signo es positivo. Esto significa que entre mayor es el tiempo que un usuario pasa viajando en metro, mayor es su probabilidad de ser frecuente. No obstante, la variable usa bus posee un signo positivo en los tres modelos. Así, los usuarios con una mayor probabilidad de ser frecuentes serán aquellos que utilizan en conjunto todos los modos de transporte pero principalmente el metro. Este resultado se ve confirmado por la variable solo bus que posee un coeficiente de signo negativo en los tres modelos. Es importante destacar que el coeficiente del porcentaje del tiempo de viaje en metro es similar en magnitud al de la variable usa bus en el tercer quintil pero no así en el cuarto quintil. Esto indica que los usuarios del cuarto quintil valoran menos este medio de transporte comparado con los del tercer quintil. En el caso del quinto quintil, el metro no es significativo. Esto puede deberse al hecho de que las comunas que pertenecen a este conjunto se encuentran más alejadas de la red de trenes.

Las variables temporales indican que los usuarios que utilizan el sistema solo en horarios punta tienen una mayor probabilidad de ser frecuentes. En cuanto a los que solo validan en días laborales, ocurre el mismo efecto pero solo en los modelos del tercer y cuarto quintil. Esto puede estar relacionado con el hecho que el Transantiago fue diseñado para satisfacer los viajes en periodos punta (buena conexión entre sectores residenciales y laborales).

La tasa de motorización es significativa en los modelos del tercer y cuarto quintil. En ambos modelos su coeficiente tiene un signo negativo, esto significa que en las comunas donde existe una mayor cantidad de autos por hogar sus residentes poseen una mayor probabilidad de disminuir su uso del sistema o abandonarlo. En cambio, el porcentaje de pobreza es significativo en los modelos del cuarto y quinto quintil. Esta variable también posee signos negativos. Este se interpreta como que en las comunas donde la pobreza es mayor existe una mayor probabilidad de que sus usuarios dejen de usar el transporte público. Esto último puede estar relacionado con la evasión. Los usuarios que deciden dejar de pagar el transporte público dejan de aparecer en la base de datos y por lo tanto son considerados como ausentes pese a seguir viajando en el sistema.

Finalmente, los usuarios que fueron catalogados como semi frecuentes y no frecuentes en el primer periodo de análisis poseen una mayor probabilidad de disminuir su uso del sistema o dejarlo en un futuro. Esto confirma la tendencia mencionada al principio del artículo.

## 5. CONCLUSIONES

### 5.1 Contribuciones

Este artículo propone una metodología para evaluar y explicar los cambios en el comportamiento de los usuarios. Primero, la clasificación de usuarios según la frecuencia de uso del sistema ha demostrado ser una herramienta útil a la hora de analizar la composición de usuarios que viajan en el sistema. Gracias a esta herramienta es posible identificar la porción de usuarios responsable de la mayoría de los viajes e ingresos de toda la red. Además, permite conocer la existencia de un conjunto importante de tarjetas que están prácticamente en desuso. Ser capaces de identificar estas diferencias entre los usuarios permite enfocar de mejor manera los recursos del sistema.

Por otra parte, el resultado de esta investigación se ha logrado gracias al esfuerzo de varios autores que previamente han enriquecido los datos pasivos de Transantiago. Estos datos poseen un potencial incalculable que puede ser utilizado para muchas otras investigaciones como ésta.

En cuanto al modelo, los resultados obtenidos son satisfactorios con respecto a la significancia y los signos de las variables. Se ha podido corroborar los resultados obtenidos por otros autores y además se han encontrado efectos propios de este escenario en materia de fuga de usuarios. Se ha encontrado que la calidad de servicio es uno de los aspectos más relevante a la hora de influenciar la decisión de los usuarios sobre su grado de uso del sistema. Cuando estos atributos se ven deteriorados, el descontento de los usuarios se ve reflejado directamente en la migración de usuarios o en la disminución en el uso del sistema. Sobre la preferencia entre los modos de transporte disponibles, el modelo indica que los usuarios más frecuentes son aquellos que combinan todos los modos (bus, metro, zonas pagas y buses expresos) pero que mayoritariamente viajan en metro. Las variables asociadas a los patrones temporales de las tarjetas dan a entender que el sistema entrega un buen servicio en horarios punta y durante días laborales ya que los usuarios que solo viajan en estos periodos poseen una mayor probabilidad de ser frecuentes. Separar la muestra en tres modelos hizo posible encontrar comportamientos distintos entre

cada estrato socio económico. Por ejemplo, los usuarios que habitan en las comunas del quinto quintil encuentran relevante la velocidad de viaje del transporte público, esto seguramente a causa de que ellos disponen de otras alternativas debido a su mayor ingreso. Por último, de las variables imputadas de la encuesta CASEN la tasa de motorización entregó los resultados esperados. En las comunas donde es mayor la tasa de motorización, también lo es la probabilidad de fuga de sus usuarios.

Los resultados obtenidos por esta investigación son de gran utilidad para los agentes planificadores de transporte público o incluso para las empresas operadoras. Mejorar los aspectos relacionados a la disminución del uso del transporte público permitiría detener la baja en su participación modal sin necesariamente disponer de muchos más recursos de los existentes hoy en día. Sino que distribuyéndolos de mejor forma.

## 5.2 Limitaciones

Una de las limitaciones más importantes de esta metodología es el criterio para definir una tarjeta ausente. Este criterio no es capaz de saber si una tarjeta está ausente por la decisión del usuario de dejar de usar el transporte público o por razones de fuerza mayor, como enfermedades, viajes de trabajo, pérdida de la tarjeta, etc. Este problema se busca solucionar al observar una ventana temporal más extensa en el segundo periodo pero aun así, esta medida está lejos de ser perfecta. Otro supuesto que no es siempre correcto, es que cada tarjeta es utilizada por un usuario. En un sistema donde las tarjetas son anónimas, muchas veces las tarjetas son compartidas entre varias personas. También, el modelo asume que los modos de transporte alternativos permanecen constantes entre los periodos. Sin embargo, la infraestructura para vehículos particulares o bicicletas puede haber cambiado durante estos seis meses, lo que afecta la decisión de los usuarios sobre qué modo utilizar. Las variables socio económicas utilizadas son limitadas (solo dos variables fueron significativas en los modelos), sería de gran utilidad conocer información demográfica sobre los portadores de las tarjetas. Por último, existe un factor climático que afecta la decisión de los usuarios. El primer periodo analizado es durante invierno y el segundo durante verano. Las temperaturas más agradables y el cese de las lluvias influencia a los usuarios a utilizar modos alternativos como la bicicleta o la caminata.

## 5.3 Trabajos futuros

Como trabajos futuros esta investigación puede ser mejorada en varias aristas. Primero, trabajar solo con ocho días para la creación de las variables explicativas puede no ser suficiente, ideal sería trabajar con un periodo de un mes donde se pueda capturar de forma más amplia el comportamiento de los usuarios y la experiencia que reciben por parte del sistema de transporte público. Por otra parte, la metodología descrita por Munizaga et al. (2017) es capaz de corregir los errores descritos por los tiempos de viaje, espera y caminata, lo que podría entregar mejores resultados con respecto a la calidad de servicio percibida. En esta investigación solo se observan dos ventanas temporales, esto puede ser expandido a más periodos haciendo un seguimiento más detallado de los usuarios y el grado de uso del sistema. Por último, el criterio para determinar la migración de usuarios en esta metodología está expuesto a errores como el de usuarios que pierdan su tarjeta. Este problema puede ser solucionado si se observan los patrones de las tarjetas y se emparejan tarjetas con una alta probabilidad de tener el mismo dueño.

## AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el proyecto Fondecyt 1161589, y por el Instituto Sistemas Complejos de Ingeniería (CONICYT – PIA – FB0816; ICM P-05-004-F).

## REFERENCIAS

- Amaya, M., Cruzat, R. y Munizaga, M. (2017) Estimating the residence zone of frequent public transport users to make travel pattern and time use analysis. *Journal of Transport Geography* (In press <https://doi.org/10.1016/j.jtrangeo.2017.10.017>)
- Bass, P., Donoso, P. y Munizaga, M. (2011) A model to assess public transport demand stability. *Transportation Research part A: Policy and Practice*, 45(8), 755-764.

- Devilleine, F., Munizaga, M. y Trépanier, M. (2012) Detection of activities of public transport users by analyzing Smart card data. *Transportation Research Record: Journal of the Transportation Research Board*, 2276, 48-55.
- DTPM (2017) Informe de Gestión 2015-2016. Disponible en [http://www.dtpm.cl/archivos/INFORME\\_GESTION\\_DTPM\\_15\\_16\\_2.pdf](http://www.dtpm.cl/archivos/INFORME_GESTION_DTPM_15_16_2.pdf).
- Imaz, A., Habib, K.M.N., Shalaby, A. y Idris, A.O. (2015) Investigating the factors affecting transit users loyalty. *Public Transport*, 7(1), 39-60.
- Langlois, G. , Koutsopoulos, H.N. y Zhao, J. (2016) Inferring patterns in the multi-week activity sequences of public transport users. *Transportation Research Part C: Emerging Technologies*, 64, 1-16.
- Munizaga, M. y Calvo, M.F. (2015) Análisis de la permanencia de usuarios frecuentes del sistema de transporte público usando datos pasivos. XVII Congreso Chileno de Ingeniería de Tránsito y Transporte, Concepción, Chile, 13 al 15 de octubre de 2015.
- Munizaga, M., Devillaine, F., Navarrete, C. y Silva, D. (2014) Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies*, 44, 70-79.
- Munizaga, M., Nuñez, C. y Gschwender, A. (2017) Smart card data for wider system evaluation. En: Kurauchi, F. y Schmöcker, J.-D. (Eds.) *Public Transport Planning with Smart Card Data*, CRC Press, 163-179.
- Munizaga, M. y Palma, C. (2012) Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, 24, 9-18.
- Muñoz, V., Thomas, A., Navarrete, C. y Contreras, R. (2016) Encuesta origen-destino de Santiago 2012: Resultados y validaciones. *Ingeniería de Transporte*, 19, 21-36.

