



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

**MODELACIÓN CONJUNTA DE  
ELECCIÓN DE LOCALIZACIÓN  
RESIDENCIAL Y ESTABLECIMIENTO  
ESCOLAR EN SANTIAGO, CHILE**

**FELIPE JOSÉ GONZÁLEZ ESPEJO**

Tesis para optar al grado de

Magíster en Ciencias de la Ingeniería

Profesor Supervisor:

**RICARDO HURTUBIA**

Santiago de Chile, Octubre, 2021.



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

# MODELACIÓN CONJUNTA DE ELECCIÓN DE LOCALIZACIÓN RESIDENCIAL Y ESTABLECIMIENTO ESCOLAR EN SANTIAGO, CHILE

**FELIPE JOSÉ GONZÁLEZ ESPEJO**

Tesis presentada a la Comisión integrada por los profesores:

**RICARDO HURTUBIA GONZÁLEZ**

**SEBASTIÁN RAVEAU FELIÚ**

**SEBASTIAN ASTROZA TAGLE**

**DENIS PARRA SANTANDER**

Para completar las exigencias del grado de

Magister en Ciencias de la Ingeniería

Santiago de Chile, Octubre 2021

A mi familia,  
incluyendo a la Mali.

## AGRADECIMIENTOS

Esta investigación no habría sido posible sin la ayuda de muchos y muchas. Primero, quiero aprovechar de darle las gracias a Ricardo, mi profesor guía, por todo el acompañamiento hecho semana a semana durante este año de trabajo. Muchas gracias por todo el apoyo intelectual y mental que me diste para llegar a la meta. ¡Ha sido un gusto trabajar juntos! También, agradezco a Sebastián Astroza, que nos acompañó también durante la investigación aportando su conocimiento en materia de modelos de elección conjunta. Además, le doy las gracias al profesor Sebastián Raveau por su apoyo durante la investigación. Muchas veces me entregó momentos de claridad cuando las cosas se ponían difíciles. Quiero aprovechar también de agradecerle por todos los semestres que lo tuve como profesor, siendo alumno o ayudante. Fue un gusto conocernos, y muchas gracias por todo lo enseñado a lo largo de mi carrera universitaria. Todos ustedes me motivaron y me enseñaron lo apasionante que puede ser la Ingeniería en Transporte. Me voy muy contento con las buenas decisiones que tomé con respecto a mi carrera.

Aprovecho también de agradecer a CONICYT y a ISCI, por sus fondos que me ayudaron durante la investigación. El apoyo a la investigación es un apoyo para el desarrollo de la sociedad y las personas. Además, quiero agradecer a FONDECYT 1180605 y a CEDEUS.

Para finalizar, y más importante, quiero agradecer a mi familia, a mis hermanas, mi papá y mi mamá. Sin ustedes no sería quien soy hoy en día, y quiero decirles que me hacen muy feliz. También, gracias a mis amigos y amigas. Su apoyo incondicional a lo largo de mi vida escolar y universitaria ha significado más de lo que imaginan para mí. Gracias por las risas, los buenos y malos ratos, por los almuerzos en la universidad, las salidas a todos los lugares que fuimos, y por estar conmigo en todo momento, buenos y malos. Por último, gracias a mi polola por estar ahí siempre. Gracias por aguantarme todo este año, con mis quejas y ansiedades, y gracias por siempre estar esperándome con una sonrisa y los brazos abiertos. Me alivianaste la carga en los momentos más difíciles. Sin duda el magíster ha sido un proceso tortuoso, pero del cual salimos victoriosos, felices y con muchos aprendizajes. No le doy las gracias al COVID-19. No se lo merece.

## INDICE GENERAL

	Pág.
AGRADECIMIENTOS .....	iii
INDICE DE TABLAS .....	vii
INDICE DE FIGURAS.....	ix
RESUMEN.....	x
ABSTRACT .....	xi
1. INTRODUCCIÓN.....	1
1.1. Hipótesis y objetivos de investigación .....	2
1.2. La tarea: ¿cómo lograr estos objetivos?.....	3
1.3. Estructura .....	3
2. REVISIÓN DE LITERATURA .....	5
2.1. Elección de establecimiento educacional escolar y el caso de Chile .....	5
2.1.1. El sistema escolar chileno.....	5
2.1.2. Atributos académicos, valóricos y socioeconómicos de la elección de establecimiento escolar.....	7
2.1.3 El rol de la localización residencial en la elección del establecimiento.....	9
2.2 Modelos de elección conjunta: localización residencial y otras elecciones.....	11
2.2.1. Modelos Logit Multinomiales y Logit Jerárquicos.....	11
2.2.2. Modelos con otras estructuras de error.....	13
2.2.3 Atributos relevantes para la elección de localización residencial..	16
2.3. Aportes de la investigación a la literatura científica .....	17
3. METODOLOGÍA.....	19
3.1 Modelo Logit Multinomial.....	19
3.2 Modelo Logit Mixto capturando correlación entre elecciones.....	20
3.2.1. Hogares priorizadores de localización residencial (PLR).....	21
3.2.2. Hogares priorizadores de establecimiento escolar (PEE).....	23

3.3	Enfoque de clases latentes.....	24
3.4	Muestreo de alternativas.....	26
4.	RECOLECCIÓN Y PROCESAMIENTO DE DATOS .....	29
4.1	Antecedentes recolectados .....	30
4.1.1	Hogares que poseen viajes por educación escolar .....	30
4.1.2	Atributos de alternativas de localización residencial.....	30
4.1.3	Atributos de alternativas de establecimientos escolares .....	31
4.2	Procesamiento de la información recolectada .....	32
4.2.1	Alternativas de localización residencial.....	32
4.2.2	Alternativas de establecimientos escolares .....	42
4.2.3	Hogares y sus elecciones observadas.....	47
4.3	Muestreo de alternativas y cálculo de distancia entre ellas. ....	50
4.3.1	Muestreo de alternativas de localización residencial.....	50
4.3.2	Muestreo de alternativas de establecimiento escolar .....	51
4.3.3	Cálculo de distancia entre alternativas.....	54
5.	Análisis de los datos: el caso de SANTIAGO, CHILE .....	55
5.1	Patrones de localización residencial.....	55
5.1.1	Valor de arriendo .....	55
5.1.2	Accesibilidad.....	56
5.1.3	Características socioeconómicas de los barrios .....	58
5.2	Patrones de alternativas de establecimientos escolares.....	60
5.2.1	Análisis muestral de alternativas de establecimientos escolares ...	63
5.3	Hogares y viajes con propósito educación escolar .....	66
5.4	Análisis de la muestra de observaciones EOD.....	70
6.	RESULTADOS Y DISCUSIÓN .....	75
6.1	Especificación de variables del modelo .....	75
6.2	Resultados para el modelo Logit Multinomial (MNL) .....	77
6.3	Modelos Logit Mixto capturando correlación entre elecciones .....	81
6.3.1	Análisis de los coeficientes del modelo .....	84
6.3.2	Factores no observados que capturan correlación entre elecciones.....	87
6.3.3	Cálculo de elasticidades y comparación de estrategias.....	88
6.4	Modelo Logit Mixto: enfoque de clases latentes.....	93

6.4.1 Clase latente .....	93
6.4.2 Análisis de los parámetros estimados .....	96
6.4.3 Cálculo de elasticidades .....	100
6.5 Comparación de modelos .....	104
7. CONCLUSIONES .....	105
BIBLIOGRAFIA.....	110
A N E X O S.....	113
Anexo A : Modelo Logit Jerárquico .....	114

## INDICE DE TABLAS

	Pág.
Tabla 4-1: Factores de impedancia iniciales considerados según uso de suelo y modo. ....	37
Tabla 4-2: Atributos que caracterizan las alternativas de localización residencial. ....	41
Tabla 4-3: Atributos que caracterizan las alternativas de establecimiento escolar. ....	46
Tabla 4-4: Factores de corrección para alternativas de localización residencial muestreadas. ....	51
Tabla 4-5: Factores de ajuste según alternativa escogida y características de la observación. ....	53
Tabla 5-1: Establecimientos educacionales y sus atributos principales. ....	60
Tabla 5-2: Descripción del universo muestral de alternativas de establecimiento escolar. .	65
Tabla 5-3: Caracterización de hogares con estudiantes escolares según quintil de ingreso. ....	66
Tabla 5-4: Características de la muestra de hogares y estudiantes. ....	71
Tabla 6-1: Variables consideradas en los modelos, tipo de variable y rango. ....	76
Tabla 6-2: Resultados modelo MNL. ....	78
Tabla 6-3: Resultados modelos Logit Mixto con estrategias mixtas. ....	83
Tabla 6-4: Coeficientes aleatorios en elección de establecimiento escolar y probabilidad de ser negativos, según estrategia modelada. ....	85
Tabla 6-5: Coeficientes aleatorios en elección de localización residencial y probabilidad de ser negativos, según estrategia modelada. ....	86
Tabla 6-6: Elasticidades de demanda para elección de establecimiento, según estrategia. .	89
Tabla 6-7: Elasticidades de demanda para elección de localización residencial, según estrategia. ....	91
Tabla 6-8: Resultados de parámetros de clase latente. ....	94
Tabla 6-9: Resultado modelo con enfoque de clases latentes. ....	97
Tabla 6-10: Coeficientes aleatorios de elección de establecimiento escolar y probabilidad de ser negativos según clase latente. ....	98
Tabla 6-11: Coeficientes aleatorios de elección de localización residencial y probabilidad de ser negativos según clase latente. ....	99

Tabla 6-12: Elasticidades de demanda para elección de establecimiento escolar, según clase latente. ....	100
Tabla 6-13: Elasticidades de demanda para elección de localización residencial, según clase latente. ....	102

## INDICE DE FIGURAS

	Pág.
Figura 4-1: Diagrama de consolidación de la base de datos de elección a generar. ....	29
Figura 4-2: Ejemplo de alternativa de localización residencial. ....	33
Figura 4-3: Accesibilidad a oficina en modo caminata según distribución de viajes por trabajo de EOD 2012. ....	36
Figura 4-4: Accesibilidad a oficina según modo. ....	38
Figura 4-5: Accesibilidad a comercio según modo. ....	39
Figura 4-6: Accesibilidad a educación según modo. ....	40
Figura 4-7: Ejemplo de alternativa de establecimiento escolar y su entorno. ....	45
Figura 4-8: Ejemplo de metodología de georreferencia de establecimientos escolares. ....	49
Figura 5-1: Valor promedio mensual de arriendo de vivienda en \$MCLP. ....	56
Figura 5-2: Accesibilidad en transporte público a educación, oficina y comercio ....	57
Figura 5-3: Porcentaje de jefes del hogar con educación superior. ....	59
Figura 5-4: Distribución de establecimientos según rango de mensualidad ....	61
Figura 5-5: Distribución de establecimientos educacionales según dependencia. ....	62
Figura 5-6: Universo muestral de alternativas escolares según dependencia. ....	64
Figura 5-7: Partición modal de viajes escolares entre las 6:01 y 9:00 am. ....	67
Figura 5-8: Distribución de los tiempos de viaje según quintil de ingreso. ....	68
Figura 5-9: Distribución de distancias entre el hogar y el colegio según de ingreso. ....	quintil 69
Figura 5-10: Distribución de distancias entre hogar y colegio según muestra. ....	73
Figura 5-11: Distribución de tiempos de viaje al colegio según muestra. ....	74
Figura 6-1: Probabilidad de escoger según estrategia PEE. ....	95
Figura A-1: Modelo Logit Jerárquico con una estrategia de decisión priorizadora de localización residencial. ....	114

## RESUMEN

Las preferencias de los hogares sobre la elección de establecimiento escolar han sido ampliamente estudiadas por la literatura, en particular en Chile debido a sus singularidades con respecto a la perspectiva internacional. Estas preferencias han demostrado ser heterogéneas y complejas. Además, han sido vinculadas de manera consistente con la elección de localización residencial. Es más, en Santiago, cada día más de 900 mil escolares viajan en hora punta mañana a estudiar. No obstante, la literatura de modelación conjunta de elecciones no ha profundizado en el vínculo intrínseco entre estas decisiones.

Este estudio busca comprender el vínculo existente entre ambas elecciones mediante modelos de elección conjunta. Así, se construyó una base de datos con observaciones de la Encuesta Origen y Destino del Gran Santiago 2012 y se generaron conjuntos de alternativas de localización residencial y escolar para cada hogar. Se estimó un modelo Logit Multinomial, dos modelos Logit Mixto (que proponen diferentes estructuras de decisión y correlacionan ambas elecciones a través de términos de error), y un modelo de clases latentes que busca caracterizar a los hogares, según las estrategias definidas.

Las conclusiones obtenidas de los modelos radican en tres aspectos. Primero, se corroboran las conclusiones de la literatura respecto de la percepción de atributos académicos, socioeconómicos y la distancia entre hogar y colegio. Segundo, se captan factores no observados asociados a atributos socioeconómicos y de accesibilidad que determinan las preferencias sobre ambas elecciones, correlacionándolas. Tercero, existe heterogeneidad caracterizable en las preferencias de los hogares. Los resultados más relevantes indican que la distancia es un factor esencial para todos los hogares, independiente de la estrategia de elección. Además, hogares que priorizan el establecimiento por sobre la localización, tienden a ser más sensibles sobre atributos socioeconómicos de ambas elecciones. Comprender este vínculo, permite diseñar políticas públicas que incluyan ambas dimensiones de elección.

Palabras clave: *Elección escolar, elección de localización, modelación conjunta, factores no observados.*

## ABSTRACT

Households' preferences about school choice have been vastly studied by literature. Specifically in Chile, due to its singularities with respect to the international picture. These preferences have demonstrated to be complex and heterogenous, as they have also been linked consistently with households' residential location choices. In Santiago, Chile, more than 900 thousand school students travel at peak hours to school. However, joint choice modelling literature has not discussed this bond between choices deep enough.

The following study aims to understand the interactions bonding both choices through discrete joint choice models. A data base containing household observations from the Santiago 2012 Origin Destination Survey was constructed, adding residential location and school alternatives for each household. A Multinomial Logit Model was estimated, in addition to two Mixed Logit Models (defining different decision structures and correlating both choices through random error terms). Finally, a Latent Segmentation Model was estimated by incorporating each Mixed Logit model decision structure, capturing preferences based on household's characteristics.

The main conclusions driven from the results rely in three aspects. First, previous studies conclusions are corroborated in terms of household's perception of academic, socioeconomic attributes, added to the relevance given to the distance between residence and school. Second, unobserved factors associated with access to job locations and neighborhoods' educational level are identified, determining, and establishing correlation between both residential location and school choice. Third, heterogeneity in household's preferences exist and is characterizable. Most relevant results indicate that distance is an essential factor for decision makers, independent of the choice strategy used. Furthermore, households that prioritize school over location choice, tend to be more sensitive to socioeconomic factors that affect both decisions. Understanding these choices together, helps to design appropriate policies that can include both decisions and their interactions.

Keywords: *School choice, location choice, joint choice modelling, unobserved factors.*

## 1. INTRODUCCIÓN

Comprender las elecciones de los hogares con respecto a sus alternativas de establecimiento escolar y de localización residencial es un problema de investigación complejo que entrega valiosa información respecto de sus preferencias y su comportamiento. El caso chileno presenta un contexto interesante para analizar estas elecciones. Desde la perspectiva de la elección de establecimiento escolar, el sistema chileno se presenta como uno singular en el mundo. Chile, durante década de 1980, se convirtió en el primer país en implementar un sistema de subvención a la demanda a nivel nacional, buscando simular de manera estricta el sistema original (Chumacero, Gómez y Paredes, 2011). Así, ha transformado al país en un caso de estudio útil para comprender las dinámicas sociales generadas por este sistema.

Tomando esto en cuenta, la literatura ha demostrado que el sistema escolar chileno ha generado marcadas desigualdades según los tipos de establecimientos y estratos socioeconómicos de las familias (Elacqua, 2012; Correa et al., 2019; Gayo, Otero y Méndez, 2019), junto con dinámicas sociales de reproducción de clases, que por cierto son heterogéneas (Bellei, Orellana y Canales, 2020). Es más, Elacqua, Schneider y Buckley (2006) estudian cómo los hogares compensan atributos académicos con aspectos socioeconómicos de los establecimientos. Esto, en un contexto donde el principio del sistema responde a entregar libertad de elección a las familias al momento de decidir la educación de sus hijos e hijas.

De la misma forma, otra razón que amerita mencionar para analizar las elecciones de localización residencial y establecimiento escolar corresponde a la relación intrínseca que las vincula. Al año 2012, en Santiago entre las 6:01 y las 9:00 AM, habían más de 927 mil viajes por estudio al colegio, realizados por más de 390 mil hogares (SECTRA, 2014). Esto corresponde a cerca de un 25% del total de viajes realizado en el período punta mañana de un día laboral normal, e involucra a 1 de cada 5 hogares del área metropolitana estudiada.

La literatura ha explorado en profundidad la relación entre la elección residencial y la elección escolar. La distancia entre el hogar y el establecimiento es un factor en la elección de colegio (Chumacero et al., 2011). Por otra parte, se ha mostrado que los hogares poseen

expectativas similares respecto de ambas elecciones y los beneficios económicos, culturales o sociales que pueden entregarles a pesar de las diferentes estrategias de elección que puedan poseer (Méndez y Gayo, 2019), en particular para una ciudad marcada por inequidades territoriales. En un contexto donde el sistema de actividades determina e influye en el sistema de transporte de las ciudades, estudiar las preferencias de los hogares con respecto a la elección de localización residencial y establecimiento escolar resulta útil para reflexionar, comprender y planificar la ciudad.

Entonces, es razonable pensar que hay factores que vinculan a ambas elecciones. Además, el sistema escolar de Chile es singular y se ha estudiado en profundidad las preferencias de los hogares respecto a esta elección. Por lo tanto, se presenta una oportunidad única de investigar con respecto a las preferencias de los hogares en torno a estas elecciones, y sus sensibilidades tanto a aspectos escolares como de localización residencial, para comprender el comportamiento de la sociedad.

### **1.1. Hipótesis y objetivos de investigación**

La hipótesis de trabajo de la investigación se fundamenta en que ambas elecciones están vinculadas entre sí. Factores observables, como la distancia, al igual que no observables, como estilos de vida, opiniones y expectativas, determinan las preferencias de los hogares con respecto a ambas elecciones, condicionándose la una a la otra. Además, es esperable que exista heterogeneidad en las preferencias y en las estrategias de los hogares respecto a las elecciones.

El objetivo de esta investigación es comprender las preferencias de los hogares de Santiago, Chile en torno a las elecciones de localización residencial y de establecimiento escolar, con el fin de determinar correlaciones entre ambas elecciones y caracterizar a los hogares según sus preferencias de elección. De este modo, los objetivos específicos de la investigación son tres. Primero, modelar en conjunto la elección de localización residencial y de establecimiento escolar a través de modelos de elección discreta. Luego, estimar correlación entre ambas elecciones a través de la estructura y especificación definida por el modelo. Tercero, caracterizar a los hogares según sus preferencias y estrategias de decisión con

respecto a ambas elecciones, comprendiendo qué hogares priorizan una elección sobre la otra.

## **1.2. La tarea: ¿cómo lograr estos objetivos?**

Para lograr estos objetivos de investigación, se elaboró una base de datos que posee información de las características de los hogares y sus elecciones observadas de localización residencial y establecimiento escolar (así como un conjunto de alternativas posibles). Esta base de datos se construyó a partir de información de la Encuesta Origen Destino 2012 (SECTRA, 2014). Además, se utilizó información del Censo 2017 (INE, 2018) y del Catastro de Bienes Raíces del SII (2014) para consolidar alternativas de localización residencial. También, las alternativas de establecimiento escolar se construyeron en base al Directorio de Establecimientos Educativos 2014 (MINEDUC, 2014) y a otras fuentes bases de datos con información de los establecimientos (Agencia Calidad de la Educación, 2006; DEMRE, 2006; MINEDUC, 2010; MINEDUC, 2006).

Esta base de datos se utilizó para estimar modelos de elección discreta (McFadden, 1974), modelando la elección conjunta de localización residencial y establecimiento educacional. Las especificaciones de los modelos, junto con las estructuras de error definidas se determinaron según lo propuesto por la literatura con respecto a ambas elecciones, estableciendo correlaciones entre ellas. Además, para caracterizar a los hogares según sus preferencias, se propuso un enfoque de clases latentes que capte diferentes sensibilidades respecto de ambas elecciones. Así, se espera poder analizar los resultados de estos modelos, comprender las correlaciones existentes, al igual que comparar las preferencias de los hogares según sus características.

## **1.3. Estructura**

La tesis se estructura de la siguiente forma. En la Sección 2, se realiza un análisis recabado de los antecedentes existentes de esta área investigativa. Por un lado, se estudia la elección de establecimiento escolar, los atributos que la determinan y el vínculo que posee con la elección de localización residencial. Por el otro, se profundiza en modelos de elección

conjunta de elecciones y las metodologías que se han aplicado para captar correlaciones y determinar dependencias entre elecciones. Se discute también sobre los aportes de la investigación en el mundo científico.

La Sección 3 describe la metodología a implementar para cuatro modelos a estimar: un modelo Logit Multinomial, dos modelos Logit Mixto que poseen diferentes estructuras de decisión, un modelo Logit Mixto con enfoque de clases latentes que caracteriza a los hogares según sus estrategias de elección y sus preferencias. Además, describe la metodología utilizada para definir el conjunto de alternativas de ambas dimensiones de elección.

Luego, la Sección 4 describe el proceso de recolección de datos y posterior procesamiento de estos, para la construcción de la base de elección. Se describen los antecedentes recolectados, junto con las decisiones tomadas y supuestos definidos. En particular, explica la consolidación de alternativas de localización residencial, de colegio y la muestra a utilizar para la estimación junto con sus elecciones observadas.

Después, la Sección 5 describe la ciudad de Santiago con respecto a su comportamiento en torno a ambas elecciones. Por una parte, describe los mercados de localización residencial y de establecimiento escolar y sus principales atributos. Luego, analiza el comportamiento de los hogares con respecto a sus características y a ambas elecciones.

La Sección 6 expone y discute los resultados para los modelos estimados utilizando la base de datos construida. Se analiza, para cada modelo la relevancia de sus resultados y las preferencias de los hogares. Además, se calculan elasticidades de demanda con respecto a ambas elecciones para poder comparar las preferencias de los hogares según las estrategias de elección que adopten.

Para finalizar, la Sección 7 concluye respecto al cumplimiento de los objetivos propuestos y la validación de la hipótesis de trabajo. Además, levanta las principales limitaciones de la investigación y reflexiona sobre el impacto que tiene sobre la comprensión de las preferencias de los hogares con respecto a ambas elecciones.

## **2. REVISIÓN DE LITERATURA**

Esta investigación pretende indagar en dos aspectos relevantes de la literatura, con el objetivo de poder aportar de manera crítica a esta. El primero corresponde a la elección de establecimiento educacional escolar. Se analizarán cómo son las preferencias de los hogares en torno a esta, en particular para el caso de la sociedad chilena. Mientras, el segundo aspecto a estudiar se vincula con la modelación conjunta de elecciones, enfocado en la elección de localización residencial y otras decisiones que los hogares toman.

Comprender en conjunto ambos problemas de investigación da pie a estimar de manera adecuada modelos conjuntos que vinculen las elecciones de localización residencial y establecimiento escolar. Además, da un sustento fundamentado para analizar de manera crítica y contextualizada los resultados obtenidos, adecuándose a la realidad de Santiago.

Primero, se presentarán revisión bibliográfica en torno a elección de establecimiento escolar y cómo esta se comporta en el contexto chileno. Luego, se analizará en detalle aspectos metodológicos de la modelación en conjunto de la localización residencial y otras variables. Para finalizar, se discutirá en qué área de la literatura esta investigación es un aporte.

### **2.1. Elección de establecimiento educacional escolar y el caso de Chile**

Esta sección busca analizar el sistema escolar chileno y el comportamiento de los hogares respecto de esta elección. Además, busca analizar el vínculo existente entre la elección escolar y la localización residencial.

#### **2.1.1. El sistema escolar chileno**

La elección de establecimiento escolar en Chile ha sido un caso de estudio ampliamente revisado en el tiempo. En particular, el sistema de vouchers (o de subvención a la demanda) implementado en la década de 1980, convirtió a Chile en el primer país en haber implementado este sistema a nivel nacional (Chumacero et al., 2011). Así, se dio origen a

los establecimientos particulares subvencionados además establecimientos municipales y particulares pagados ya existentes.

Gallego y Sapelli (2007) resumen de manera adecuada el sistema. A grandes rasgos, este propone financiar total o parcialmente la mensualidad escolar de los estudiantes, a través de un “cheque” a la familia. Este cheque, para el caso chileno, se traduce en una subvención directa a los establecimientos por cada alumno o alumna que atiende a ellos. De esta forma, como la subvención es para el estudiante, los hogares tienen la libertad de escoger un establecimiento que se adecúe a sus intereses educativos. Mientras, los colegios deben competir por atraer alumnos, lo que se logra mediante mejoras a la calidad y eficiencia de la educación ofrecida.

Con el tiempo se ha evidenciado que se genera segregación social en la educación (Elacqua, 2012; Correa et al., 2019). Los mecanismos que producen esta segregación son variados. Por ejemplo, el sistema de selección escolar permite a los establecimientos elegir entre una familia u otra, lo que genera discriminación de los estudiantes según la afinidad con el establecimiento con respecto características socioeconómicas, académicas o valóricas (Gayo et al., 2019; Valenzuela, Bellei y Ríos, 2014). Otro mecanismo discriminatorio radica en el copago escolar, que permite a los establecimientos subsidiados por el estado a cobrar una mensualidad a las familias, segregando según sus capacidades de pago (Valenzuela et al., 2014). Estas falencias han sido en parte mitigadas a través de la Ley de Inclusión Escolar (BCN, 2015), que pretendía acabar con la discriminación en los procesos de entrada a los establecimientos.

Un último aspecto discriminatorio que permite el sistema escolar, el cual esta revisión bibliográfica busca reportar y comprender, es el comportamiento de los hogares respecto a esta elección. La esencia del sistema escolar chileno radica en que las familias tengan la capacidad, de decidir entre un establecimiento u otro. Lo cual ha incentivado mayor estratificación producto de la capacidad de hogares de mayor nivel socioeconómico a escoger los establecimientos de mayor prestigio (Valenzuela et al., 2014). Entonces, comprender las preferencias de los hogares en torno a esta elección, permite entender también el rol que las familias juegan en la dinámica social presente. En particular, es útil

analizar cómo las familias en Chile aprecian aspectos académicos y valóricos de los establecimientos, pero también consideran aspectos socioeconómicos en su elección, al igual que aspectos con respecto a su lugar de residencia. Estas preferencias se discuten a continuación.

### **2.1.2. Atributos académicos, valóricos y socioeconómicos de la elección de establecimiento escolar.**

La literatura científica profundiza en los *trade-offs* que las familias hacen entre diferentes atributos al momento de elegir. Elacqua et al. (2006) detalla en particular el vínculo existente entre la percepción de la calidad escolar, y aspectos socioeconómicos de los establecimientos. Mediante entrevistas a padres y madres de estudiantes, logra demostrar diferencias entre las preferencias declaradas como las reveladas por las familias. Por un lado, la mayoría de los apoderados declaran valorar primero el ambiente académico que ofrecen los colegios. Otros atributos declarados como relevantes estaban asociados a la localización de los colegios, junto con la oferta valórica impartida. Esta última, ganaba prioridad hacia hogares con mayores niveles de ingreso. En el estudio de Elacqua et al. (2006) rara vez se reportó el nivel socioeconómico de las familias del establecimiento como un atributo relevante. Sin embargo, al observar los establecimientos que cada hogar consideró al momento de escoger, los resultados indicaban lo contrario. En promedio, sólo 1 de cada 4 hogares escogió el establecimiento con mejor rendimiento académico, dentro de sus alternativas. Mientras, un 70% de los hogares consideraron establecimiento de similares valores, y más de un 85% consideraron establecimientos similares en cuanto al niveles socioeconómico del cuerpo estudiantil.

Estas diferencias reflejan cómo, a pesar de que los hogares declaran valorar principalmente aspectos académicos de los establecimientos, tienden a preferir establecimientos que se adecúen a sus requerimientos socioeconómicos. Esto podría estar asociado a cómo las familias compensan un atributo por sobre otro.

Este comportamiento tiende a exacerbarse hacia los niveles socioeconómicos de mayor ingreso. Bellei et al. (2020), a través de entrevistas y grupos de discusión, caracteriza a los

hogares de la clase media alta chilena. Por una parte, define a aquellas familias que escogen los establecimientos por un sentido de pertenencia a su comunidad, valorando en casos extremos de manera lexicográfica este atributo en la elección. En ciertos contextos religiosos, se visibilizó cómo esta visión comunitarista se exacerbaba. Siempre, desde un enfoque hacia el vínculo colegio – familia, y cómo la elección te inserta dentro de una comunidad. Por otro lado, reconoce la existencia de hogares que valoran el éxito académico por sobre el sentido comunitario que entrega el establecimiento, en miras de entregarle a sus hijos e hijas educación que les asegure un futuro como profesionales exitosos. Aspectos como el nivel de inglés enseñado eran muy valorados por estos hogares, por sobre pruebas estandarizadas de rendimiento. Así, demuestra que a pesar de la marcada preferencia por colegios que entreguen un sentido de comunidad en la clase alta, puede existir heterogeneidad en los hogares y valoraciones distintas entre lo académico, valórico y social.

Cabe mencionar, que este comportamiento de pertenencia y reproducción social (Bellei et al., 2020) es razonable en el contexto chileno. Se ha demostrado la correlación entre el colegio de procedencia (y su “status”) con el éxito profesional y universitario (Zimmerman, 2019). Por lo tanto, es razonable que un hogar valore la demografía del alumnado y la relevancia que puede tener para el éxito de sus hijos e hijas. Esto es un reflejo de las preferencias de los hogares sobre los atributos de los establecimientos educacionales, en el contexto chileno.

En línea con la valoración de aspectos académicos en la elección, Gómez, Chumacero y Paredes (2012) discuten cómo padres y madres declaran valorar la calidad académica, como pruebas estandarizadas de enseñanza. A través de la estimación de modelos con dos cortes temporales, demuestran cómo, al tener conocimiento de los resultados de estas pruebas, la valoración por ellas aumenta. Esto es relevante para la investigación, ya que demuestra que la calidad académica es relevante para los hogares de manera transversal. No obstante, los autores ponen en cuestión el acceso a la información que se tiene con respecto a la calidad de la educación, y cómo los hogares compensan este atributo sobre otros, como la distancia. Esta última variable será discutida en profundidad en la subsección siguiente.

### **2.1.3 El rol de la localización residencial en la elección del establecimiento**

Dentro del análisis hay que considerar el rol que juega la localización residencial en la elección del establecimiento escolar. Chumacero et al. (2011) compara atributos y sus impactos sobre la probabilidad de elección. Por un lado, demuestra que la probabilidad de escoger un establecimiento aumenta a medida que se encuentra más cerca de la residencia. Similar es lo que ocurre con la calidad del establecimiento, que aumenta su probabilidad de ser seleccionado. Mientras, estima que, a mayor mensualidad del colegio, *ceteris paribus*, menor es la probabilidad de elección.

Los modelos estimados por Chumacero et al. (2011) explican que la valoración de la distancia es heterogénea entre hogares. Los hogares de mayor ingreso son menos propensos a escoger el establecimiento más cercano a ellos. También, aquellos hogares que escogen establecimientos municipales tienden a escoger el más cercano, mientras que los hogares que eligen establecimientos privados (i.e. de mayor ingreso) poseen mayor dispersión en términos de la distancia. Estos resultados van en línea con las estimaciones de Elacqua et al. (2006) donde, de los hogares encuestados, aquellos que eligen establecimientos públicos declaran tres veces más la variable de distancia como la más importante, en comparación con los hogares que escogen colegios del sector privado.

Además de la distancia, existen otros aspectos de la localización residencial que afectan en la elección de establecimiento educacional. Méndez y Gayo (2019) profundizan en el vínculo existente entre ambas elecciones, en particular para los hogares de nivel socioeconómico alto. En su revisión bibliográfica, mencionan cómo la elección de localización residencial puede determinar estilos de vida e identidad social, donde las características socioeconómicas juegan un rol clave en la elección, pues entregan un sentido de pertenencia a la comunidad (Savage et al., 2010, citado en Méndez y Gayo, 2019). Esto va de la mano con lo que plantea Valenzuela et al. (2014), donde la segregación que existe a nivel de localización residencial tiende a reproducir la segregación presente a nivel de elección escolar. Bajo esta mirada, la interpretación es similar a lo expuesto por la literatura visitada en los párrafos previos, donde se visibiliza que existen aspectos socioeconómicos, tanto del

colegio como de la residencia, que determinan la elección de establecimiento escolar. Así, es razonable pensar que ambas elecciones pueden estar vinculadas, y que una puede depender de la otra.

Los principales resultados de la investigación de Méndez y Gayo (2019) exploran cómo las expectativas de los hogares con respecto a la localización residencial y al establecimiento escolar están relacionadas. Así, aquellos hogares que poseen altas expectativas en torno a la localización residencial y los beneficios que entrega (como buena oferta de transporte, calidad de vida o éxito económico/profesional), se pueden agrupar con aquellos hogares con altas expectativas en torno a los beneficios que entrega la elección escolar (como éxito económico/profesional, buena oferta de transporte, valores, entre otros). Esto es similar para hogares con bajas o medianas expectativas. De esto, se desprende que ambas elecciones están vinculadas por los beneficios esperados que estas entregan a cada hogar, y cómo las características de cada tomador de decisión pueden determinar estas expectativas, y por ende sus preferencias por una u otra alternativa.

Cabe mencionar que, todo este análisis está sujeto a las capacidades reales que poseen los hogares de acceder tanto a un establecimiento como a una localización. En particular para la elección de colegio, las expectativas no siempre se cumplen debido a que existen mecanismos de selección por parte de los colegios. Así, hogares con intenciones de escoger un establecimiento, no siempre logran acceder a este (Elacqua, 2012; Gayo et al., 2019). Además, las capacidades de pago de los hogares pueden influir en la capacidad de acceder tanto al establecimiento como a una localización. La investigación no profundiza en estos aspectos (i.e. asume que las elecciones reveladas son un reflejo de las preferencias de los hogares), aunque a futuro debería retomarse.

Tomando en cuenta la literatura presente sobre la elección de establecimiento educacional, se puede concluir que es una elección compleja y de muchas dimensiones. Los hogares compensan atributos como el cobro de mensualidad, la calidad de la educación, la oferta valórica, las características sociodemográficas del cuerpo estudiantil, al igual que la distancia desde el hogar, para poder determinar la elección más beneficiosa para sus familias. Además, es una elección heterogénea entre los hogares que, dependiendo de sus

características, podrán tener diferentes preferencias y expectativas con respecto a la elección y deben además ser al mismo tiempo “seleccionados” por el establecimiento.

## **2.2 Modelos de elección conjunta: localización residencial y otras elecciones**

Los modelos de elección conjunta han sido estudiados de manera extensa por la literatura de modelación de transporte y uso de suelo. Los tomadores de decisiones deben elegir alternativas en diferentes dimensiones de elección, ya sea a nivel de localización residencial, como a nivel de modo de viaje, destino o horario, entre otras. Estas decisiones tienden a determinarse e influenciarse entre sí (Lerman, 1976), lo que hace razonable modelarlas en conjunto. Esta sección pretende describir diferentes modelos propuestos por la literatura, desde las estructuras más básicas a otras con mayor complejidad, en torno a modelos de elección conjunta y al rol de la localización residencial en estos.

### **2.2.1. Modelos Logit Multinomiales y Logit Jerárquicos**

Las primeras aproximaciones hechas por la literatura en esta materia se enfocaron en la utilización del modelo Logit Multinomial (McFadden, 1974) (MNL) para modelar en conjunto dos o más elecciones vinculadas entre sí. Lerman (1976), modela en conjunto la elección de localización, de residencia, posesión vehicular y modo de viaje, en lo que determina como un *mobility bundle*. Es decir, cada alternativa de elección equivale a una combinación de una alternativa de localización, número de vehículos y modo. Por lo mismo, existen tantas alternativas posibles como el múltiplo entre el número de alternativas de cada dimensión. Así, cada alternativa combinada está determinada por los atributos de cada una de las categorías que componen la alternativa. De manera similar, Ben-Akiva y Lerman (1985) describen de manera formal los modelos de elección multidimensionales, a través de la combinación de alternativas de diferente índole. Estas elecciones están determinadas por atributos de cada dimensión, como también por atributos observados compartidos. Por ejemplo, la elección de localización residencial y de trabajo puede estar determinada por atributos de la localización, del trabajo, y por la distancia entre ambos.

En este sentido, existen varias alternativas que comparten atributos similares con respecto a una dimensión de elección. Por ejemplo, aquellas alternativas que combinen la misma elección de localización residencial con diferentes modos de viajes compartirán los mismos atributos observados que determinan la elección residencial. De la misma manera, estas alternativas pueden compartir factores no observados asociados a la localización residencial que las determinan (Ben-Akiva y Lerman, 1985). El modelo MNL, supone una estructura de error simple donde las alternativas son independientes entre sí, por lo que no es capaz de captar estos factores no observados que correlacionan las alternativas.

Frente a esto, el modelo Logit Jerárquico (HL) surge como un enfoque de modelación que logra captar estos factores compartidos no observados. Esto se logra mediante un término de error que se repite para aquellas alternativas que comparten una misma dimensión de elección (e.g. comparten la misma alternativa de localización residencial). Estos términos de error determinan qué dimensión de elección es condicional a la otra, lo que establece una jerarquía de decisiones, entre alternativas de distintas dimensiones (Ben-Akiva y Lerman, 1985), como por ejemplo elección de modo y horario de viaje. Dado que el modelo MNL supone independencia entre alternativa, no logra captar estas correlaciones en la elección.

Ben-Akiva (1973), propone modelos HL entre la elección conjunta de destino y modo de viaje. Se estiman tres modelos. Primero, un modelo MNL donde no existen factores no observados entre alternativas. Luego un modelo HL donde las alternativas que comparten el mismo destino comparten también este término de error. Después, un modelo HL donde las alternativas que comparten modo comparten atributos no observados. Los resultados sugieren que la estructura de decisión, definida por los atributos no observados, sí es importante al momento de elegir elecciones multidimensionales. Waddell (1993), realiza un ejercicio similar, estudiando la elección de localización residencial, tenencia de vivienda y lugar de trabajo. A partir de los resultados, concluye que no existen estructuras jerárquicas de decisión significativas, por lo que la elección de una dimensión de alternativas no condiciona la elección de otra.

El atractivo que poseen estos modelos yace en la tractabilidad que presentan al momento de estimar. Esto se debe a sus estructuras de covarianza simples, determinadas por términos de

error que distribuyen Generalized Extreme Value (Ben-Akiva y Lerman, 1985). El cálculo de las probabilidades con estos términos de error permite formulaciones cerradas en la integración, a diferencia de términos de error con distribución Normal, por ejemplo. No obstante, al tratar con combinaciones de conjuntos de alternativas, puede ser necesario simplificar los conjuntos de alternativas. Pues, el número de alternativas a tratar consiste en la combinación de cada dimensión de elección. Esto es particularmente cierto, para elecciones que consideran grandes universos muestrales de alternativas, como lo es la elección de localización residencial y de establecimiento escolar.

Además, el HL es capaz de establecer correlación entre alternativas que comparten una misma dimensión de elección, pero no logra estimar modelos con estructuras de covarianza más complejas. Es importante tener en cuenta estas ventajas y limitaciones al momento de proponer un modelo a estimar para la investigación de elección de localización residencial y establecimiento escolar.

### **2.2.2. Modelos con otras estructuras de error.**

Dadas las limitaciones de los modelos MNL y HL al momento de modelar elecciones multidimensionales, cabe analizar otras posibles aristas de la literatura. En ese sentido, Miller y Lerman (1981) presentan un modelo de elección multidimensional a través de un modelo Probit Multinomial (MP). Este posee la misma estructura que el HL propuesto por Ben-Akiva (1973), con la excepción de que los términos de error distribuyen Normal. Esta diferencia relaja las restricciones con respecto a la matriz de covarianza del modelo. No obstante, debido a que los términos de error con distribución Normal presentan un desafío computacional producto de que la probabilidad a calcular no posee una formulación cerrada, este tipo de modelos se suelen limitar a un conjunto de alternativas restringido.

Han existido otras aproximaciones a la modelación conjunta de elecciones a través de modelos Probit. Paleti, Bhat y Pendyala (2013) utilizan un modelo generalizado que combina elecciones multidimensionales. A grandes rasgos, la metodología plantea diferentes funciones de utilidad, para las alternativas de cada dimensión de elección. Estas, se encuentran vinculadas entre sí a través de términos de error compartidos con distribución

Normal, lo que permite vincular las elecciones a través de atributos no observados. Así, este estudio utiliza este modelo para estimar la elección conjunta de localización residencial, localización del trabajo, posesión vehicular, distancia de viaje, modo y número de etapas de viaje.

Astroza et al. (2019) utilizando una metodología similar, estima en conjunto las elecciones de localización residencial, disponibilidad de vehículo privado y uso de modos compartidos (e.g. *share a ride* o taxi). Además, incorpora un enfoque de clases latentes según las estrategias que caracterizan a cada tomador de decisión (i.e. el orden en que escoge cada alternativa). Estas estrategias se ven determinadas por las elecciones observadas, incorporadas en cada función de utilidad. Por ejemplo, una de las estrategias definidas determina que la localización residencial condiciona la disponibilidad de vehículo, y ambas condicionan el uso de modos compartidos. Así, las funciones de utilidad de disponibilidad vehicular incorporan un parámetro asociado a la alternativa de localización escogida por el individuo. Mientras, las funciones de utilidad de uso de modos compartidos incorporan parámetros asociados a la localización escogida, y a la disponibilidad de vehículo del individuo. Se estima la sensibilidad que las elecciones observadas del individuo tienen sobre la utilidad de la alternativa de elección. Así también, se estiman los términos de error que comparten las funciones de utilidad.

A diferencia de los modelos MNL, HL o MP, en los modelos Probit de Paleti et al. (2013) y Astroza et al. (2019) no se requiere de conjuntos combinados de alternativas. Es decir, cada elección posee funciones de utilidad distintas, que se vinculan a través de términos de error en común. Sin embargo, presentan el desafío de estimar las covarianzas entre las distintas funciones de utilidad. Esto, debido a su estructura Probit, se realiza para un número acotado de alternativas, lo que para universos muestrales de mayor tamaño puede ser alto en costos computacionales. Así también, al momento de trabajar con alternativas no etiquetadas, puede ser difícil interpretar el sentido de estas covarianzas en el modelo.

Cabe mencionar que el enfoque de clases latentes en modelos de elección conjunta se ha presentado en otras investigaciones, además del de Astroza et al. (2019). Por ejemplo, Waddell et al. (2007) caracteriza a dos segmentos de hogares, modelándolos a través de

Logit Mixtos: aquellos que escogen localización residencial condicionado a la elección observada de trabajo, y aquellos que escogen el trabajo, condicionado a la elección observada de localización. El modelo logra captar heterogeneidad en los tomadores de decisión, sin embargo, no estima en conjunto ambas elecciones. Más bien, para cada segmento estima una elección condicionada a la otra elección ya observada. De todas formas, es un avance para determinar para qué tipos de hogares, qué elección condiciona a la otra.

Dadas las limitaciones de estos modelos, resulta atractivo analizar el trabajo propuesto por Pinjari et al. (2011). Este, modela la elección de localización residencial, posesión vehicular, posesión de bicicleta y elección de modo, a través de modelos Logit Multinomiales y Logit Ordinales. A estos, les incorpora términos de error aleatorios con distribución Normal, que logran captar correlaciones entre alternativas de diferentes dimensiones. Estos términos de error pueden ser específicos de la alternativa, como estar asociados a atributos similares entre distintas elecciones (e.g. el tiempo de viaje al trabajo desde la localización residencial, que afecta tanto a la elección de localización como la elección de modo). Esta estructura, permite incorporar correlación entre alternativas, sin perder la tractabilidad que entrega un modelo Logit. La estructura resultante es adecuada a la de un modelo Logit Mixto. Por otro lado, el orden de las decisiones se define de manera similar a lo propuesto por Astroza et al. (2019), incorporando las elecciones observadas del individuo (o atributos asociados a estas) dentro de las funciones de utilidad de las alternativas evaluadas.

Esta modelación resulta atractiva, ya que logra estimar factores no observados que comparten alternativas de diferentes elecciones, aspecto que no logra captar un MNL. Por otra parte, posee la tractabilidad de un modelo Logit, que los modelos Probit descritos por Astroza et al. (2019) o Paleti et al. (2013) no tienen. Además, permite definir una estrategia de decisión, similar a lo que logra capturar un modelo HL, sin la necesidad de definir explícitamente combinaciones de alternativas. No obstante, los factores no observados estimados tienen la restricción que están asociados exclusivamente a las alternativas que los comparten, o en su defecto a los atributos a los que están asociados. También, para universos muestrales de gran tamaño, puede ser difícil determinar estructuras de correlación sofisticadas.

En general, la literatura para modelar en conjunto varias elecciones es extensa, muchas veces siendo vinculada a la elección de localización residencial. Cada modelo presenta ventajas y desventajas, desde los supuestos que los determinan hasta la capacidad de estimar e interpretar los parámetros estimados.

### **2.2.3 Atributos relevantes para la elección de localización residencial**

Un último aspecto a notar, aunque no menor, corresponde a comprender qué atributos son relevantes para la elección de localización residencial. Esto da pie a poder modelar de manera adecuada esta elección con la de establecimiento escolar. Primero, el valor del suelo como atributo de la localización es esencial para determinar la utilidad que le entrega al hogar. Esto es razonable desde la perspectiva de modelos clásicos de localización residencial, como el propuesto por Alonso (1964) que plantea a un consumidor racional que compensa el precio del suelo a consumir con otros atributos que entregan utilidad.

En la misma línea, medidas de acceso a educación, salud, servicios, comercio e industria son determinantes al momento de elegir localización (Hurtubia, Gallay y Bierlaire, 2010). Son una buena medida del nivel oportunidades que presenta la localización residencial con respecto al costo de viajar hacia ellas. Esto, de la mano con la infraestructura pública, como por ejemplo acceso a áreas verdes (Clark et al., 2006, citado en Hurtubia et al., 2010).

De la misma forma, aspectos socioeconómicos tienden a determinar las elecciones de la localización, que pueden estar asociados por ejemplo a la criminalidad del barrio (Weisbrod, Ben-Akiva y Lerman, 1979). Además, otras variables de ingreso medio del barrio (Clark y Onaka, 1983, citado en Hurtubia et al., 2010) también pueden determinar las preferencias por una localización u otra. Esto tiene sentido desde la perspectiva que la localización residencial se vincula con el sentido de pertenencia a la comunidad sobre la que el hogar participa (Savage et al., 2010, citado en Méndez y Gayo, 2019).

Por último, otros atributos relevantes se vinculan a la infraestructura de transporte, que permite un mejor acceso a las oportunidades de la ciudad y entregan utilidad a la elección. Aspectos como presencia de líneas de metro, ciclovías o autopistas pueden entregar valor a

las localizaciones, al igual que generar externalidades negativas que afecten la elección del hogar (Hurtubia et al., 2010). Todos estos, son sólo algunos aspectos relevantes que determinan la elección de localización. Por cierto, serán también factores heterogéneos que pueden variar según el hogar y sus sensibilidades.

### **2.3. Aportes de la investigación a la literatura científica**

El principal aporte que esta investigación busca hacer a la literatura se focaliza en estimar modelos multidimensionales de elección de localización residencial y de elección de establecimiento escolar. La literatura de elección de establecimiento escolar es amplia, y de manera recurrente concluye en que esta elección está vinculada a la localización residencial. Es más, autores como Méndez y Gayo (2019) demuestran cómo ambas decisiones son motores que explican la forma en que la sociedad se reproduce y comporta. Chumacero et al. (2011) por otra parte demuestra cómo la distancia desde la localización residencial cobra vital importancia al momento de escoger colegio. También, autores como Elacqua et al. (2006) y Gómez et al. (2012) demuestran que la elección de establecimiento escolar es heterogénea, cuyas preferencias varían según las características de los hogares y sus preferencias con respecto a atributos académicos como sociales de las alternativas. En fin, la elección de establecimiento escolar se presenta como una elección compleja, estrechamente vinculada con la localización residencial.

Por el otro lado, la localización residencial ha sido modelada en conjunto con un gran número de elecciones: localización de trabajo, elección modal, posesión vehicular, de bicicleta, tenencia de vivienda, entre otras. Así también, diferentes metodologías se han propuesto e implementado (Astroza et al., 2019; Ben-Akiva y Lerman, 1985; Lerman, 1976; Paleti et al., 2013;), demostrando las diferentes estrategias de decisión que pueden existir. Es más, la heterogeneidad en estas decisiones ha sido captada mediante el uso de clases latentes (Astroza et al. 2019; Waddell, 2007).

Sin embargo, a la luz de la revisión realizada y desde el conocimiento del investigador, a la fecha no existen estudios que modelen en conjunto la elección de localización residencial y de elección de establecimiento escolar. Además, los artículos que exploran la elección de

localización residencial con otras elecciones (modo, destino, trabajo), la mayoría trabaja con alternativas etiquetadas. Esto, en un contexto donde la literatura demuestra el fuerte vínculo entre ellas, y en un contexto donde existe amplio desarrollo metodológico de modelos de elección multidimensional.

### 3. METODOLOGÍA

Como se mencionó en el capítulo anterior, existen diversas formas para modelar en conjunto dos o más elecciones. En este capítulo se desglosarán dos modelos. El primero, un Logit Multinomial que pretende ser un punto de partida para el análisis de elección de localización residencial y de establecimiento escolar. Después, se hará una descripción metodológica dos modelos Logit Mixtos, cada uno con una estrategia de decisión diferente: una estrategia que prioriza la elección de localización residencial por sobre la elección de establecimiento escolar, y otra estrategia que prioriza la elección de establecimiento por sobre la elección de localización residencial. Luego, a través de un enfoque de clases latentes se buscará modelar en conjunto ambas estrategias de decisión. Para finalizar, se explicará en detalle la metodología de muestreo de alternativas para poder abarcar el universo de alternativas posibles de ambas elecciones.

#### 3.1 Modelo Logit Multinomial

Este modelo va en línea con la literatura expuesta en la sección anterior. Sea un hogar  $q = \{1, \dots, Q\}$ , que debe elegir entre las alternativas de localización residencial  $r_q = \{1, \dots, R_q\}$  y de establecimiento educacional  $c_q = \{1_q, \dots, C_q\}$ . Dados ambos conjuntos de alternativas, la elección conjunta posee un set combinado de alternativas  $r_q \times c_q$ . La función de utilidad asociada a la elección conjunta de las alternativas  $r$  y  $c$  está dada por:

$$U_{qrc} = \sum_i \theta_i \cdot z_{iqr} + \sum_j \theta_j \cdot z_{jqc} + \sum_k \theta_k \cdot z_{kqrc} + \varepsilon_{qrc} \quad (3.1)$$

Donde  $\theta_i$  corresponde al coeficiente que acompaña al  $i$ -ésimo atributo de la alternativa de localización residencial  $r$  del hogar  $q$ ,  $z_{iqr}$ . Por otra parte,  $\theta_j$  es el parámetro que acompaña al  $j$ -ésimo atributo de la alternativa de establecimiento escolar  $c$  del hogar  $q$ ,  $z_{jqc}$ . Mientras que  $\theta_k$  es el coeficiente que acompaña al  $k$ -ésimo atributo asociado a la combinación de alternativas  $r, c$  (por ejemplo, la distancia de  $r$  a  $c$ ), para el hogar  $q$ ,  $z_{kqrc}$ . Por último,  $\varepsilon_{qrc}$  corresponde a un término de error aleatorio que distribuye i.i.d.  $Gumbel(0, \lambda)$ .

Este modelo de elección discreta es parsimonioso al momento de trabajar en conjunto varias elecciones, dado que su estructura multinomial supone varios aspectos que entregan simpleza al modelo. Uno de estos está asociado a su estructura de error simplificada, donde las alternativas de residencia y establecimiento educacional son independientes entre sí (sólo vinculadas por atributos como la distancia entre ellas). Esto implica, en un contexto de elecciones conjuntas, que este modelo asume simultaneidad en la toma de decisiones, lo que es un supuesto fuerte que merece ser revisado y levantado.

En ese sentido, otra posible arista a investigar es un modelo Logit Jerárquico de elección conjunta. En este, se añade un término de error aleatorio que comparten aquellas funciones de utilidad asociadas a una misma alternativa. Por ejemplo, en un contexto de elección de modo de viaje y horario de viaje, todos los sets de alternativa que contengan el modo auto compartirán un término de error adicional. A través de esta correlación, se puede definir una jerarquía de decisiones de una elección por sobre la otra (Ben-Akiva y Lerman, 1985). Se intentó modelar estrategias de decisión con elecciones de localización residencial donde se agrupan alternativas que comparten la misma alternativa de localización, o la misma alternativa de establecimiento escolar. Sin embargo, los resultados indicaron que las estructuras jerárquicas supuestas eran incorrectas. Por lo tanto, se optó en no profundizar en la explicación del modelo. El Anexo A presenta un diagrama que explica los nidos utilizados, junto con un mayor detalle respecto a las estructuras jerárquicas definidas y estimadas.

### **3.2 Modelo Logit Mixto capturando correlación entre elecciones**

Este modelo propone estimar en conjunto las elecciones de localización residencial y de establecimiento educacional escolar, correlacionando cada decisión a partir de parámetros aleatorios. El modelo considera alternativas no etiquetadas para ambos conjuntos de elección.

Se asume que existen dos estrategias posibles para la elección de localización residencial y de establecimiento educacional. La primera corresponde a la de hogares priorizadores de localización residencial (PLR de ahora en adelante). Estos son hogares que escogen localización residencial, y luego condicionado a esa elección, determinan el establecimiento

escolar. Por el otro lado, la segunda estrategia corresponde a la de hogares priorizadores de establecimiento escolar (PEE de ahora en adelante). Estos hogares escogen establecimiento escolar, y luego en función de esa elección determinan la localización residencial a escoger.

### 3.2.1. Hogares priorizadores de localización residencial (PLR)

Para una estructura de decisión de hogares PLR, la función de utilidad asociada a las alternativas de localización residencial  $r$  para el hogar  $q$  viene dada por:

$$U_{qr} = \sum_i (\theta_i + \theta_{iq} \cdot x_q) \cdot z_{iqr} + \sum_k \gamma_{kq} \cdot z_{kqr} + \varepsilon_{qr} \quad (3.2)$$

Donde el coeficiente  $\theta_i$  capta la sensibilidad con respecto a  $i$ -ésimo atributo de la alternativa de localización  $r$  del hogar  $q$ ,  $z_{iqr}$ . El parámetro  $\theta_{iq}$  capta variaciones sistemáticas en torno al atributo  $z_{iqr}$ , según características  $x_q$  del hogar<sup>1</sup>. Por otro lado,  $\gamma_{kq}$  corresponde a un término de error aleatorio con que distribuye i.i.d Normal(0,  $\sigma_k$ ), donde el parámetro a estimar es la desviación estándar de la distribución. Este término de error busca captar factores no observados del hogar  $q$  con respecto al  $k$ -ésimo atributo  $z_{kqr}$ . La distinción entre  $i$ -ésimos y  $k$ -ésimos atributos, se debe a que el segundo conjunto corresponde a variables presentes tanto en la elección de localización residencial como en la elección de establecimiento educacional. Por ejemplo, atributos socioeconómicos de los hogares que viven en la localización y de los hogares que viven en el entorno del colegio. Como plantea la literatura, es razonable pensar que existen factores no observados asociados a las expectativas de los hogares, con respecto a características socioeconómicas, que determinen ambas elecciones (Méndez y Gayo, 2019). Por último,  $\varepsilon_{qr}$  corresponde a un término aleatorio que distribuye i.i.d. Gumbel(0,  $\mu$ ).

Por otra parte, la función de utilidad asociada a la elección de la alternativa de establecimiento escolar  $c$  para el hogar  $q$  está dada por:

---

<sup>1</sup> Se podrían plantear otras especificaciones que incorporen términos de error aleatorios que capten factores no observados sobre estos atributos. Por parsimonia, las ecuaciones presentes no lo hacen.

$$U_{qc|r^*} = \sum_j (\beta_j + \beta_{jq}x_q) \cdot z_{jqc} + \sum_k \gamma_{kq}z_{kqc} + \phi \cdot d_{qc|r^*} + \varepsilon_{qc} \quad (3.3)$$

Donde  $\beta_j$  es el coeficiente que acompaña al  $j$ -ésimo atributo de la alternativa escolar  $c$  del hogar  $q$ ,  $z_{jqc}$ . Mientras, el parámetro  $\beta_{jq}$  es el que capta variaciones sistemáticas del atributo  $z_{jqc}$ , según características socioeconómicas del hogar  $q$ ,  $x_q$ . Por otra parte, el término de error aleatorio  $\gamma_{kq}$  con distribución i.i.d. Normal  $(0, \sigma_k)$  capta factores no observados del  $k$ -ésimo atributo  $z_{kqc}$ . Como se explicó, este atributo tal que está presente tanto en localización residencial  $r$ , y las alternativas de establecimiento escolar  $c$ . Así, a través del parámetro  $\gamma_{kq}$ , que es igual para ambas funciones de utilidad, se captan factores no observados del atributo que determinan ambas elecciones. Este parámetro es el que busca establecer una correlación entre ambas elecciones, a través de la covarianza existente en los atributos de cada alternativa (Train, 2009).

Además de estos atributos asociados a la alternativa escolar, el parámetro  $\phi$  acompaña al atributo que corresponde a la distancia entre la alternativa de localización residencial escogida y la alternativa de establecimiento escolar evaluada. Este término, es el que determina una dependencia en las elecciones, ya que condiciona la elección de establecimiento a la localización residencial ya escogida por el hogar. Por último, el término de error aleatorio  $\varepsilon_{qc}$  posee una distribución i.i.d. *Gumbel* $(0, \lambda)$ .

Así, para un hogar  $q$  que elige según una estrategia PLR, la probabilidad de escoger una localización residencial  $r$  y un establecimiento escolar  $c$  está dada por:

$$P_q(r, c) = \int \frac{\exp(V_{qr}(\gamma_q))}{\sum_{l=1}^R \exp(V_{ql}(\gamma_q))} \cdot \frac{\exp(V_{qc|r^*}(\gamma_q))}{\sum_{n=1}^C \exp(V_{qn|r^*}(\gamma_q))} \cdot f(\gamma_q) d\gamma_q \quad (3.4)$$

Donde  $V_{qr}(\gamma_q)$  y  $V_{qc|r^*}(\gamma_q)$  corresponde a las utilidades sistemáticas de las funciones de utilidad de localización residencial y establecimiento escolar, respectivamente ( $V_{qr} = U_{qr} - \varepsilon_{qr}$ ). Estas están en función de  $\gamma_q$ , vector que posee los  $k$  términos de error  $\gamma_{kq}$ .  $f(\gamma_q)$  corresponde a la función de densidad de probabilidad del vector  $\gamma_q$ . Cabe mencionar que los

parámetros de escala  $\lambda$  y  $\mu$  acompañan a los coeficientes estimados, pues no son identificables de estos modelos. No obstante, por parsimonia se asumieron iguales a 1, por lo que no aparecen en las ecuaciones.

Así, la log-verosimilitud a maximizar está dada por:

$$l(\theta, \beta, \gamma, \phi) = \sum_{q \in Q} \sum_{r \in R} \sum_{c \in C} g_{qr} \cdot g_{qc} \cdot \ln(P_q(r, c)) \quad (3.5)$$

Donde  $g_{qr}$  y  $g_{qc}$  toman valor 1 si el hogar  $q$  escoge la alternativa  $r$  y la alternativa  $c$ , respectivamente.  $\theta, \beta, \gamma$  y  $\phi$  corresponden a los vectores que contienen los diferentes parámetros a estimar.

### 3.2.2. Hogares priorizadores de establecimiento escolar (PEE)

Para aquel modelo que asume una estrategia de elección PEE, la función de utilidad de una alternativa de establecimiento escolar está dada por:

$$U_{qc} = \sum_j (\beta_j + \beta_{jq} \cdot x_q) \cdot z_{jqc} + \sum_k \gamma_{kq} \cdot z_{kqc} + \varepsilon_{qc} \quad (3.6)$$

Donde los parámetros poseen el mismo significado que los expuestos en la ecuación 3.3. Además, como esta estructura de decisión propone una elección de establecimiento escolar no condicionada a la localización residencial, no posee el atributo de distancia entre la localización y el establecimiento.

Por otra parte, la función de utilidad asociada a la alternativa de localización residencial, condicionada al establecimiento escolar que el hogar  $q$  escogió, viene dada por:

$$U_{qr|c^*} = \sum_i (\theta_i + \theta_{iq} \cdot x_q) \cdot z_{iqr} + \sum_k \gamma_{kq} \cdot z_{kqr} + \phi \cdot d_{qr|c^*} + \varepsilon_{qr} \quad (3.7)$$

Así, los parámetros poseen el mismo significado que los expuestos en la ecuación 3.2. Además, dada que la elección está condicionada a la elección de establecimiento escolar observada, se incorpora el atributo de distancia entre la alternativa de localización residencial y el establecimiento escogido, con su coeficiente respectivo.

De esta manera, la probabilidad de escoger un establecimiento educacional y una localización residencial, para un hogar  $q$  priorizador de establecimiento escolar PEE, está dada por:

$$P_q(c, r) = \int \frac{\exp(V_{qc}(\gamma_q))}{\sum_{n=1}^C \exp(V_{qn}(\gamma_q))} \cdot \frac{\exp(V_{qr|c^*}(\gamma_q))}{\sum_{l=1}^R \exp(V_{ql|c^*}(\gamma_q))} \cdot f(\gamma_q) \cdot d\gamma_q \quad (3.8)$$

Donde  $V_{qc}(\gamma_q)$  y  $V_{qr|c^*}(\gamma_q)$  corresponden a las utilidades sistemáticas de las funciones de utilidad de establecimiento educacional y localización residencial, respectivamente. Estas, al igual que en la sección previa, están en función de  $\gamma_q$ , vector que posee los términos de error aleatorios que capta el modelo.  $f(\gamma_q)$  corresponde a la función de densidad de probabilidad del vector  $\gamma_q$ . Cabe mencionar que los parámetros de escala  $\lambda$  y  $\mu$  acompañan a los coeficientes estimados, pues no son identificables de estos modelos. No obstante, por parsimonia se asumieron iguales a 1, por lo que no aparecen en las ecuaciones.

Así, la log-verosimilitud a maximizar está dada por:

$$l(\theta, \beta, \gamma, \phi) = \sum_{q \in Q} \sum_{r \in R} \sum_{c \in C} g_{qc} \cdot g_{qr} \cdot \ln(P_q(c, r)) \quad (3.9)$$

Donde  $g_{qc}$  y  $g_{qr}$  toman valor 1 si el hogar  $q$  escoge la alternativa  $c$  y la alternativa  $r$ , respectivamente.  $\theta, \beta, \gamma, \phi$  corresponden a los vectores que contienen los diferentes parámetros a estimar.

### 3.3 Enfoque de clases latentes

Las estructuras descritas en las secciones previas asumen que un hogar tomará una estrategia o la otra. Sin embargo, es razonable suponer que algunos hogares optarán por la estrategia de elegir primero la localización residencial, mientras que otros optarán por elegir primero el establecimiento escolar, mientras que otros optarán por estrategias mixtas, representadas como combinaciones de las dos anteriores. Esto puede modelarse mediante un enfoque de clases latentes, el cual pretende incorporar ambas estructuras de decisión, y asigna una probabilidad de pertenencia a cada segmento en función de características del hogar.

Sin pérdida de generalidad, se supone que la función de utilidad latente de un hogar  $q$  para pertenecer a la clase  $m = PEE$  está dada por:

$$U_{qPEE} = \alpha_0 + \sum_{t=1}^T \alpha_{tq} \cdot x_q + \varepsilon_{qPEE} = V_{qPEE} + \varepsilon_{qPEE} \quad (3.10)$$

Donde  $\alpha_0$  corresponde a un parámetro constante de pertenencia a la clase, mientras que  $\alpha_{tq}$  son coeficientes que acompañan a las características  $x_q$  del hogar  $q$ . Estos determinan la sensibilidad que cada característica del hogar tiene sobre la pertenencia a la clase  $PEE$ , con respecto a la clase  $PLR$ .  $V_{qPEE}$  corresponde al componente sistemático de la función de utilidad. Mientras que  $\varepsilon_{qPEE}$  corresponde a un término de error aleatorio con distribución i.i.d.  $Gumbel(0, \varphi)$ . De esta forma, la probabilidad de pertenecer a la clase PEE, y de pertenencia a la clase PLR están dadas por:

$$P_{qPEE} = \frac{\exp(V_{qPEE})}{1 + \exp(V_{qPEE})} \quad (3.11)$$

$$P_{qPLR} = 1 - P_{qPEE} \quad (3.12)$$

Entonces, para un hogar  $q$ , la probabilidad de escoger una localización residencial  $r$  y un establecimiento educacional  $c$  está formulada de la siguiente manera:

$$P_q(r, c) = P_{qPLR} \cdot P_{qPLR}(r, c) + P_{qPEE} \cdot P_{qPEE}(c, r) \quad (3.13)$$

Donde las probabilidades  $P_{qPLR}(r, c)$  y  $P_{qPEE}(c, r)$  corresponden a las descritas por las ecuaciones 3.4 y 3.8, respectivamente. Además, esto significa que cada segmento posee coeficientes distintos para cada atributo. Por lo tanto, la función de log-verosimilitud a maximizar está dada por:

$$l(\theta, \beta, \gamma, \phi, \alpha) = \sum_{q \in Q} \sum_{r \in R} \sum_{c \in C} g_{qc} \cdot g_{qr} \cdot \ln(P_q(r, c)) \quad (3.14)$$

Donde  $\theta, \beta, \gamma, \phi, \alpha$  corresponde a vectores de todos los parámetros a estimar, para ambos segmentos latentes. Para estimarlos se puede realizar simulación de Montecarlo, aproximando la integral necesaria para estimar los parámetros aleatorios de distribución

Normal.  $g_{qc}$  y  $g_{qr}$  toman valor 1 cuando la alternativa  $r$  y  $c$  es escogida por el individuo  $q$ , respectivamente.

### 3.4 Muestreo de alternativas

Como se explicó en la sección 3.2, los conjuntos de alternativas  $R_q$  y  $C_q$  corresponden a alternativas no etiquetadas para cada hogar. Esto se debe a que debido que el universo de alternativas posibles de localización residencial puede ser de gran tamaño (lo mismo ocurre con el universo de alternativas de establecimiento educacional). Para evitar los altos costos computacionales en que se incurriría al considerar todas las alternativas posibles, se opta por realizar un muestreo de alternativas para generar el conjunto que cada hogar  $q$  va a considerar en el modelo. Esto permite que el modelo se pueda estimar a un menor costo, obteniendo estimadores consistentes (McFadden, 1978, citado en Ben-Akiva y Lerman, 1985).

El muestreo de alternativas realizado para esta investigación es el descrito por Ben-Akiva y Lerman (1985). Este, extrae una muestra de alternativas, según la elección que se observe del hogar. De esta manera, puede extraer más alternativas similares a la escogida (que podrían ser más probables a ser escogidas), y menos de otras alternativas que no sean similares (o potencialmente menos probables). Además, corrige según el número de alternativas muestreadas de cada estrato. La formulación, que utiliza la misma notación propuesta por Ben-Akiva y Lerman, se explica a continuación.

Sea un hogar  $q$ , que tiene un conjunto de posible alternativas  $\tilde{J}_q$ . La probabilidad conjunta que este hogar elija una alternativa  $i$ , y posea un subconjunto de alternativas  $D$ , está dada por:

$$\pi_q(i, D) = \pi_q(D|i) \cdot P_q(i), \forall i = 1, \dots, J \quad (3.15)$$

Donde  $\pi_q(D|i)$  corresponde a la probabilidad de tener una muestra de alternativas  $D$ , dado que el hogar elige  $i$ . Mientras  $P_q(i)$  corresponde a la probabilidad de que el hogar escoja la alternativa  $i$ , según un modelo tipo Logit con una función de utilidad sistemática,  $V_{qi}$ . Según

el Teorema de Bayes, la probabilidad condicionada de escoger la alternativa  $i$ , dada una muestra de alternativas es:

$$\pi_q(i|D) = \frac{\pi_q(D|i) \cdot P_q(i)}{\sum_{j \in D} \pi_q(D|j) \cdot P_q(j)}, \forall i = 1, \dots, J : \pi_q(D|j) > 0 \forall j \in D \quad (3.16)$$

McFadden (1978, citado en Ben-Akiva y Lerman, 1985), establece que para obtener estimadores consistentes de esta probabilidad condicionada, es necesario cumplir la condición fijada en la ecuación anterior. De este modo, como  $P_q(i)$  es la probabilidad asociada a un modelo Logit multinomial, la ecuación 3.16 se puede reescribir como:

$$\pi_q(i|D) = \frac{\pi_q(D|i) \cdot \exp(\mu \cdot V_{qi})}{\sum_{j \in D} \pi_q(D|j) \cdot \exp(\mu \cdot V_{qj})}, \forall i = 1, \dots, J \quad (3.17)$$

Así, al incorporar la probabilidad  $\pi_q(D|i)$  dentro del exponente, se obtiene la siguiente probabilidad condicionada:

$$\pi_q(i|D) = \frac{\exp(\mu \cdot V_{qi} + \ln(\pi_q(D|i)))}{\sum_{j \in D} \exp(\mu \cdot V_{qj} + \ln(\pi_q(D|j)))}, \forall i = 1, \dots, J \quad (3.18)$$

Finalmente, la función de log-verosimilitud condicionada a maximizar, que genera estimadores consistentes, está dada por:

$$l = \sum_q \sum_i g_{qi} \cdot \ln(\pi_q(i|D)) \quad (3.19)$$

Donde  $g_{qi}$  toma valor 1 si  $i$  es la alternativa escogida por el hogar  $q$ .

Entonces, para poder estimar de manera apropiada la probabilidad condicionada, es necesario corregirla por una probabilidad de escoger una muestra de alternativas  $D$ , dado que el hogar escogió la alternativa  $i$ . Sean  $J$  el tamaño del universo de alternativas posibles para un hogar. Este, puede dividirse en  $N$  subconjuntos posibles, tal que:

$$\sum_{n=1}^N J_{nq} = J \quad (3.20)$$

Donde  $J_{nq}$  corresponde al universo de alternativas del subconjunto  $n$  para el hogar  $q$ . Luego, sea  $\widetilde{J}_{nq}$  el tamaño de la muestra del estrato  $n \in N$ , mientras que  $n(i)$  corresponde al subconjunto al cual pertenece la alternativa  $i$ . Así, para cada estrato donde no pertenezca  $i$ , se pueden realizar  $\widetilde{J}_{nq}$  extracciones del subconjunto  $n$ , mientras que de  $n(i)$ , se realicen  $\widetilde{J}_{n(i)q} - 1$ . Si se suman todas las alternativas muestreadas, más la alternativa  $i$ , se obtiene una muestra  $D$  de tamaño:

$$\widetilde{J}_q = \sum_n \widetilde{J}_{nq} \quad (3.21)$$

Además, se asumen que para todos los hogares el tamaño de la muestra  $\widetilde{J}_q$  es fijo e igual a  $\widetilde{J}$ . Así, la probabilidad de escoger una muestra  $D$ , condicional a que la alternativa escogida es  $i$ , se puede calcular como:

$$\pi_q(D|i) = \frac{J_{n(i)q}}{\widetilde{J}_{n(i)q}} \cdot \prod_{n=1}^N \left( \frac{J_{nq}}{\widetilde{J}_{nq}} \right)^{-1} \quad (3.22)$$

Es así cómo se puede extraer una muestra, basada en la elección observada del hogar, para estimar parámetros consistentes que logren modelar una elección discreta. El factor  $\pi_q(D|i)$  definido es el que logra corregir según la muestra y las alternativas extraídas. Este se incorpora como un término adicional a la función de utilidad de cada alternativa, aplicado con un logaritmo según lo explicado en la ecuación 3.18.

Para efectos de esta investigación, se realizó para cada individuo un muestreo de alternativas de localización residencial, y un muestreo de alternativas de establecimiento escolar. En la sección 4, se explicará en función de los datos a utilizar, el muestreo realizado para ambas elecciones, el universo de alternativas, tamaños muestrales y factores de corrección aplicados.

#### 4. RECOLECCIÓN Y PROCESAMIENTO DE DATOS

En esta sección se describen los antecedentes recolectados que determinan los datos a utilizar. Esto implica comprender sus fuentes, junto con las decisiones tomadas en torno a procesamiento de información. Para eso, se analizarán tres grupos de información que componen el estudio: la información con respecto a las alternativas de localización residencial, las fuentes que definen las alternativas de establecimientos educacionales escolares y la información con respecto a los hogares y los viajes que realizan por educación escolar. El diagrama de la Figura 4-1 relaciona estos tres componentes de información, indicando las fuentes de información para cada uno, consolidando una base de datos de elección.

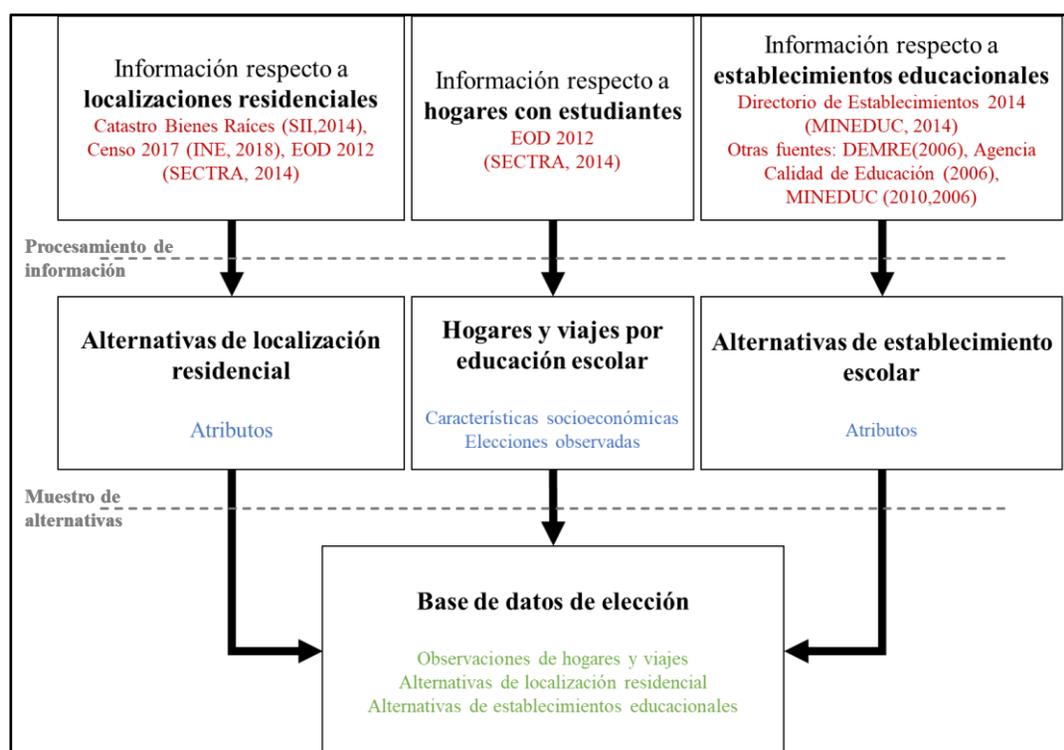


Figura 4-1: Diagrama de consolidación de la base de datos de elección a generar.

La siguiente subsección detalla la recolección de información de los tres grupos de información mencionados. Después, se detalla las metodologías aplicadas para procesar la información recolectada para cada componente y consolidar los universos de alternativas de

localización residencial, establecimientos escolares al igual que las observaciones a utilizar en la muestra. Por último, la tercera sección profundiza en el muestreo de alternativas realizado acorde a los universos de alternativas definidos y la metodología planteada en la Sección 3.4.

## **4.1 Antecedentes recolectados**

### **4.1.1 Hogares que poseen viajes por educación escolar**

Ya que se busca modelar la elección de un hogar sobre su localización residencial y el establecimiento educacional para los estudiantes de este, la EOD 2012 (SECTRA, 2014) se presenta como un antecedente esencial para el estudio. Esta entrega información con respecto a hogares que poseen viajes a establecimientos educacionales escolares, realizado por estudiantes nacidos entre 1994 y 2007.

Cada observación posee características del hogar, del estudiante y del viaje que permiten comprender el comportamiento de estos en la elección. Además, entrega las coordenadas de la localización residencial y del establecimiento educacional, lo que permite utilizar la distancia entre alternativas como variable relevante del proceso de elección. La información entregada por esta encuesta es la que determina las observaciones de elecciones realizadas (tanto residenciales como de establecimiento) a utilizar en el modelo de elección discreta.

### **4.1.2 Atributos de alternativas de localización residencial**

Con el fin de describir apropiadamente las alternativas localización residencial, se obtuvo información de diferentes fuentes, para abarcar variables socioeconómicas y de uso de suelo. Las fuentes utilizadas, que poseen información a nivel de manzana corresponden al Catastro de Bienes Raíces del Servicio de Impuestos Internos (2014) junto con información población del Censo 2017 (INE, 2018). Luego, esta información se procesó a nivel de hexágonos regulares 72,5 m de lado. Así, cada hexágono contenía variables como  $m^2$  de suelo construidos, calidad de las edificaciones, número de viviendas, habitaciones promedio por

viviendas, personas según edad y número de jefes de hogar según el nivel educacional, entre otros.

Además, a partir de la EOD 2012 (SECTRA, 2014), se obtuvo información a nivel zonal con respecto al valor medio de arriendo de vivienda en la Región Metropolitana. Luego, esta información se incorporó a las unidades hexagonales según la pertenencia de estos a cada zona.

#### **4.1.3 Atributos de alternativas de establecimientos escolares**

Para definir el conjunto de alternativas de establecimientos escolares, se recolectó información del Directorio de Establecimientos Educativos 2014 (MINEDUC, 2014). Esta base de datos posee información de todos los establecimientos educacionales por región, según dependencia, niveles de enseñanza impartidos, cobro de mensualidades, orientación religiosa y localización geográfica. Se optó por este año, dado que es el año más cercano al de la EOD 2012 que posee establecimientos georreferenciados. Esto fue de utilidad para utilizar su ubicación con respecto al hogar como atributo relevante de la alternativa de elección. En esta base, se consideraron únicamente establecimientos municipales, particulares subvencionados y particulares pagados, que impartieran al menos enseñanza básica o enseñanza media para jóvenes.

Por sobre esta base, se incorporó información adicional asociada a matrículas de hombres y mujeres, número de docentes y subvenciones (MINEDUC, 2006). También se recolectó información con respecto a resultados SIMCE por establecimiento (Agencia de Calidad de la Educación, 2006) y PSU por establecimiento (DEMRE, 2006). Por último, se incorporaron antecedentes con respecto al índice de vulnerabilidad (IVE) para enseñanza básica y media de cada colegio (MINEDUC, 2010). El año 2006 se definió como año de referencia, para una elección que tienen diferentes temporalidades según el año de ingreso de los estudiantes. En el caso del IVE, el año 2010 es el primer año del cual se tiene registro de esta información. Cada uno de estos antecedentes aportaron variables relevantes, según la revisión de literatura, para modelar la elección de establecimiento escolar.

## **4.2 Procesamiento de la información recolectada**

### **4.2.1 Alternativas de localización residencial**

Tomando en cuenta los antecedentes recolectados, se consolidó la información, procesando las variables y filtrando aquellas observaciones con atributos erróneos o faltantes. Se procedió primero a eliminar todas aquellas observaciones que tuviesen información poco razonable o que no tuviese información. Por ejemplo, hexágonos sin información de usos de suelo, o hexágonos con más jefes de hogar que personas, entre otros.

Después, con la información ya filtrada, se tomó la decisión de agrupar hexágonos, para generar alternativas que abarquen una mayor área del entorno. Así, cada alternativa se conformó de un hexágono, junto con los 36 hexágonos más cercanos a su alrededor. Esta decisión fue tomada dado que así la alternativa puede representar a una localización y sus atributos con respecto a su entorno. Por una parte, el centroide del hexágono central siguió correspondiendo a la localización geográfica de la alternativa, por lo que atributos asociados a la distancia desde la localización se harían desde este punto. La Figura 4-2 ejemplifica una alternativa de localización, considerando los hexágonos descritos y el centroide de la alternativa.



Figura 4-2: Ejemplo de alternativa de localización residencial.

La imagen describe cómo cada alternativa de localización se compone de un centroide, asociado a 1 hexágono, junto con otros 36 hexágonos y sus atributos agregados. Cabe mencionar que, este proceso se repitió para todos los hexágonos. Esto implica que no se redujeron la cantidad de alternativas. Por lo tanto, dos hexágonos colindantes comparten una porción de los atributos. Esto es un aporte positivo, ya que entrega movilidad a las alternativas de elección. Dos localizaciones en un mismo sector de la ciudad pueden tener características muy similares en su entorno, a pesar de que la localización geográfica cambie. Esto es una ventaja en comparación con alternativas zonales fijas.

Aspectos asociados a los  $m^2$  construidos según la calidad de vivienda, al igual que el número de personas, viviendas y jefes de hogar según nivel educacional aprobado, fueron agrupados según la información hexagonal proveniente del catastro de bienes raíces o del Censo 2017. Mientras, información como el valor medio de arriendo de vivienda y distintos indicadores

de accesibilidad de uso de suelo fueron construidos según la información disponible. Este proceso de información se explica a continuación.

a) Valor del arriendo mensual promedio de vivienda

Un atributo relevante no considerado en la información disponible era el valor del m<sup>2</sup> suelo. Esta variable ha demostrado ser significativa en los modelos de elección de localización residencial, por lo que era necesario agregarlo como atributo de las alternativas a considerar.

Para incorporarla, se utilizó el valor del arriendo de la vivienda declarado por los encuestados de la EOD 2012. Este dato tiene dos limitaciones. Por un lado, hay hogares encuestados que no arriendan la vivienda, por lo que declararon el valor que creían sería el valor de arriendo. Esto puede diferir del valor real al cual se arrendaría la vivienda. Segundo, es un valor que no ajusta por el tamaño de la vivienda. Por lo tanto, existe una distorsión debido a que casas de mayor tamaño tendrán una tendencia a vale más. Una medida normalizada puede ser más adecuada para medir el impacto real del valor del suelo.

La segunda limitación no fue tratada, pero la primera sí, con el objetivo de obtener valores de arriendo más confiables en la medida que fuese posible. Primero, se contó el número de hogares por zona EOD que son arrendatarios. Si este número era mayor o igual a 10 observaciones, entonces el valor del arriendo se calculó con el promedio de hogares que arriendan en la zona, ponderado por el factor de expansión respectivos de la EOD. Mientras, si el número de observaciones de hogares que arriendan era menor a 10, no se consideró como representativa la muestra. Por lo tanto, se consideraron todas las observaciones de hogares, arrendatarios o no, en esa zona para el cálculo del valor promedio del arriendo, ponderadas por los factores de expansión. Esta medida se tomó para que, en aquellas zonas con pocas observaciones de arrendatarios, se evitara tener valores poco representativos, pero en la medida de lo posible utilizar sólo la información con respecto al nivel de arriendo real de las viviendas. Por último, el atributo zonal del valor de arriendo promedio se asignó a cada alternativa de localización según la pertenencia del centroide de esta sobre la zona EOD.

b) Accesibilidades a usos de suelo

Otro conjunto de atributos generados para cada alternativa corresponde a las accesibilidades a distintos usos de suelo. Se utilizó el concepto de accesibilidad gravitacional (Hansen, 1959), que propone una medida basada en las oportunidades presentes en un lugar y el costo generalizado de viajar en cierto modo hacia esa localización, como lo explica la ecuación 4.1. La localización tendrá mejor accesibilidad, en la medida que pueda aprovechar más oportunidades en un menor costo.

$$Acc_{modo,op,i} = \sum_j N_{op,j} \cdot \exp(\phi \cdot C_{ij}) \quad (4.1)$$

Los índices  $i, j$  corresponden a la localización de origen y de destino.  $N_{op,j}$  corresponde a la cantidad de cierta oportunidad presente en el destino  $j$  (e.g. m<sup>2</sup> de comercio).  $C_{ij}$  corresponde al costo generalizado de viajar de  $i$  a  $j$  en cierto modo, lo que se interpretó como el tiempo de viaje. Por último,  $\phi$  corresponde a un parámetro de impedancia, asociada a la sensibilidad al costo de viajar.

Para calcular las accesibilidades, se calibró el parámetro de impedancia al costo de viaje. Esta calibración inicial buscó reproducir las distribuciones observadas de los largos de viaje de la EOD según propósito educación, industria, trabajo, salud, de compras y recreación, en modos motorizados y no motorizados. No obstante, y dado que corresponde a un ajuste visual, se hizo un análisis de sensibilidad alterando el valor del parámetro (es decir, se sensibilizó la accesibilidad) al momento de estimar los modelos. Esto, con el fin de maximizar la verosimilitud a estimar. La Figura 4-3 muestra un ejemplo de la distribución del largo de viaje con propósito Trabajo en modo caminata según las observaciones EOD, y la curva de accesibilidad según el parámetro estimado.

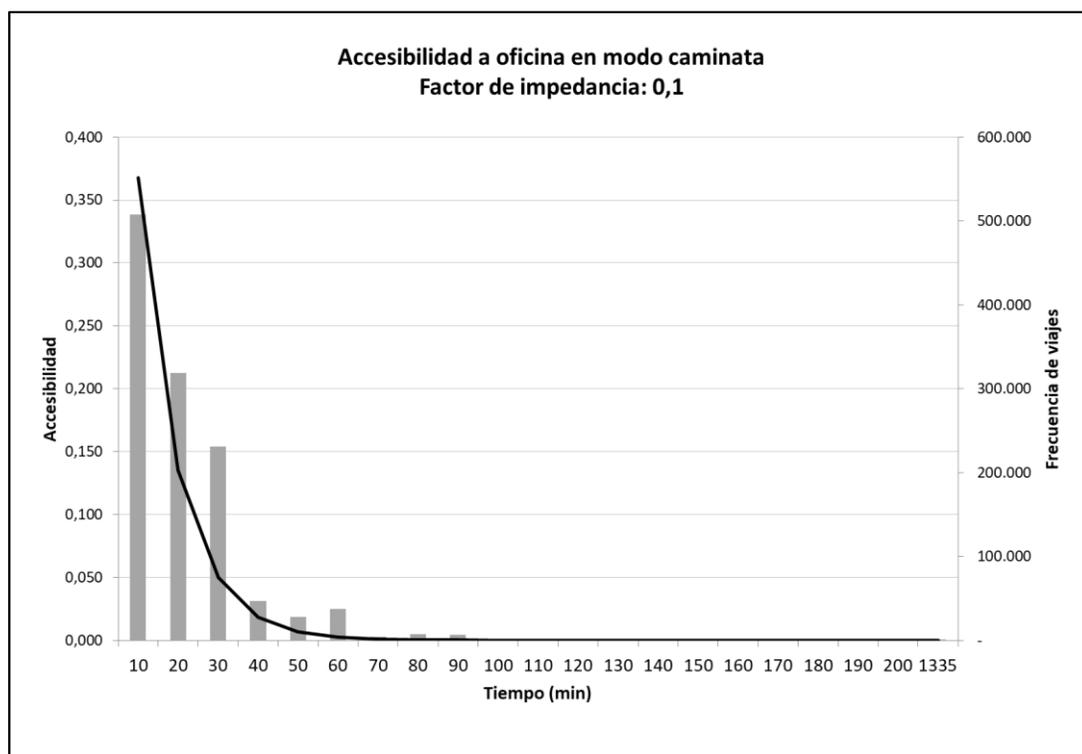


Figura 4-3: Accesibilidad a oficina en modo caminata según distribución de viajes por trabajo de EOD 2012.

Fuente: Elaboración propia en base a EOD 2012 (SECTRA, 2014).

Los parámetros de impedancia calibrado para las distintas oportunidades se muestran en la Tabla 4-1. Cabe mencionar que se consideró el mismo parámetro de impedancia para viajes en modo privado como en transporte público, mientras que se consideró uno distinto para viajes en caminata. Además, los modos considerados corresponden a los modos usados en más del 80% de los viajes diarios para un día laboral normal (SECTRA, 2014). Por lo tanto, se consideró razonable calcular accesibilidades sólo para estos. Se diferenció entre motorizado y caminata, ya que la caminata puede incluir otros aspectos del viaje como el clima, capacidad de caminata y seguridad.

Tabla 4-1: Factores de impedancia iniciales considerados según uso de suelo y modo.

Uso de suelo	Modo motorizado (transporte público o auto)	Caminata
Comercio	-0,15	-0,2
Oficina	-0,05	-0,1
Educación	-0,15	-0,2
Industria	-0,3	-0,3
Salud	-0,05	-0,1
Deporte	-0,1	-0,2

Por el lado del costo generalizado de viaje, se estimó el tiempo de viaje en modo privado, público y caminata. Para el primero, se utilizó la API de Open Route Service ([www.openrouteservice.org](http://www.openrouteservice.org)), que estima el tiempo de viaje en auto desde un punto a otro, utilizando la red vial disponible. Mientras, para el transporte público y la caminata se dispuso del software OpenTripPlanner ([www.opentripplanner.org](http://www.opentripplanner.org)) el cual, para un conjunto de pares origen-destino entregados, logra estimar los tiempos de viaje por modo, y utilizando la infraestructura disponible. En el caso del transporte público, utiliza los itinerarios dispuestos por la autoridad de transporte, estimando tiempo programado de viaje.

Debido al alto costo computacional de calcular tiempos entre pares origen – destino a nivel hexagonal, se optó por considerar los centroides de las zonas EOD para definir los tiempos a estimar por los softwares. Por lo tanto, los m<sup>2</sup> de uso de suelo dispuestos a nivel hexagonal se agregaron también a nivel de zona EOD para lograr tener una medida más realista de las accesibilidades. Así, se calcularon los tiempos de viaje entre cada par de zonas EOD para los modos caminata, transporte público y automóvil.

Con esta información, se procedió a estimar las accesibilidades por zona y modo. Las Figuras 4-4, 4-5 y 4-6 muestran las accesibilidades a oficina, comercio y educación, respectivamente, según modo. Cabe mencionar que la accesibilidad fue normalizada según el valor máximo zonal.

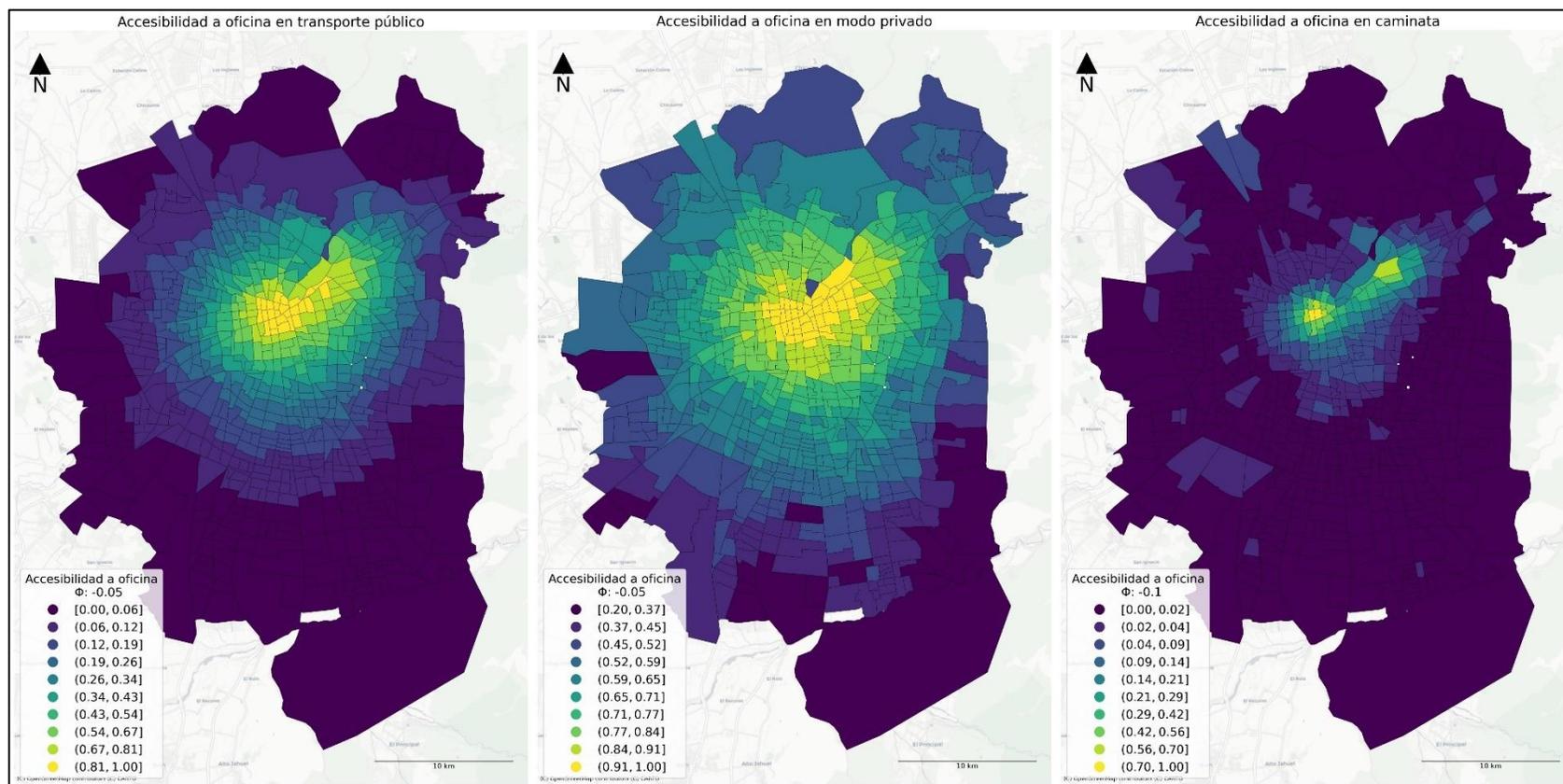


Figura 4-4: Accesibilidad a oficina según modo.

Fuente: Elaboración propia en base a SECTRA (2014), SII (2014).

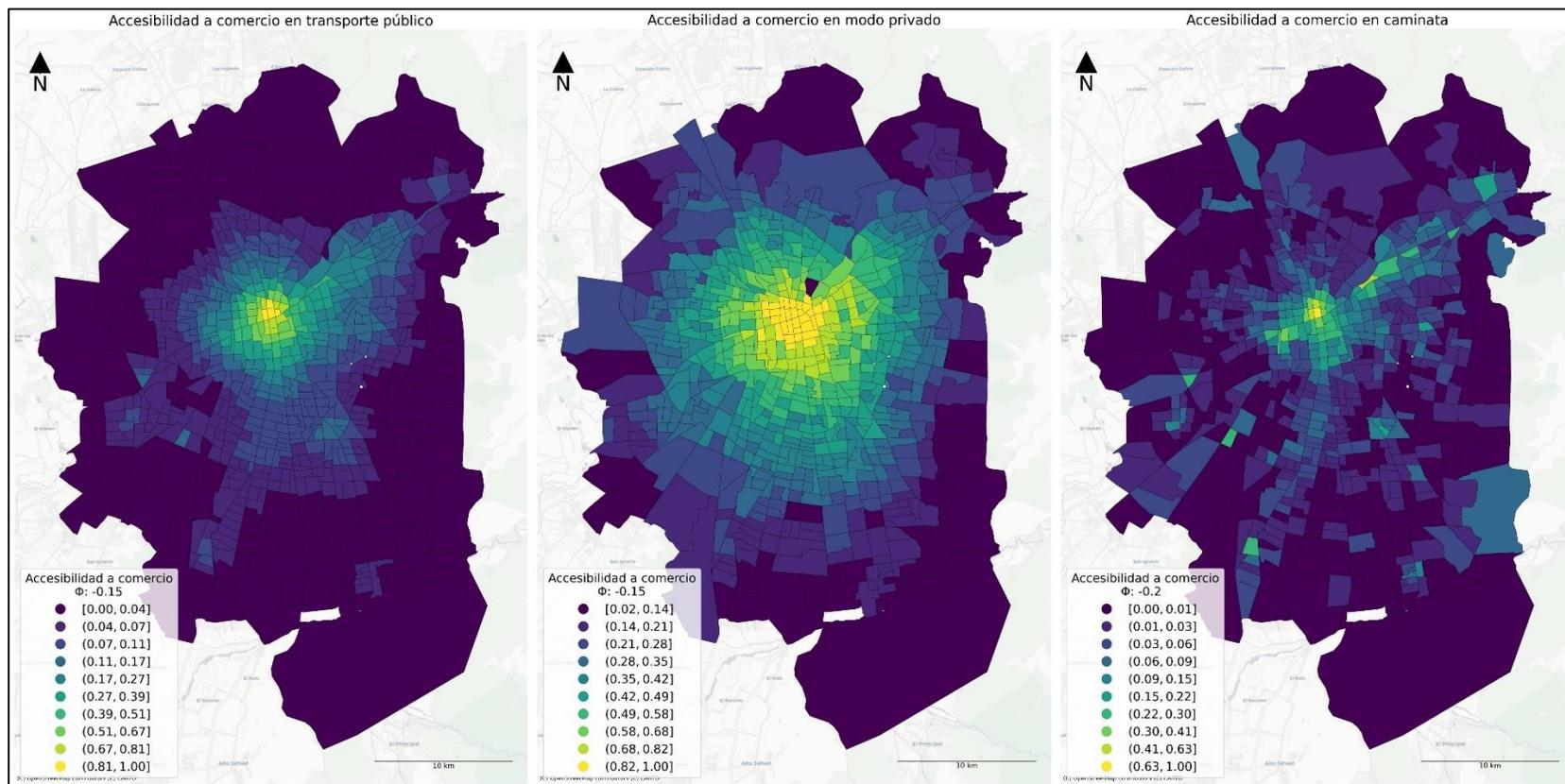


Figura 4-5: Accesibilidad a comercio según modo.

Fuente: Elaboración propia en base a SECTRA (2014), SII (2014).

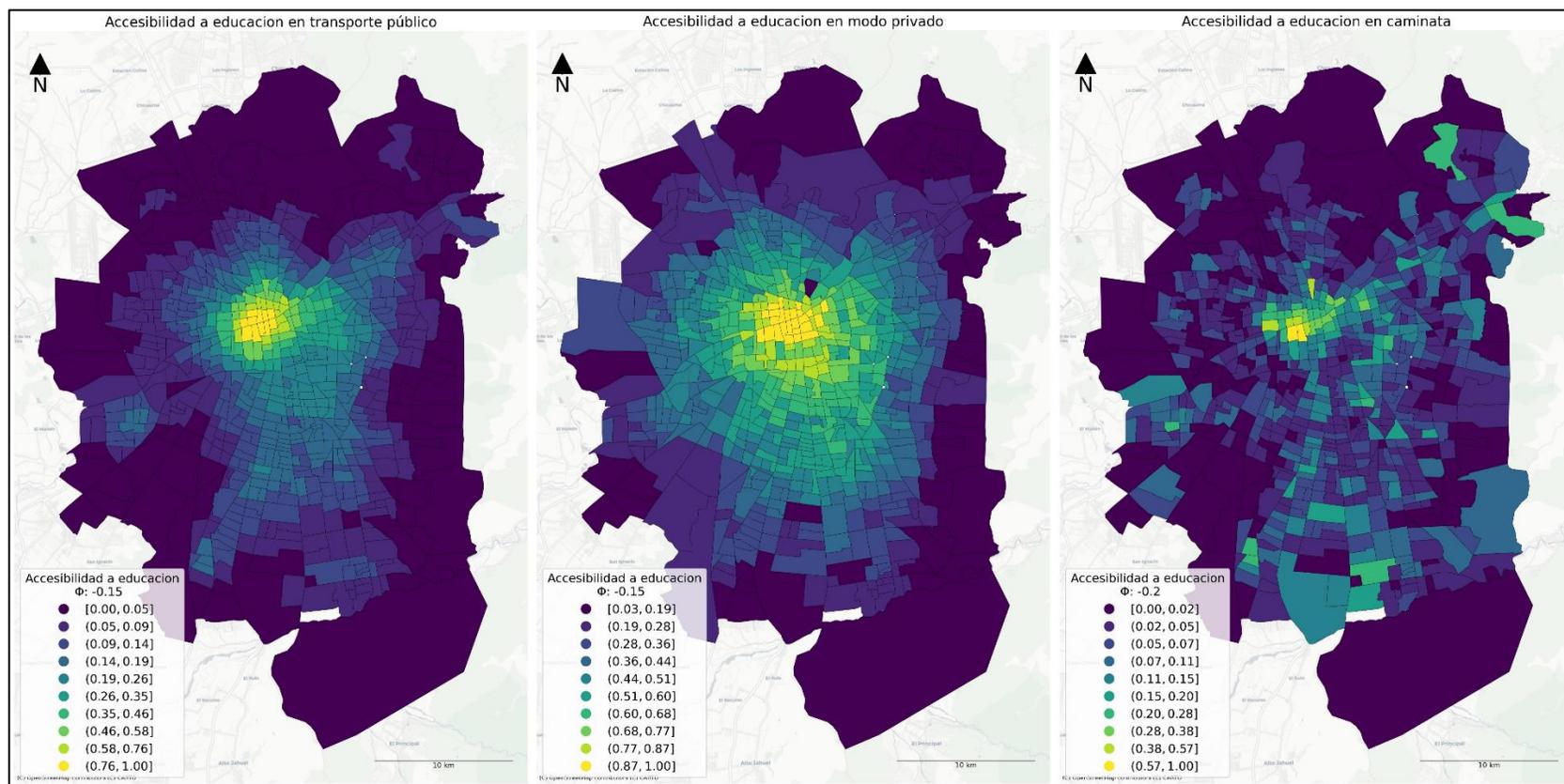


Figura 4-6: Accesibilidad a educación según modo.

Fuente: Elaboración propia en base a SECTRA (2014), SII (2014).

A modo de resumen, la Tabla 4-2 describe todos los atributos considerados para las alternativas de localización residencial, según la información obtenida de las fuentes mencionadas. Se expone el atributo, su unidad, nivel de agregación que posee y la fuente de la que proviene.

Tabla 4-2: Atributos que caracterizan las alternativas de localización residencial.

<b>Atributo</b>	<b>Unidad</b>	<b>Nivel de agregación</b>	<b>Fuente</b>
Valor medio de arriendo de vivienda	\$CLP	Zona EOD	SECTRA, 2014
Accesibilidad* a m2 de comercio	Entre 0 y 1	Zona EOD	SII, 2014
Accesibilidad a m2 de oficinas	Entre 0 y 1	Zona EOD	SII, 2014
Accesibilidad a m2 de salud	Entre 0 y 1	Zona EOD	SII, 2014
Accesibilidad a m2 de educación	Entre 0 y 1	Zona EOD	SII, 2014
Accesibilidad a m2 de industria	Entre 0 y 1	Zona EOD	SII, 2014
Accesibilidad a m2 de deportes	Entre 0 y 1	Zona EOD	SII, 2014
m2 construidos según calidad de vivienda**	m2	37 hexágonos	SII, 2014
Metro cerca***	0 ó 1	37 hexágonos	-
Mall cerca	0 ó 1	37 hexágonos	-
Feria cerca	0 ó 1	37 hexágonos	-
Número de personas	Persona	37 hexágonos	INE, 2018
Número de viviendas	Vivienda	37 hexágonos	INE, 2018
Número de jefes de hogar según nivel educacional aprobado****	Persona	37 hexágonos	INE, 2018

\*Las accesibilidades se calcularon para modos de transporte público, privado y caminata.

\*\*Se desglosa en cinco categorías: inferior, media - inferior, media, media - superior y superior

\*\*\*Cerca hace alusión a si está dentro de los hexágonos que componen la alternativa de localización. Estos atributos se obtuvieron de cruces con información pública georreferenciada.

\*\*\*\*Los niveles educacionales son: Educación diferencial, kínder, educación básica, científico humanista, técnico profesional, superior técnico profesional, superior profesional, magíster y doctorado.

Así, se obtuvo un universo muestral de 48.724 alternativas de localización residencial, cada una con atributos asociados a la información mencionada. Esto corresponde a un 88,19% del

área cubierta por la información recolectada, lo que puede entregar un buen nivel de confianza a nivel muestral.

#### **4.2.2 Alternativas de establecimientos escolares**

La información recolectada da pie a consolidar un conjunto de alternativas de establecimientos educacionales escolares. Primero, se tomó del Directorio de Establecimientos 2014 (MINEDUC, 2014) aquellos establecimientos municipales, particular subvencionados y particulares. Por simpleza, establecimientos con otras dependencias como Corporaciones de Administración Legal, o Servicio Legal de Educación fueron omitidos. De todas formas, el número de establecimientos era mínimo. Segundo, de ese conjunto de establecimientos, se consideraron sólo alternativas que impartieran al menos enseñanza básica o media para jóvenes, tanto científico humanista como técnico profesional. No se consideraron establecimientos como jardines infantiles, escuelas exclusivamente diferenciales, o exclusivamente para adultos. Por último, fueron eliminados aquellos establecimientos sin referencia geográfica, nivel del valor del cobro mensual, dependencia ni orientación religiosa. Así, desde la base de este conjunto de alternativas, se incorporó la información adicional provista por los datos de MINEDUC, DEMRE y Agencia Calidad de Educación.

Se utilizaron los datos del año 2006 en miras de poseer información representativa sobre los atributos escolares. Dado que la encuesta fue realizada el año 2012, implica que los hogares encuestados escogieron el establecimiento en el pasado con información de al menos entre los años 1999 y 2007. Se definió el año 2006 pues se encontraba dentro de este rango, y era el año que más información se tenía disponible. La única información no disponible para ese año fue la referida al Índice de Vulnerabilidad, cuya fecha más cercana era el 2010. La siguiente subsección detalla la incorporación de la información y los procesamientos necesarios.

a) Incorporación de información adicional de establecimientos

A los establecimientos filtrados, se les incorporó información con respecto de las matrículas por sexo, número de docentes, monto de subvenciones, resultados SIMCE y PSU, además del IVE. Para las primeras dos variables, aquellos establecimientos sin matrículas, docentes o sin información respecto de estos, fueron eliminados. En el caso de la subvención, esta sólo aplica a establecimientos municipales o particulares subvencionados. Por lo tanto, a los establecimientos particulares pagados se les imputó una subvención estatal de \$0 CLP. Se eliminaron aquellos establecimientos sin información de subvención escolar, en caso de haber debido tenerla.

Por otra parte, los resultados SIMCE de básica y de media se consideraron como dos atributos diferentes, diferenciando también según prueba de lenguaje y matemática. Se eliminaron establecimientos sin información con respecto a ninguno de los resultados. Para los resultados PSU, se consideró únicamente los resultados promedio del establecimiento para Lenguaje y Matemáticas. Se elaboró también una variable que indica si el colegio rindió la PSU o no. En caso de no haber información, se tomó el supuesto que no había rendido la prueba de selección.

Para finalizar, se incorporó información respecto al Índice de Vulnerabilidad Escolar de los establecimientos. Este índice recupera el porcentaje de estudiantes considerados como prioritarios para recibir el Programa de Alimentación Escolar (PAE), según las características socioeconómicas de estos (BCN, 2020). A mayor número de alumnos prioritarios, mayor será el IVE, lo que se puede interpretar como una variable asociada al nivel socioeconómico de las familias del establecimiento.

El IVE se calcula para la enseñanza básica y para la enseñanza media. Por lo tanto, para incorporar una variable única se procesó la información de la siguiente manera. Si el establecimiento no imparte enseñanza media, se le asignó el IVE para educación básica. Si el establecimiento no imparte enseñanza básica, se le asignó el IVE de educación media. No obstante, si el establecimiento imparte ambos niveles de enseñanza, el indicador calculado corresponde a un promedio simple de ambos indicadores. Esto fue útil para utilizar el IVE

como atributo único asociado a la vulnerabilidad del establecimiento, sin tener que utilizar dos atributos por separado. Cabe mencionar que, a los establecimientos particulares, en caso de no tener información asociada a esta variable, se les asignó un IVE igual a 0. Esto parece un supuesto razonable, dado que los hogares que atienden a estos colegios son de mayor nivel de ingreso.

b) Atributos de localización del establecimiento escolar

Sumado a los atributos del establecimiento, se optó por incorporar información de la localización en la cual se ubica cada establecimiento. Esto podría ser útil para considerar variables de uso de suelo o del nivel socioeconómico del barrio que determinen la elección del hogar. Por ejemplo, que el colegio se encuentre en un barrio de buen nivel socioeconómico, o que tenga buena accesibilidad a servicios u oficina. Para esto, se tomaron las coordenadas de cada establecimiento, y se buscó el hexágono de las alternativas de localización sobre el cual el colegio estuviera. Al establecimiento se le asignó atributos de la alternativa de localización encontrada. Es decir, la alternativa de establecimiento escolar considera también atributos de localización asociados a los 37 hexágonos vinculados con esta.

La Figura 4-7 ejemplifica esto para un establecimiento A y una localización B. Se observa que al establecimiento A, pertenece a la alternativa de localización B. Por ende, se le pueden asignar atributos de esta alternativa. Aquellos establecimientos a los que no se les pudo asignar una localización, fueron eliminados.

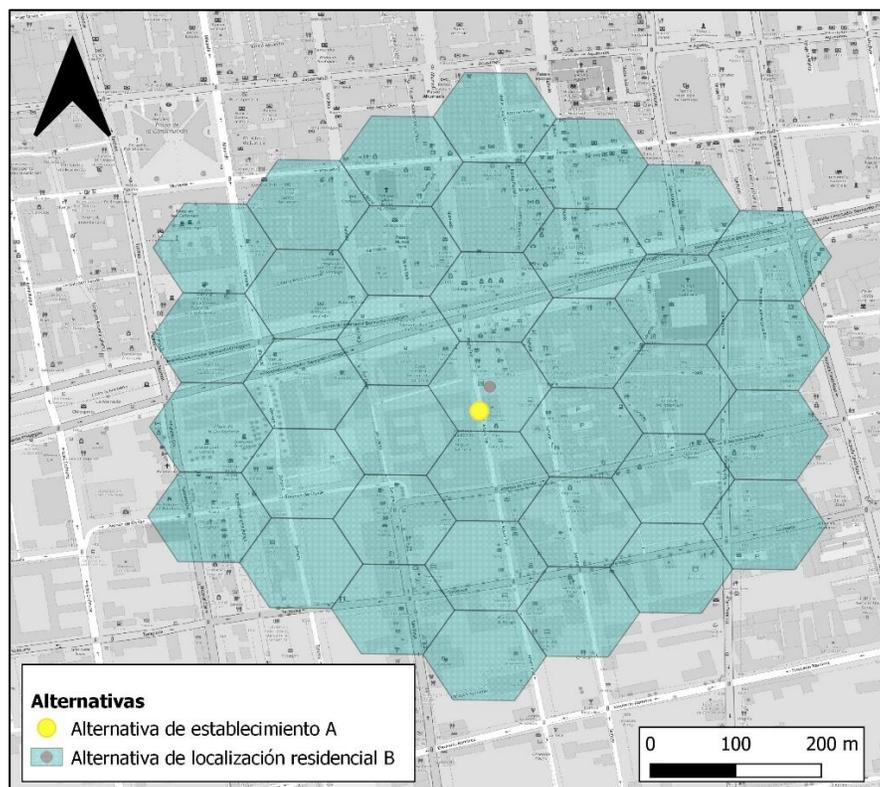


Figura 4-7: Ejemplo de alternativa de establecimiento escolar y su entorno.

A partir de la información procesada para las alternativas de establecimiento educacional, junto con los filtros realizados y variables consideradas, se consolidó un universo muestral de 938 establecimientos educacionales escolares. Esto corresponde a un 53,45% del total posible según la información recolectada. Cada una de estas alternativas posee atributos asociados a la información mencionada en este subcapítulo. La Tabla 4-3 detalla los atributos consolidados, su unidad de medida, nivel de agregación espacial y la fuente de información.

Tabla 4-3: Atributos que caracterizan las alternativas de establecimiento escolar.

<b>Atributo</b>	<b>Unidad</b>	<b>Nivel de agregación</b>	<b>Fuente</b>
Dependencia del establecimiento: Municipal, Particular Subvencionado o Particular	0 ó 1	Establecimiento	MINEDUC, 2014
Rango de pago de mensualidad*	0 ó 1	Establecimiento	MINEDUC, 2014
Códigos de educación impartida**	0 ó 1	Establecimiento	MINEDUC, 2014
Orientación religiosa***	Entre 0 y 1	Establecimiento	MINEDUC, 2014
Matrículas según sexo de estudiante	Estudiantes	Establecimiento	MINEDUC, 2006
Valor de subvención escolar	\$CLP	Establecimiento	MINEDUC, 2006
Número de docentes	Docentes	Establecimiento	MINEDUC, 2006
Índice de Vulnerabilidad	Porcentaje	Establecimiento	MINEDUC, 2010
Puntaje PSU Lenguaje y Matemáticas	0 a 850	Establecimiento	DEMRE, 2006
Resultados SIMCE básica	0 a 100	Establecimiento	Agencia Calidad de Educación, 2006
Resultados SIMCE media	0 a 100	Establecimiento	Agencia Calidad de Educación, 2006
Accesibilidad a m2 de oficinas	Entre 0 y 1	Zona EOD	SII, 2014
Número de jefes de hogar según nivel educacional aprobado****	Persona	37 hexágonos	INE, 2018

\*Los rangos son: Gratuito, \$0-\$10 mil CLP, \$10-\$25mil CLP, \$25-\$50mil CLP, \$50-\$100mil CLP, más de \$100mil CLP

\*\*Estos corresponden a: Párvulo, Básica, Media Científico Humanista y Media Técnico Profesional

\*\*\*Las orientaciones consideradas son: Católica, Laica, Otras

\*\*\*\*Los niveles educacionales son: Educación diferencial, kínder, educación básica, científico humanista, técnico profesional, superior técnico profesional, superior profesional, magíster y doctorado.

### **4.2.3 Hogares y sus elecciones observadas**

Definidos los universos muestrales tanto de las alternativas de localización residencial, como de las de establecimiento escolar, se procedió a definir las observaciones a considerar para la estimación del modelo. Estas, como se mencionó previamente, corresponden a todos aquellos viajes realizados por estudiantes escolares, nacidos entre los años 1994 y 2007, con origen en el hogar y destino en el establecimiento escolar, durante las 6:01 y 9:00 am. Se eliminaron observaciones que indicaran viajes del establecimiento al hogar, al igual que observaciones sin información respecto del origen o del destino del viaje. Además, se consideró el rango de edad entre esos años ya que corresponderían a estudiantes de no más de 18 años, y no menos de 5 años de edad. A través de la información de la EOD 2012 asociada a dónde estudia, se filtró de manera efectiva que los viajes correspondieran a estudiantes escolares o de párvulo en colegios.

El siguiente paso fue definir las alternativas escogidas para la observación, pues son necesarias para poder estimar el modelo. Para esto, se utilizaron las coordenadas entregadas por la información de la encuesta para el origen (asociado a la localización residencial), y para el destino (asociado al establecimiento escolar). Con esa información, se georreferenciaron las alternativas escogidas de la siguiente manera.

#### **a) Georreferencia de localización residencial escogida**

Para localizar la alternativa residencial en la cual el estudiante que viaja vive, se analizó para cada observación el hexágono sobre el cual están las coordenadas. De esta manera, a esa observación, se definió como alternativa escogida el hexágono al que pertenece. Este proceso es sencillo en la medida que se logre coincidir el punto geográfico del hogar y el hexágono al cual pertenece. Aquellas observaciones, y por ende hogares, a los que no se le pudo asignar una alternativa de localización residencial escogida, fueron eliminadas.

b) Georreferencia de establecimiento escolar escogido

A diferencia de la georreferenciación realizada para las alternativas de localización residencial, el proceso de localización de los establecimientos educacionales requiere de más pasos. La información imputada en la EOD con respecto al destino del viaje, no siempre coincide con las localizaciones de las alternativas escolares según la base de datos del MINEDUC (2014). Por una parte, al ser coordenadas geográficas pueden existir variaciones mínimas que generen diferencias. Por otra, a veces las referencias utilizadas en la encuesta hacen alusión a calles o intersecciones cercanas al establecimiento. Esto hizo necesario definir una metodología para georreferenciar los establecimientos educacionales escogidos, para cada observación.

Primero, se toman las coordenadas del destino del viaje, asociadas al establecimiento educacional. En torno a esa localización se define un área de radio 200 m. Se definió así, pues una distancia mayor hacía poco confiable la elección de establecimiento. Si dentro de ese perímetro de 200 m de radio no existía ninguna alternativa de establecimiento, entonces la observación se eliminaba ya que no se podía garantizar una elección de manera confiable. Mientras, si existía sólo 1 establecimiento dentro de este perímetro, este se consideraba como la alternativa escogida. Cerca de un 25% de los viajes no lograron asignar un establecimiento dentro de este perímetro.

Existe la excepción que dentro de este radio de búsqueda se encuentren dos o más establecimientos. En ese caso, se disminuía el radio en un 10%, y se volvía a iterar hasta que hubiese sólo 1 establecimiento por escoger. En caso de que, al reducir el radio, todos los establecimientos dejaran de pertenecer al perímetro, se eliminaba la observación pues la metodología no era capaz de escoger un establecimiento de manera confiable. De todas formas, esto no sucedió en ningún caso. La Figura 4-8 ilustra esta situación, considerando un radio de 200 m, otro de 180 m y dos establecimientos educacionales.

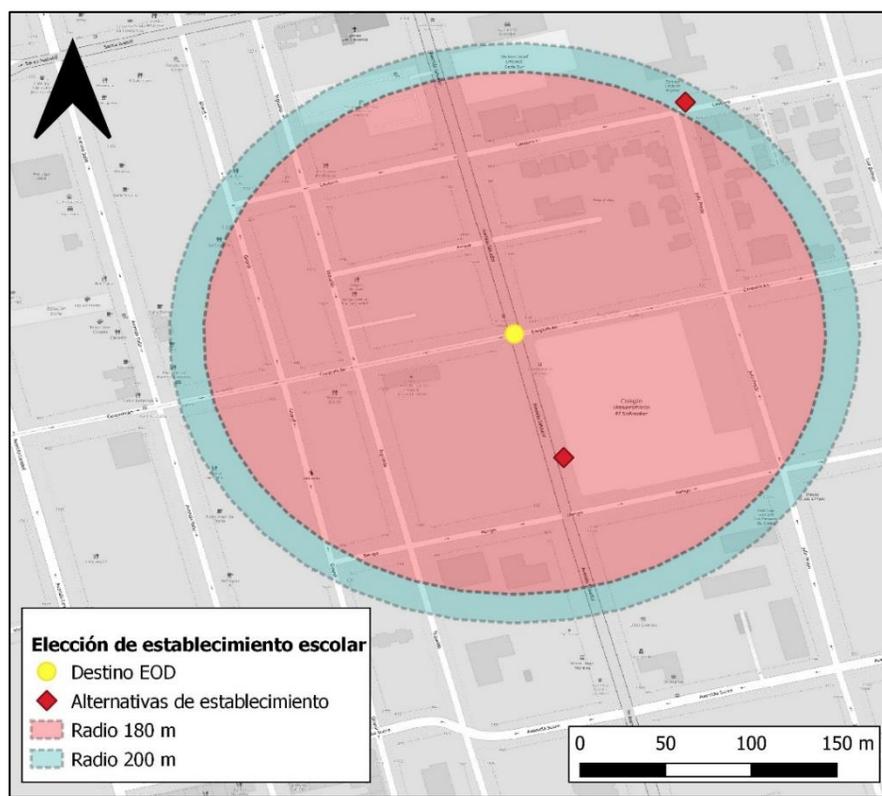


Figura 4-8: Ejemplo de metodología de georreferencia de establecimientos escolares.

Como se observa en la figura, el destino asociado al viaje se encuentra referenciado en una intersección. Al considerar el área de 200 m de radio, existen dos establecimientos. Entonces, al reducir en un 10% el radio de búsqueda sólo queda un establecimiento por considerar. Cabe mencionar que dentro de la metodología se consideraron filtros por sexo del estudiante y la matrícula del establecimiento. Dentro de las posibles alternativas, se consideró si estas aceptan estudiantes del sexo del viajero, según las matrículas del establecimiento. En caso de no aceptar estudiantes de ese sexo, se eliminaba la alternativa y se procedía aplicando la metodología con los colegios restantes hasta obtener una alternativa escogida.

De esta manera, tomando en cuenta todas las observaciones filtradas, sumado a aquellas observaciones eliminadas porque no se les asignó una alternativa de localización o de establecimiento educacional escogida, se obtuvo una base de datos con 2.582 observaciones de viajes de estudiantes al colegio. A nivel de hogares, esto corresponde a 1.981 hogares

diferentes, cada uno con una alternativa de localización residencial y una alternativa de establecimiento educacional observada. El detalle con respecto a las características de los hogares y viajes se realizará en la Sección 5, junto con la descripción del universo muestral de alternativas de localización residencial y de establecimiento escolar.

### **4.3 Muestreo de alternativas y cálculo de distancia entre ellas.**

Teniendo definidas las observaciones de la encuesta EOD a utilizar, junto con ambos universos de alternativas, se procedió a generar los conjuntos de alternativas para cada observación. Esto se realizó mediante la técnica de muestreo de alternativas descrita en la Sección 3.4. Para cada elección, se definieron los estratos de alternativas para generar el muestreo, el número de alternativas a extraer, y se calcularon los factores de corrección respectivos. Las siguientes secciones describen este proceso para ambas elecciones.

#### **4.3.1 Muestreo de alternativas de localización residencial**

Para definir la muestra de alternativas que evalúa cada observación con respecto a la elección de localización residencial, se determinó trabajar con las macrozonas de la EOD. La muestra consta de 20 alternativas para cada observación. La primera, corresponde a la alternativa escogida por el individuo. Luego, en función de la macrozona correspondiente a esa alternativa, se extraen 9 alternativas de las posibles que pertenecen a ese sector. A estas 10 alternativas, se les aplica un factor de corrección  $\ln(\pi(D|i))$  según la ecuación 3.23 explicada en la Sección 3.4. Luego, las siguientes 10 alternativas se extraen del estrato asociado a todas las alternativas que pertenecen a las otras macrozonas diferentes de la de la alternativa escogida. Esto genera 20 alternativas, las primeras 10 con un factor de corrección, las segunda 10 con otro factor de corrección. La Tabla 4-4 detalla, para cada macrozona a la que puede pertenecer la alternativa escogida, los factores de ajuste aplicados a las alternativas.

Tabla 4-4: Factores de corrección para alternativas de localización residencial muestreadas.

Macrozona	N° Alternativas por estrato		Factor de ajuste aplicado	
	Macrozona Escogida	Resto de las Macrozonas	Alternativas 1 - 10	Alternativas 11 - 20
Norte	6.150	42.574	-157,19	-155,26
Poniente	9.937	38.787	-160,58	-159,22
Oriente	9.933	38.791	-160,58	-159,22
Centro	1.676	47.048	-146,47	-143,14
Sur	9.914	38.810	-160,57	-159,20
Sur - Oriente	11.102	37.622	-161,28	-160,06
Ext. Sur - Poniente	12	48.712	-96,84	-885,30

Cabe mencionar que debido a que el universo muestral es de más de 48 mil alternativas, la probabilidad de extraer una muestra de 20 alternativas es muy pequeña. Esto implica que, para cualquier alternativa el factor sea muy negativo (pues es el logaritmo de una probabilidad pequeña). Esto no presenta ser un problema, pues lo relevante son las diferencias relativas entre alternativas. La única situación donde existen diferencias significativas entre el primer y segundo conjunto de la muestra ocurre en el caso que la alternativa escogida pertenezca al sector que se extiende hacia el Sur - Poniente. Sin embargo, en la muestra no hay hogares que vivan en esta macrozona.

#### 4.3.2 Muestreo de alternativas de establecimiento escolar

En el caso del muestreo de alternativas escolares, se aplicaron estratos según la dependencia de la alternativa escogida: municipal, particular subvencionada y particular pagada. Ahora, previo a realizar el muestreo, se determinaron las alternativas disponibles para la observación según 2 características: el sexo del estudiante y la edad del estudiante. Para el caso del sexo del estudiante, se consideraron sólo las alternativas que aceptaran estudiantes matriculados de esa orientación. Por ejemplo, si el estudiante que realiza el viaje es hombre, se consideraron alternativas que sólo aceptaran hombres, o que aceptaran hombres y mujeres.

Mientras, para el filtro respecto a la edad, el conjunto de alternativas posibles se consolidó en función del año de nacimiento, el lugar de estudio y el último nivel de enseñanza aprobado, según la información de la EOD.

Primero, si la observación indica que la persona estudia en un Prekínder o Kínder, se consideran todas las alternativas que al menos tengan educación parvularia. Luego, para las observaciones que indican que el estudiante estudia en escuela, liceo o colegio, la separación se realizó según la edad y el último nivel de enseñanza aprobado. Dado que esta información fue imputada por los encuestados, puede tener errores. Por ejemplo, estudiantes de 7 años que deben cursando enseñanza básica, pero que indiquen ya haberla cursado. Dado que no se puede confiar con certeza en esta información, el procedimiento fue el siguiente. Para aquellos alumnos que indiquen haber cursado enseñanza media, se les consideraron alternativas que al menos cursen educación media. Esto es razonable, pues indicaron en la encuesta estar estudiando en un establecimiento educacional, y la enseñanza media es el mayor nivel posible. Por otro lado, si el estudiante nació entre 1994 y 1998, pero indicó haber cursado enseñanza básica, se consideraron alternativas que al menos enseñaran educación básica o media, ya que no se puede asegurar en qué curso se encuentran. En el caso que la observación dice haber cursado enseñanza básica, pero nació después de 1998, se consideran establecimientos que al menos posean educación básica. Pero, si nació después de 1998 y dice haber cursado enseñanza parvularia, se considerarán colegios que enseñen al menos uno de ambos niveles de educación. Por último, si el estudiante dice estudiar en una escuela, liceo o colegio, y nació el 2007, se le considerarán establecimientos que al menos tengan educación parvularia.

Con los posibles universos de alternativas a considerar, se aplicó el factor de ajuste  $\ln(\pi(D|i))$  según la ecuación 3.23 explicada en detalle en la Sección 3.4. Este factor se realizó según la dependencia del establecimiento escogido, obteniendo una muestra de 20 alternativas. Por ejemplo, si la alternativa escogida corresponde a un establecimiento municipal, las primeras 9 alternativas se extraerán del universo de colegios municipales. Las próximas 5 serán extraídas del universo de establecimientos particulares subvencionados, y

las últimas 5 del universo de establecimientos particulares. La Tabla 4-5 detalla, los factores aplicados por estrato.

Tabla 4-5: Factores de ajuste según alternativa escogida y características de la observación.

Dependencia de alternativa escogida	Características de la observación		Factor de ajuste según dependencia de alternativa muestreada		
	Sexo	Nivel de enseñanza cursada	Municipal	Subvencionado	Particular
Municipal	F	Párvulo	-77,57	-76,20	-77,74
		Párvulo y Básica	-80,82	-79,38	-80,82
		Básica	-80,82	-79,38	-80,82
		Básica y Media	-82,12	-80,64	-82,19
		Media	-66,34	-63,97	-65,14
	M	Párvulo	-77,01	-75,63	-77,15
		Párvulo y Básica	-80,24	-78,77	-80,21
		Básica	-80,24	-78,77	-80,21
		Básica y Media	-81,45	-79,96	-81,49
		Media	-65,68	-63,33	-64,47
Subvencionado	F	Párvulo	-80,36	-80,37	-81,22
		Párvulo y Básica	-83,96	-83,90	-84,66
		Básica	-83,96	-83,90	-84,66
		Básica y Media	-85,43	-85,34	-86,20
		Media	-74,50	-73,52	-74,00
	M	Párvulo	-79,86	-79,86	-80,69
		Párvulo y Básica	-83,48	-83,41	-84,14
		Básica	-83,48	-83,41	-84,14
		Básica y Media	-84,84	-84,73	-85,57
		Media	-73,77	-72,81	-73,26
Particular	F	Párvulo	-72,32	-71,64	-73,88
		Párvulo y Básica	-76,48	-75,73	-77,87
		Básica	-76,48	-75,73	-77,87
		Básica y Media	-77,43	-76,65	-78,89
		Media	-68,39	-66,72	-68,58
	M	Párvulo	-71,90	-71,21	-73,43
		Párvulo y Básica	-76,08	-75,31	-77,43
		Básica	-76,08	-75,31	-77,43
		Básica y Media	-76,90	-76,11	-78,33
		Media	-67,81	-66,15	-67,99

Tal como ocurre para las alternativas de localización residencial, dado que el tamaño universal es grande en comparación con la muestra extraída, las probabilidades asociadas a definir la muestra son bajas, lo que se refleja en factores de corrección negativos. No obstante, el efecto del factor de ajuste se refleja en las diferencias relativas entre estos.

#### **4.3.3 Cálculo de distancia entre alternativas**

Un último aspecto relevante para mencionar fue el cálculo de la distancia entre alternativas de localización residencial y alternativas de establecimiento educacional (tanto observadas como no seleccionadas). Esta variable, como se mencionó en la revisión de literatura, puede ser esencial para las preferencias de los individuos con respecto a las elecciones estudiadas. Para esto, se tomaron los pares de alternativas de localización y establecimiento posibles. Para cada par, se obtuvieron las localizaciones geográficas de cada alternativa y se calculó la distancia euclidiana entre ambos puntos. Es decir, para cada observación conjunta de colegio (20 alternativas) y localización residencial (20 alternativas) se calcularon 400 (20x20) distancias.

Se optó por una distancia euclidiana debido a su simpleza y porque se asocia a una proximidad física entre ambas elecciones. Dada la cantidad de pares a calcular, estimar los tiempos de viaje por modo, en vez de las distancias, resultaba costoso en tiempo y recursos. Esta limitante se puede abordar parcialmente al considerar parámetros distintos según el modo utilizado al incorporar las distancias en las funciones de utilidad de los modelos.

## **5. ANÁLISIS DE LOS DATOS: EL CASO DE SANTIAGO, CHILE**

El objetivo de esta sección es caracterizar a la ciudad que se pone en estudio: Santiago de Chile. El capítulo analiza los tres componentes que determinan la investigación: los patrones de localización residencial, patrones de establecimientos educacionales escolares y el comportamiento de los hogares respecto al viaje con propósito educación. Además, en cada subsección, si corresponde, se analizará la muestra de observaciones a utilizar. Este análisis da paso a comprender variables relevantes que pueden determinar las preferencias de los hogares sobre ambas elecciones consideradas, además de estudiar si la muestra logra captar aquellos atributos de manera correcta.

### **5.1 Patrones de localización residencial**

Varios atributos pueden caracterizar la oferta de localización residencial. Dentro de estos se encuentran: su valor o precio, las características físicas de su entorno (accesibilidad, infraestructura, usos de suelo y actividades) y las características socioeconómicas asociadas al barrio de la localización. El análisis realizado se elabora a partir de la información recolectada y procesada para las alternativas de localización residencial, según lo descrito en la Sección 4. Se consideró razonable analizar el caso de Santiago de esta manera, pues las alternativas cubren un área extensa de la ciudad. Debido a esto, el análisis de la ciudad como caso de estudio se hace a partir de la muestra de alternativas de localización. Cabe mencionar que, dado que el número de alternativas posibles de localización es ligeramente superior a 45 mil, el análisis se presenta a nivel de zonas EOD (SECTRA, 2014).

#### **5.1.1 Valor de arriendo**

La Figura 5-1, muestra los valores promedio por zona. De esta se desprende que hacia el sector oriente existe un mayor valor de la vivienda con respecto a otros sectores de la capital. En las comunas de Vitacura, Lo Barnechea, Las Condes y Providencia es donde se encuentra un mayor valor de la vivienda. Esto se extiende por la cordillera a barrios de las comunas de La Reina, Ñuñoa, Peñalolén y La Florida. Mientras, hacia el sector sur, donde se ubican las

comunas de San Ramón, La Pintana y La Cisterna, se sitúan las viviendas con un menor valor de arriendo. Las zonas de la Región Metropolitana que no se visualizan, se debe a falta de información en el procesamiento de datos.

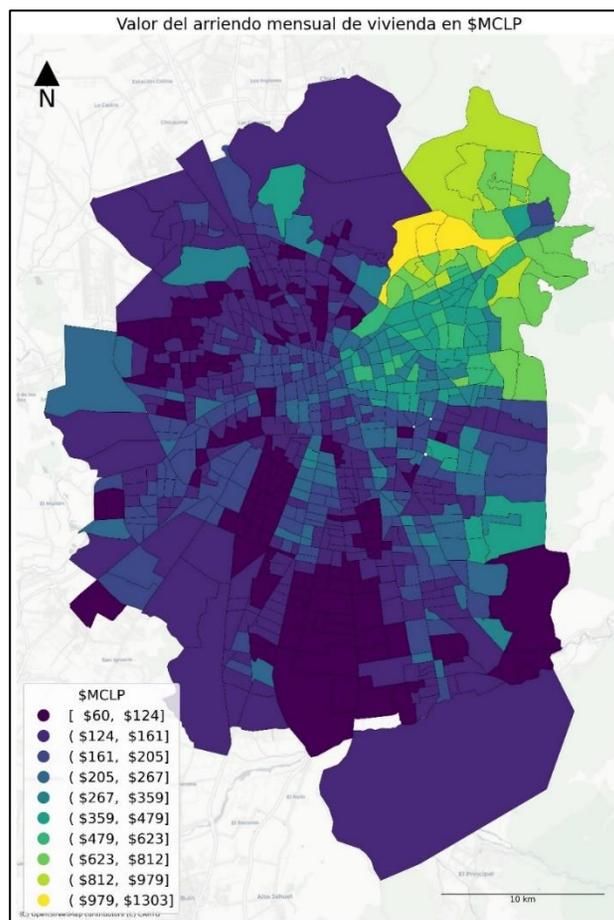


Figura 5-1: Valor promedio mensual de arriendo de vivienda en \$MCLP.

Fuente: Elaboración propia a partir de SECTRA (2014).

### 5.1.2 Accesibilidad

La Figura 5-2 muestra la accesibilidad, calculada según la Sección 4.2.1, en transporte público, a metros cuadrados de los usos de suelo de oficina, comercio y educación.

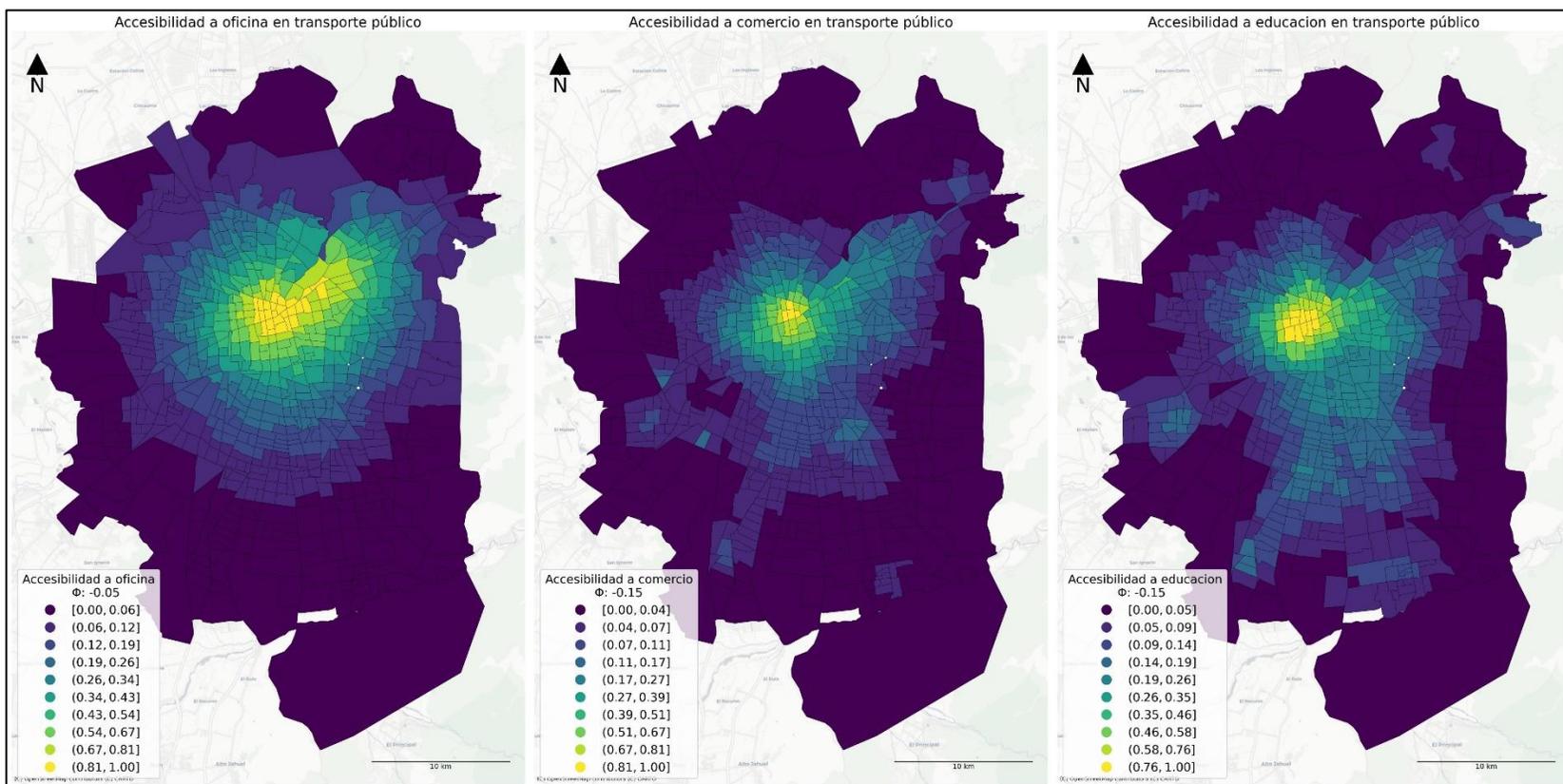


Figura 5-2: Accesibilidad en transporte público a educación, oficina y comercio

Fuente: Elaboración propia en base a SII (2014), SECTRA (2014).

De la figura se desprenden aspectos relevantes de la ciudad de Santiago. Primero, una configuración esencialmente monocéntrica de la ciudad, donde la mayoría de las oficinas y el comercio se aglomeran en el centro. Para las oficinas, existe una tendencia de más oportunidades hacia el sector oriente. El comercio muestra ser accesible en transporte público desde el centro, centro oriente y hacia las periferias de la ciudad, ya sea a través de arterias de la capital, o debido a concentraciones de comercio en subcentros de comunas como Maipú y La Florida. Por otro lado, al observar la accesibilidad a educación, esta se encuentra más dispersa a lo largo de las zonas analizadas. Primero, el centro aún posee buena accesibilidad, lo que puede deberse a adecuada infraestructura de transporte público como a oportunidades cercanas en distancia. No obstante, sectores más distanciados aún logran poseer mediana accesibilidad a educación, lo que se refleja por ejemplo hacia el sector sur de la ciudad en comunas como San Joaquín, Macul, Lo Espejo, La Cisterna, entre otras.

### **5.1.3 Características socioeconómicas de los barrios**

Las alternativas de localización residencial se caracterizan también por aquellos hogares que las integran. En este sentido, conocer la composición sociodemográfica de la oferta de localización residencial puede determinar también las preferencias de los hogares. Uno de estos atributos está asociado al nivel educacional del barrio, el cual está vinculado con el nivel de ingresos de los hogares.

La Figura 5-3 muestra para las zonas analizadas, el porcentaje de jefes de hogar que tienen estudios universitarios (i.e. título universitario o postgrados, sin considerar carreras técnicas profesionales). De esta se desprende que al igual que el valor del arriendo de vivienda, existe una mayor concentración hacia el oriente de zonas con mayor nivel de jefes de hogar con educación universitaria. No obstante, hacia el suroriente, poniente y el centro de la ciudad también existen índices elevados de jefes de hogar con altos niveles educacionales. Mientras, los sectores sur, al igual que norponiente presentan los menores porcentajes de jefes de hogar con título profesional o postgrado. Esta variable puede ser relevante también para comprender desde qué sectores pueden existir viajes en modo privado, pues a mayor nivel

socioeconómico, sectorizado en la zona oriente de la capital, mayor tiende a ser la tasa de motorización (SECTRA, 2014).

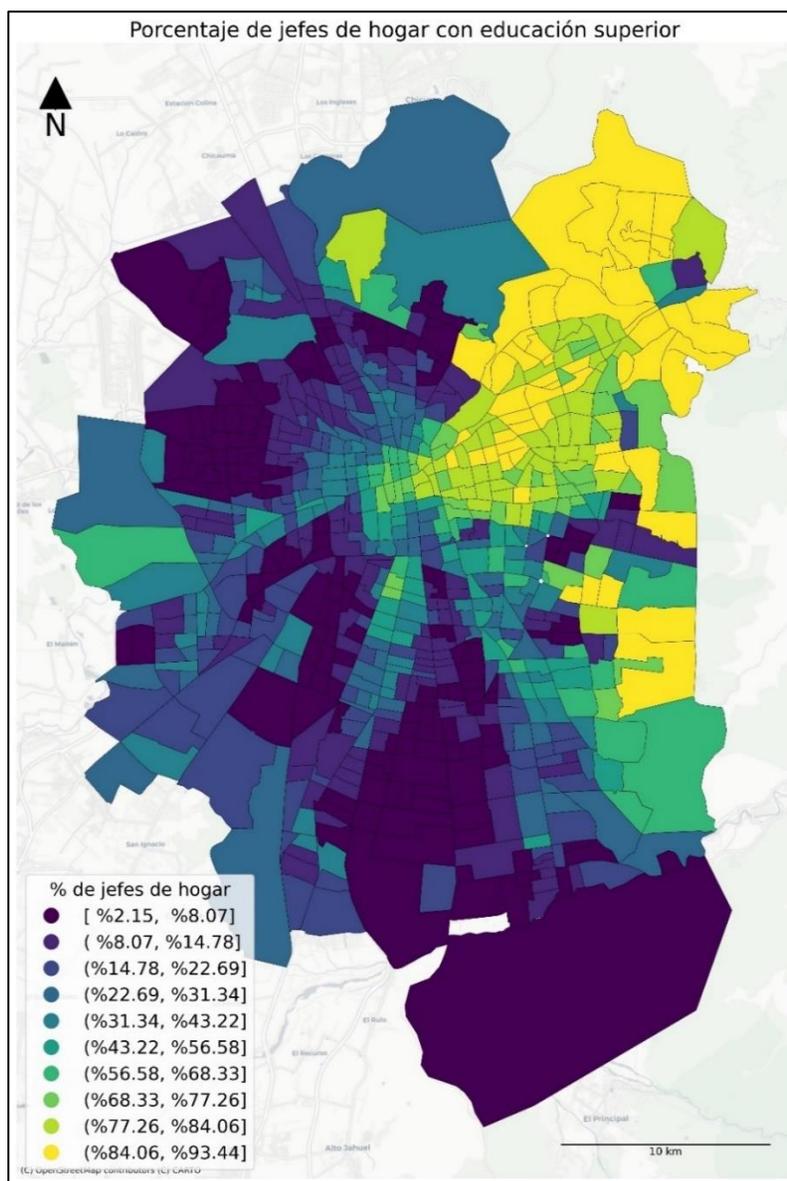


Figura 5-3: Porcentaje de jefes del hogar con educación superior.

Fuente: Elaboración propia en base a INE (2018), SECTRA (2014).

## 5.2 Patrones de alternativas de establecimientos escolares

En las zonas analizadas, comprendidas por las expuestas en la Sección 5.1, se pueden analizar los establecimientos educacionales, sus características y su distribución en la ciudad. A partir del Directorio de Establecimientos 2014 (MINEDUC, 2014), se analizan todos los establecimientos que al menos imparten educación básica o media para niños y jóvenes. Estos establecimientos considerados corresponden a un número mayor al de la muestra, ya que representan información sin procesar, que es útil para comprender en el contexto de la ciudad de Santiago. La Tabla 5-1 hace un resumen de estos establecimientos según dependencia, tipos de educación impartidas y cobro mensual.

Tabla 5-1: Establecimientos educacionales y sus atributos principales.

Fuente: Elaboración propia en base a MINEDUC (2014).

<b>Nro. Establecimientos</b>	1755	100%		1755	100%
<b>Dependencia</b>			<b>Mensualidad</b>		
Municipal	478	27%	Gratuito	705	40%
Subvencionado	1071	61%	\$1.000 - 10.000 CLP	156	9%
Particular	206	12%	\$10.001 - 25.000 CLP	320	18%
<b>Enseñanza</b>			\$25.001 - 50.000 CLP	232	13%
Parvularia	1332	76%	\$75.001- \$100 MCLP	125	7%
Básica	1554	89%	\$100.001CLP o más	193	11%
Científico Humanista	818	47%	Sin información	24	1%
Técnico Profesional	437	25%			

Se observa una mayor proporción de establecimientos particulares subvencionados que municipales, mientras que los establecimientos particulares no superan el 15% del total analizado. Por otro lado, la mayoría de los establecimientos imparten al menos educación básica y parvularia, mientras que la mitad poseen educación media científico humanista. Solo 1 de cada 4 establecimientos imparte educación técnico profesional. Además, 2 de cada 5 establecimientos poseen mensualidad gratuita. La Figura 5-4 presenta esta distribución según dependencia escolar: municipal, particular subvencionada y particular pagada.

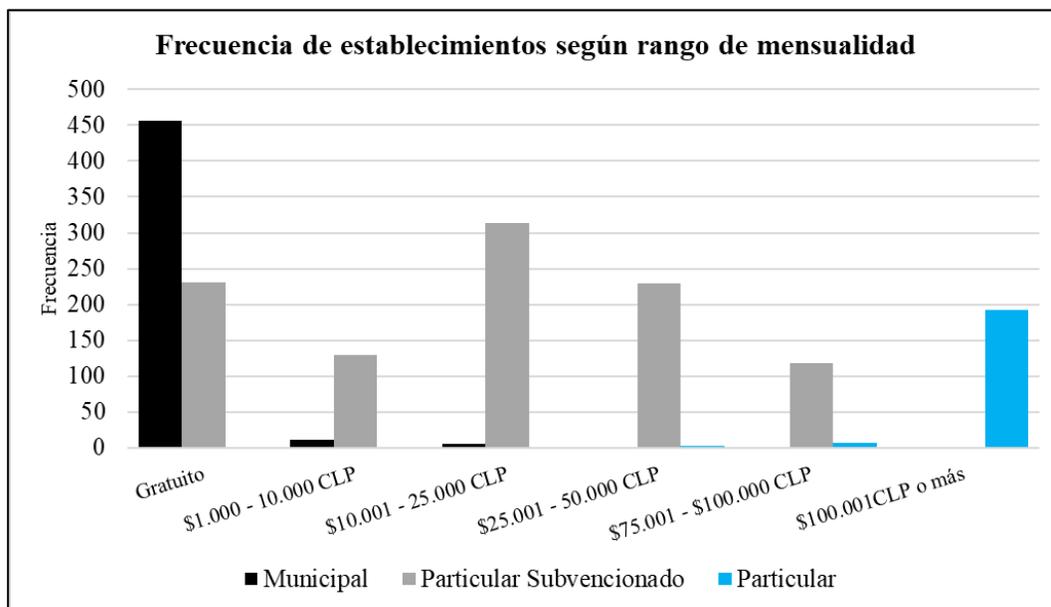


Figura 5-4: Distribución de establecimientos según rango de mensualidad

Fuente: Elaboración propia en base a MINEDUC (2014).

De la figura se observa que el 97% de los establecimientos municipales son gratuitos, y el 95% de los particulares pagados cobran sobre 100 mil pesos mensuales. Los establecimientos particulares subvencionados poseen una mayor distribución donde 2 de cada 3 colegios con esta dependencia cobran entre \$0 y 25.000 CLP, mientras que el otro tercio no supera los \$100 mil CLP.

La Figura 5-5 describe la distribución de establecimientos educacionales según dependencia en la ciudad. En el fondo se presenta el nivel educacional zonal, medido como el porcentaje de jefes de hogar con educación universitaria o superior.

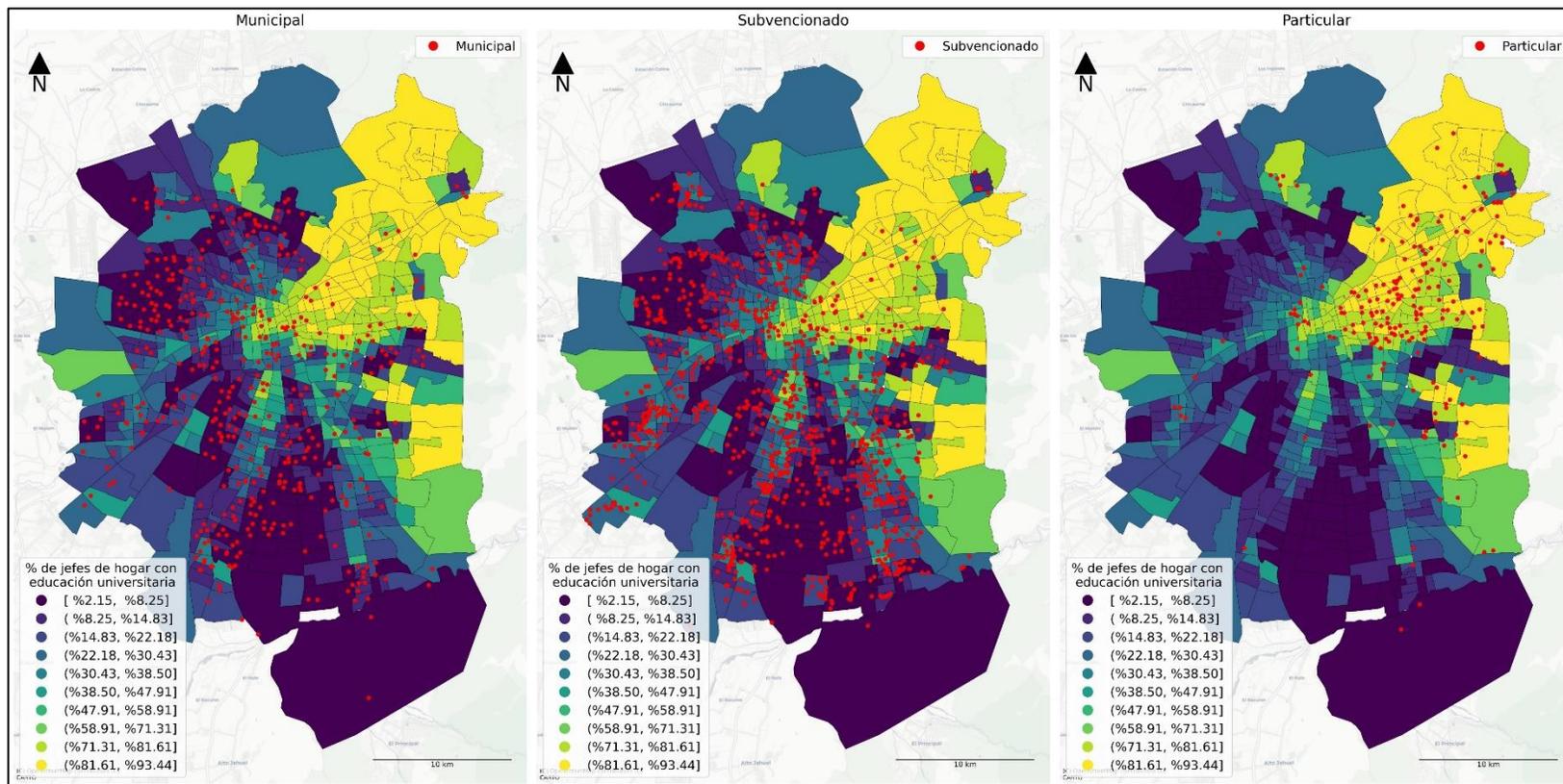


Figura 5-5: Distribución de establecimientos educacionales según dependencia.

Fuente: Elaboración propia en base a SECTRA (2014), INE (2018).

En la figura se aprecian diferencias con respecto a la distribución de los establecimientos en la ciudad. De manera evidente, los establecimientos particulares tienden a posicionarse hacia sectores de nivel socioeconómico alto (asociado al nivel educacional de los jefes de hogar). Esto tiene sentido ya que al no tener subvención estatal y ser de mayor costo<sup>2</sup>, solo son asequibles para hogares de ingresos altos. Por otra parte, los establecimientos municipales y particular subvencionado tienden a distribuirse sobre el resto de las zonas, en un patrón que pareciera ser espacialmente complementario, pero que evidencia la segregación socioespacial de la oferta educacional en Santiago. Además, se aprecia cómo los establecimientos subvencionados tienen la tendencia a aglomerarse sobre ciertos sectores. Estos sectores son cercanos arterias viales de la ciudad, o subcentros dentro de estas. Esta distribución de los diferentes establecimientos puede ser esencial para la comprensión de las preferencias de los hogares sobre esta elección, debido a aspectos como el nivel socioeconómico del barrio y la distancia al hogar.

### **5.2.1 Análisis muestral de alternativas de establecimientos escolares**

De los posibles establecimientos, se consideró un universo muestral de 938 establecimientos educacionales<sup>3</sup>. Es decir, un 53,45 % de la información inicial dispuesta. La Figura 5-6 grafica los establecimientos que sí se consideraron, y cuales no fueron tomados en cuenta debido a falta de información relevante para el estudio. La imagen logra reflejar buena cobertura espacial de los establecimientos, para las tres dependencias escolares.

---

<sup>2</sup> La Tabla 5-2 muestra cómo más del 90% de los establecimientos escolares (de la muestra) poseen mensualidades por sobre los 100 mil pesos.

<sup>3</sup> Este valor se obtiene del procesamiento de la información descrito en la Sección 4.

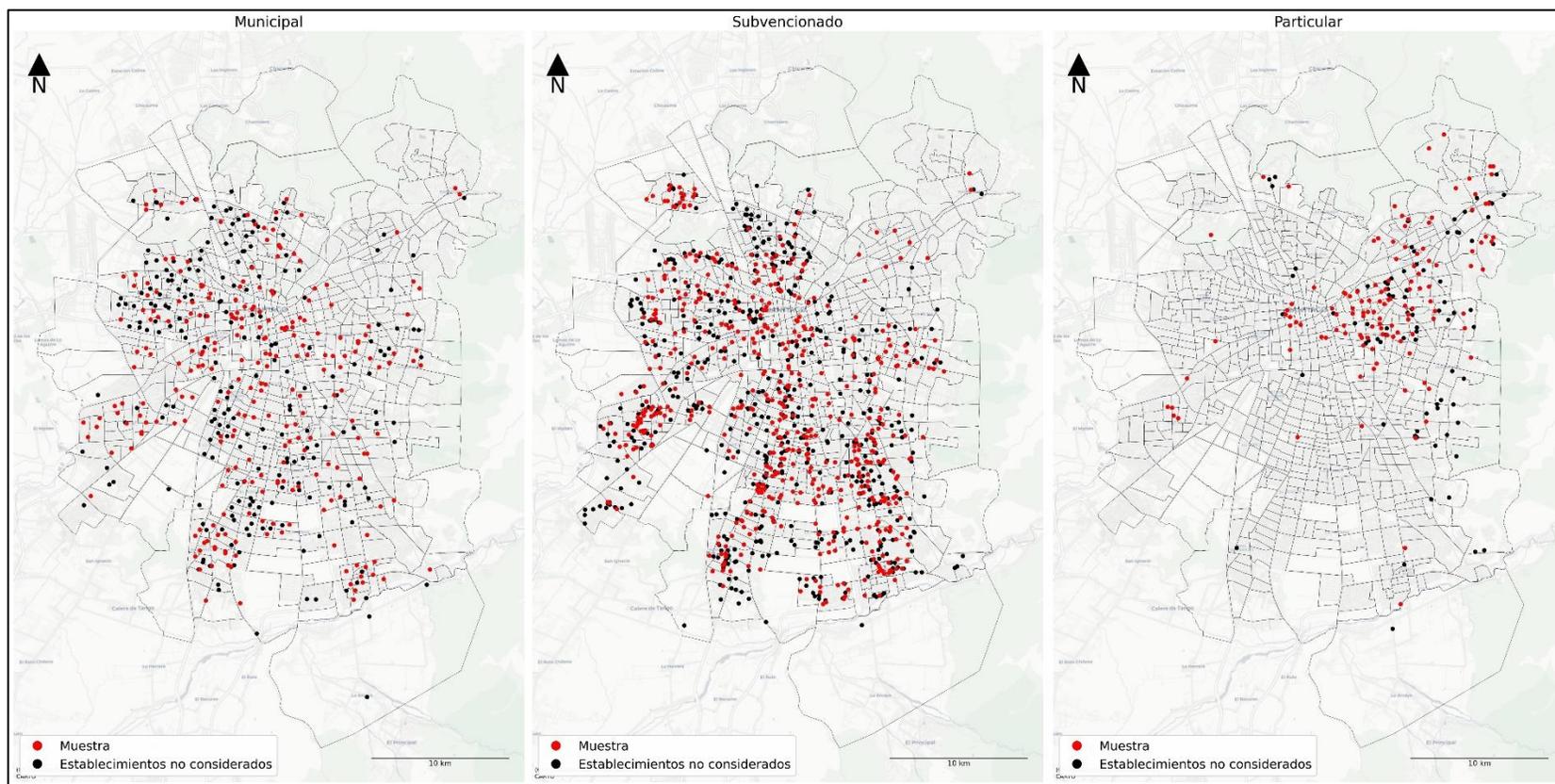


Figura 5-6: Universo muestral de alternativas escolares según dependencia.

Fuente: Elaboración propia en base a SECTRA (2014), MINEDUC (2014).

La Tabla 5-2 detalla el número de establecimientos educacionales por dependencia, y nivel de pago y otros atributos interesantes de las alternativas.

Tabla 5-2: Descripción del universo muestral de alternativas de establecimiento escolar.

<b>Dependencia</b>	<b>Municipal</b>		<b>Particular Subvencionado</b>		<b>Particular</b>	
<b>Variable</b>	<b>Cantidad</b>	<b>%</b>	<b>Cantidad</b>	<b>%</b>	<b>Cantidad</b>	<b>%</b>
<b>Establecimientos educacionales</b>	255	27%	560	60%	123	13%
<b>Educación impartida</b>						
Parvularia	209	82%	414	74%	93	76%
Básica	236	93%	498	89%	122	99%
Media	69	27%	353	63%	112	91%
Científico - Humanista	60	24%	313	56%	111	90%
Técnico Profesional u otras	27	11%	127	23%	1	1%
<b>Mensualidad</b>						
Gratuito	244	96%	82	15%	0	0%
\$1.000 - \$10.000 CLP	8	3%	71	13%	0	0%
\$10.001 - \$25.001 CLP	3	1%	182	33%	1	1%
\$25.001 - \$50.000 CLP	0	0%	122	22%	0	0%
\$50.001 - \$100.000 CLP	0	0%	81	14%	6	5%
\$100.001 CLP o más	0	0%	0	0%	116	94%
<b>Orientación religiosa</b>						
Laico	223	87%	322	58%	61	50%
Católico	28	11%	210	38%	58	47%
Otras religiones	4	2%	28	5%	4	3%
<b>Acepta hombres y mujeres</b>	228	89%	509	91%	105	85%
<b>Índice de Vulnerabilidad</b>						
0%	0	0%	5	1%	122	99%
]0%, 20%]	0	0%	1	0%	1	1%
]20%, 40%]	7	3%	75	13%	0	0%
]40%, 60%]	43	17%	213	38%	0	0%
]60%, 80%]	149	58%	232	41%	0	0%
]80%, 100%]	56	22%	34	6%	0	0%

De esta se desprenden aspectos relevantes sobre los establecimientos. Por un lado, se observa cómo los colegios particulares subvencionados corresponden a la mayor parte de la muestra.

El tipo de educación es bien similar entre establecimientos, con la excepción que los particulares ofrecen casi siempre educación media científico – humanista, con solo 1 establecimiento que imparte técnico profesional. En cuanto a rangos de mensualidad se observa cómo el valor de los precios crece desde los colegios municipales hacia los privados, por lo que se comporta de manera similar al universo descrito en la sección previa. La mayor cantidad de establecimientos católicos son privados, aunque la mayor porción de los establecimientos acepta hombres y mujeres, sin importar el tipo de dependencia. Por último, la vulnerabilidad en establecimientos municipales es siempre mayor a un 20% para la muestra, siendo similar el comportamiento para los particulares subvencionados, aunque con mayor concentración hacia el intervalo de 40%-60%. Mientras que los establecimientos particulares tienen casi en su totalidad estudiantes sin prioridad.

### 5.3 Hogares y viajes con propósito educación escolar

Según la EOD 2012 (SECTRA, 2014), existen más de 927 mil viajes al colegio o escuela entre las 6:01 y las 9:00 am, realizados por estudiantes nacidos entre 1994 y 2007, pertenecientes a 390 mil hogares. Esta información proviene de un total de 3.671 hogares encuestados, compuestos en conjunto por 5171 estudiantes. La Tabla 5-3 describe los cinco quintiles de ingreso per cápita de hogares, junto con los tamaños de hogar promedio y la posesión vehicular y el número de viajes medio por hogar.

Tabla 5-3: Caracterización de hogares con estudiantes escolares según quintil de ingreso.

Fuente: Elaboración propia en base a SECTRA (2014).

Quintil	Ingreso per cápita	Tamaño de hogar	Número de estudiantes escolares por hogar	Tasa de Motorización
Q1	[0; \$73.333]	4,8	1,6	0,2
Q2	(\$73.333; \$108.000]	4,8	1,4	0,4
Q3	(\$108.000; \$151.920]	4,8	1,4	0,5
Q4	(\$151.920; \$228.742]	4,6	1,4	0,7
Q5	> \$228.742	4,4	1,4	1,4
<b>Nº de hogares</b>	<b>390.309</b>	<b>4,7</b>	<b>1,5</b>	<b>0,7</b>

Los tamaños de los hogares en la muestra son mayores al tamaño promedio del hogar en Santiago, que es de 3,24 personas por hogar (SECTRA, 2014), lo que es razonable pues el tamaño mínimo considerando 1 jefe de hogar y 1 estudiante es de 2 personas. Hacia los quintiles de menor ingreso hay una tendencia a tener más estudiantes escolares que quintiles más altos. La tasa de motorización se duplica entre el 20% más adinerado y cuarto quintil, mientras que se triplica con respecto al tercer quintil.

La Figura 5-7 representa la partición modal de viajes con propósito educación escolar, según la pertenencia del hogar de la persona a cada quintil de ingreso. Los viajes no motorizados corresponden a un 40% del total con propósito educación escolar. El transporte público y privado corresponden a cerca de un 20% cada uno. Además, el modo “Otro” forma cerca de un 20% del total de los viajes. Esto último se debe a que esta categoría puede corresponder a transporte escolar<sup>4</sup> (SECTRA, 2014).

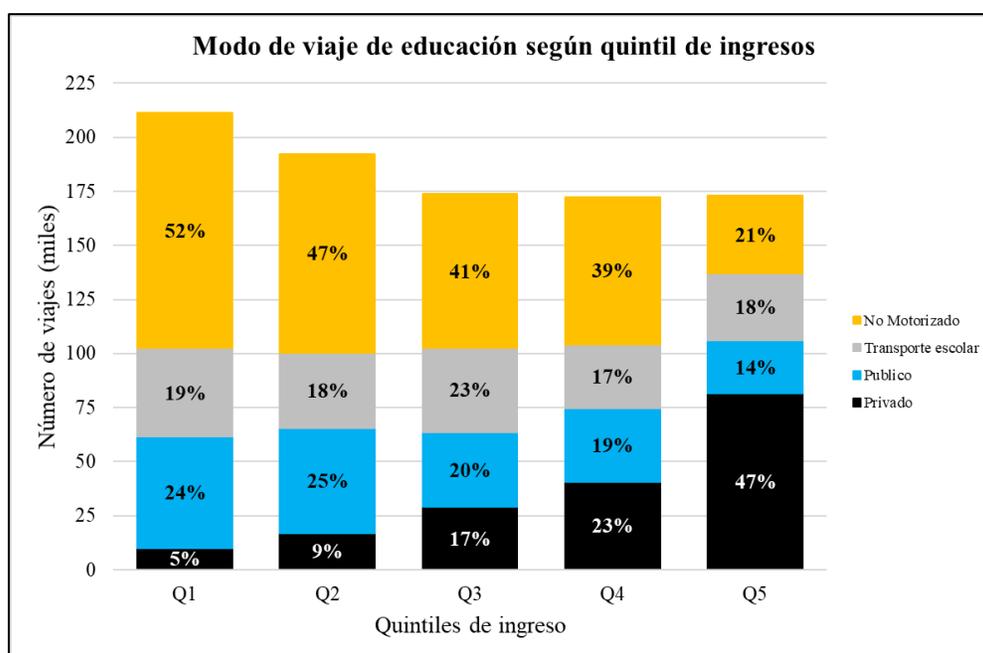


Figura 5-7: Partición modal de viajes escolares entre las 6:01 y 9:00 am.

Fuente: Elaboración propia en base a SECTRA (2014).

<sup>4</sup> Corresponden a buses contratados directamente por los apoderados (o el establecimiento) para movilizar a los estudiantes entre sus casas y los colegios. Otros modos considerados son: moto, tren o buses institucionales. Por simpleza, se utilizará “Transporte escolar” para referirse al modo.

Se observa que, a menor el quintil de ingreso, mayor el número de viajes. Esto es razonable desde la perspectiva que los hogares de menores ingresos tienden a tener mayor número de estudiantes escolares. Además, hacia los quintiles más bajos de ingreso, mayor parte de los usuarios tienden a movilizarse en modos no motorizados al establecimiento educacional. A medida que aumenta el nivel de ingreso, esta tendencia disminuye, lo que podría estar asociado a una mayor realización del viaje en auto. Similar es el caso para los viajes en transporte público. Mientras que, para los viajes en modo privado, el comportamiento es opuesto y posee una relación similar a las tasas de motorización de cada quintil. El furgón escolar se mantiene para todos los niveles de ingreso cerca del 20% del total de los viajes.

Para analizar el tiempo de viaje hacia el establecimiento educacional se consideró la variable de tiempo imputada por los encuestados de la EOD2012, y se analizaron las frecuencias de viajes según intervalos de tiempo. La Figura 5-8 muestra la distribución de la duración de los viajes para cada quintil.

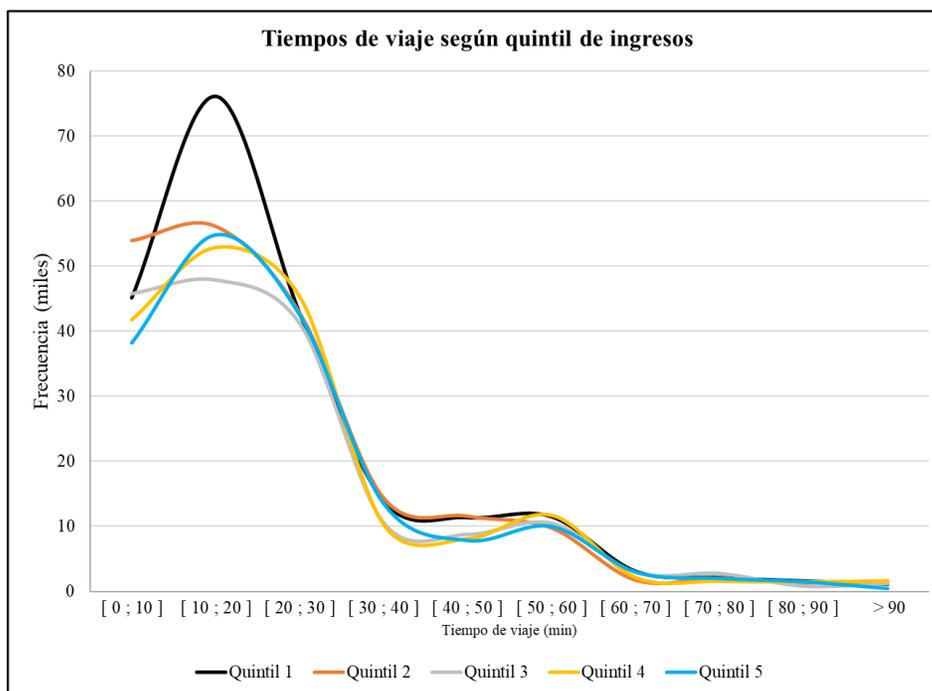


Figura 5-8: Distribución de los tiempos de viaje según quintil de ingreso.

Fuente: Elaboración propia en base a SECTRA (2014).

Se aprecia cómo, a pesar de ciertas diferencias, los hogares tienen una distribución de tiempos de viaje similar. Existe una concentración de viajes de no más de 30 minutos en los cinco quintiles. Por parte del primer quintil, es razonable que sus viajes sean menores, debido a que el 52% de estos son en modos no motorizados. Distinta es la situación de quintiles de mayor ingreso que, a pesar de realizar viajes principalmente cortos, utiliza más modos motorizados. Esto podría indicar una menor sensibilidad a la distancia al establecimiento para hogares con mayores ingresos.

La Figura 5-9 pretende acompañar este análisis, observando la distribución de las distancias euclidianas entre la localización residencial del hogar y el establecimiento educacional. La información es la entregada por la EOD2012.

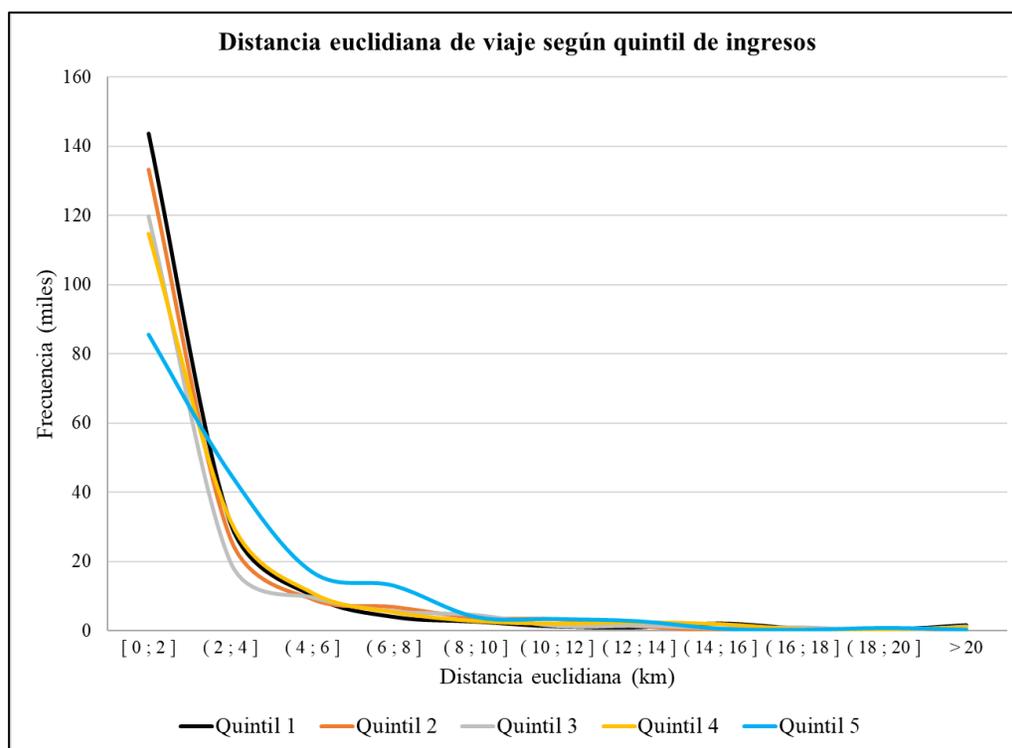


Figura 5-9: Distribución de distancias entre el hogar y el colegio según quintil de ingreso.

Fuente: Elaboración propia a partir de SECTRA (2014).

La figura corrobora el análisis previo. Los hogares, a pesar de tener frecuencias similares en tiempos de viaje, poseen comportamientos relativamente diferentes en cuanto a distancias entre localizaciones. Por una parte, es razonable que dados los tiempos de viaje la mayoría de los colegios estén cerca de los hogares, pero a medida que aumenta el nivel de ingreso, la distancia entre residencia y colegio tiende a aumentar. En particular, se aprecia para el quintil más alto, que tienden a hacer viajes más largos que los otros quintiles, aunque eso no se refleje en mayores tiempos, necesariamente. Esto puede deberse en parte a una mayor tasa de motorización, como también a que pueden vivir en sectores donde el transporte público no es atractivo para los usuarios, por lo que optan por modos privados primero.

#### **5.4 Análisis de la muestra de observaciones EOD**

La muestra a utilizar en la estimación considera 2582 estudiantes que pertenecen a un total de 1981 hogares y que asisten a 853 colegios, de 938 alternativas posibles. Esto corresponde a un 50% del total de los estudiantes encuestados, y a un 54% del total de hogares encuestados que realizan viajes por educación escolar en la EOD 2012. Tomando en cuenta los factores de expansión de la EOD 2012, estos porcentajes son de un 49% y de un 54%, respectivamente. La Tabla 5-4 recoge los aspectos más relevantes de esta muestra, tanto para los hogares como para los estudiantes.

Por parte de los hogares de la muestra, se observa una mayor composición de familias de 4 o más personas. Es más, los hogares que se componen de 4 o 5 personas corresponden a un 59,2% del total de la muestra. Esto tiene sentido con respecto al análisis de la EOD, donde el tamaño del hogar promedio es de 4,7 integrantes. Mientras, los quintiles de ingreso aparentan estar bien representados, ya que existe una distribución homogénea sobre estos siendo el cuarto quintil el con mayor representación. Además, cerca del 55% de los hogares poseen automóvil, y 3 de cada 4 hogares son propietarios de sus viviendas. Por otra parte, un 27,9% del total de hogares observados tiene 2 o más estudiantes escolares en su familia. Dos tercios de estos hogares tienen a sus hijos e hijas en el mismo establecimiento, mientras que un 10% del total de la muestra escogió colegios distintos para ellos. Los 1981 hogares se distribuyen sobre 6 de las 7 macrozonas analizadas para el estudio. Esto se debe a que, en

la macrozona Extensión Sur – Poniente, el número de alternativas residenciales era muy bajo producto de falta de información (12 alternativas de más de 48 mil). Por lo tanto, ningún hogar de la muestra se localizó en ellas.

Tabla 5-4: Características de la muestra de hogares y estudiantes.

Fuente: Elaboración propia a partir de SECTRA (2014), MINEDUC (2014).

Hogares	Cantidad	%	Estudiantes escolares	Cantidad	%
<b>Observaciones</b>	1981	100	<b>Observaciones</b>	2582	100
<b>Integrantes</b>			<b>Sexo</b>		
2 personas	56	2,8	Femenino	1261	48,8
3 personas	385	19,4	<b>Edad</b>		
4 personas	683	34,5	5 - 6 Años	289	11,2
5 personas	489	24,7	7 - 13 años	1280	49,6
6 o +	378	19,1	14 - 18 años	1013	39,2
<b>Quintil de ingreso</b>			<b>Dependencia</b>		
[0; \$73.333]	376	19,0	Municipal	603	23,4
(\$73.333; \$108.000]	374	18,9	Particular Subvencionado	1586	61,4
(\$108.000; \$151.920]	415	20,9	Particular	393	15,2
(\$151.920; \$228.742]	422	21,3	<b>Enseñanza Media</b>	<b>839</b>	<b>32,5</b>
> \$228.742	394	19,9	Científico - Humanista	769	29,8
<b>Posee vehículo</b>	<b>910</b>	<b>45,9</b>	Técnico Profesional	70	2,7
1	711	35,9	<b>Religión establecimiento</b>		
2 o +	199	10,0	Laica	1681	65,1
<b>2 o más escolares</b>	<b>553</b>	<b>27,9</b>	Católica	901	34,9
En el mismo establecimiento	342	17,3	<b>Modo de viaje al colegio</b>		
En establecimientos diferentes	211	10,7	Motorizado	1917	74,2
<b>Casa Propia</b>	<b>1510</b>	<b>76,2</b>	Auto	516	20,0
<b>Ubicación geográfica</b>			Bus Transantiago	702	27,2
Norte	205	10,3	Metro	138	5,3
Poniente	486	24,5	Taxi	56	2,2
Oriente	252	12,7	Bus (No Transantiago)	10	0,4
Centro	77	3,9	Otro modo	593	23,0
Sur	468	23,6	No Motorizado	665	25,8
Sur - Oriente	493	24,9	Caminata	630	24,4
Extensión Sur - Poniente	0	0,0	Bicicleta	35	1,4

Por el lado de los estudiantes, que viajan día a día a estudiar, las proporciones de sexo se encuentran cerca del 50%, lo que es de esperar. Estos, están compuestos mayoritariamente en estudiantes de entre 7 y 13 años, quienes deberían estar estudiando en la educación básica. Además, de los establecimientos a los que asisten, sobre un 60% lo hace en colegios particulares subvencionados, un 23% en colegios municipales y sólo un 15% sobre el sector particular. Esto se condice con las proporciones presentes de los patrones de establecimientos escolares. También, cerca de dos tercios de los establecimientos a los que atienden son laicos, mientras que el resto pertenece tiene una orientación católica. No se observaron estudiantes que asistieran a establecimientos de otras religiones en la muestra, lo que es aceptable ya que a nivel absoluto no corresponde a más de un 3% del total de los establecimientos. Cabe mencionar que de los estudiantes que cursan educación media, que corresponden a un 30% de la muestra, la gran mayoría optó por una educación científico humanista.

Por último, de los modos utilizados en el viaje al colegio, priman los motorizados. Esto puede ser un potencial sesgo al momento de modelar, ya que la EOD demuestra que un 40% de los viajes al colegio son en modos no motorizados, a diferencia del 25% expuesto en la muestra. Dentro de los modos motorizados, prima el auto, el transporte público y el modo “Otros”, donde se ubica el transporte escolar. Esto tiene sentido según lo expuesto en la sección anterior. A grandes rasgos, el comportamiento es similar al descrito por la EOD, a pesar de que hay que mirar con detención el potencial sesgo de los viajes en modos no motorizados.

Para finalizar este análisis, cabe comparar las distancias de viaje a los establecimientos entre la muestra y la EOD. La Figura 5-10 presenta la distribución de las distancias euclidianas entre los establecimientos y los hogares, según la muestra y la EOD. Dada las diferentes magnitudes, se utilizó el porcentaje de viajes para cada intervalo de distancia.

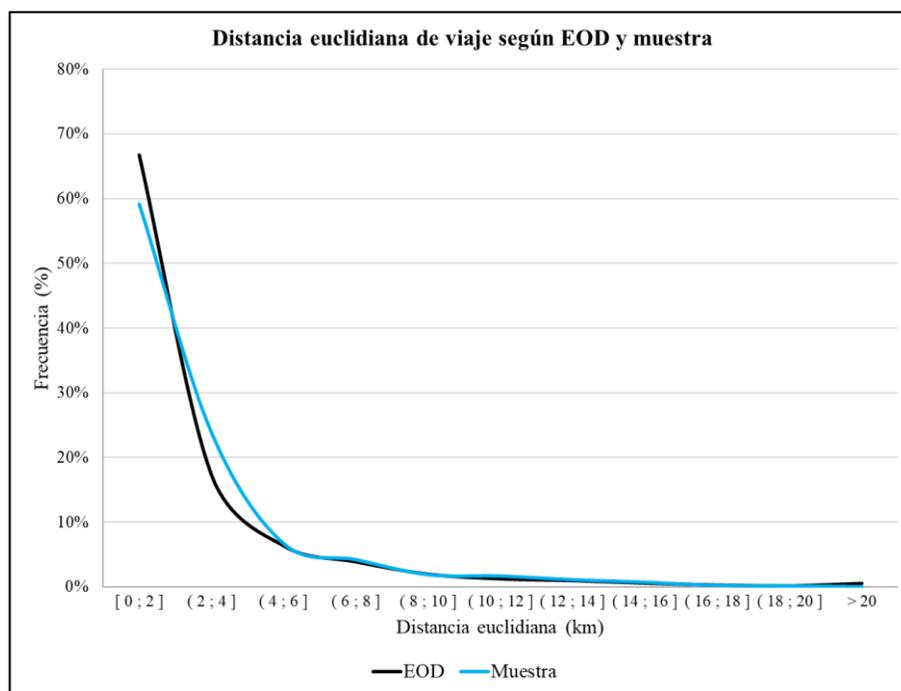


Figura 5-10: Distribución de distancias entre hogar y colegio según muestra.

Fuente: Elaboración propia a partir de SECTRA (2014).

De la imagen se observa un comportamiento similar en distancias entre la EOD, con un leve aumento entre los 2 y 6 kilómetros para las observaciones de la muestra. Además, se observa como cerca de un 10% menos de los viajes se realiza entre los 0 y 2 kilómetros, lo que es de esperar dado que la muestra posee menor partición modal de viajes no motorizados en comparación con la EOD. La Figura 5-11 pretende aportar a este análisis, comparando esta vez los tiempos de viaje entre ambos conjuntos de observaciones.

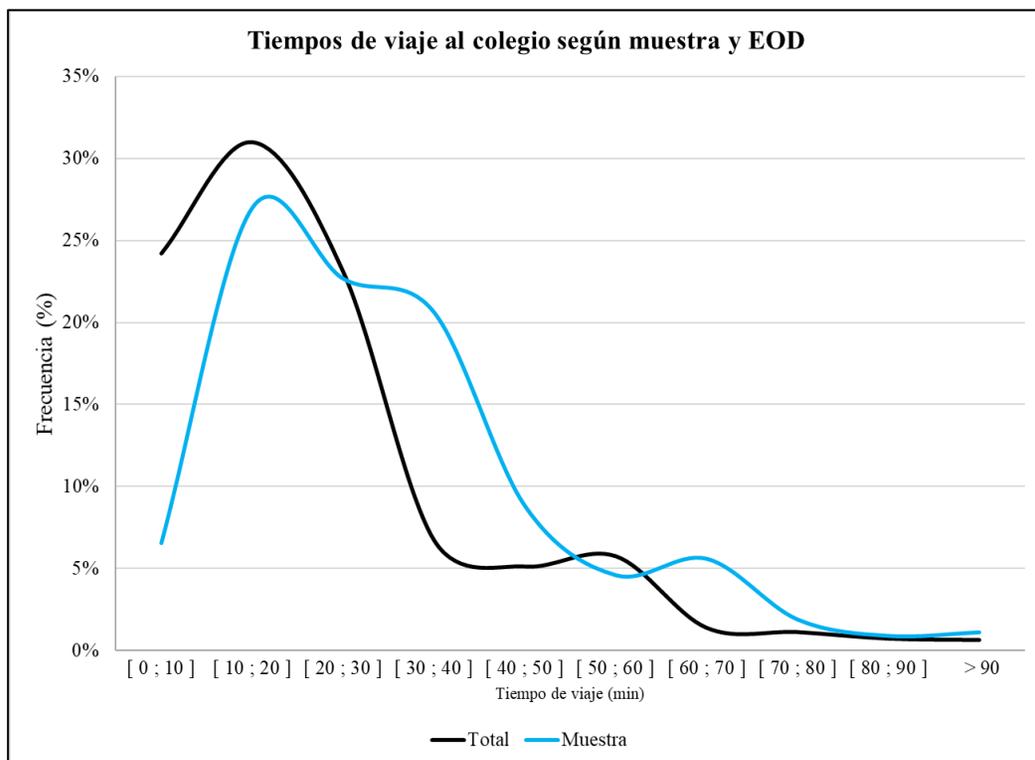


Figura 5-11: Distribución de tiempos de viaje al colegio según muestra.

Fuente: Elaboración propia a partir de SECTRA (2014).

Como era de esperar, existen diferencias en los tiempos de viaje reportados por las observaciones de la muestra. Por un lado, la muestra tiende a hacer menos viajes de corta duración, mientras que hay más frecuencia de viajes más extensos en las observaciones. Esto puede tener sentido en el contexto de mayor viajes motorizados, que están asociados a más distancias de viaje. Aunque, las distancias no presentaron tener diferencias drásticas en sus histogramas. Desde esa perspectiva, resulta conveniente utilizar la distancia euclidiana como medida pues refleja de mejor manera el comportamiento de los hogares. No obstante, hay que tener en cuenta en el análisis los potenciales sesgos que pueden existir debido a la diferencias en la composición modal.

## **6. RESULTADOS Y DISCUSIÓN**

En los capítulos previos se ha analizado la estructura de la ciudad en términos de alternativas de localización residencial y establecimientos escolares, junto con cómo se comportan los hogares en torno a estas elecciones. Se ha indagado también respecto de las variables más relevantes para estas decisiones, según la literatura. Con esta información, se procedió a estimar los modelos propuestos en la metodología descrita en la Sección 3. El objetivo de este capítulo es presentar y discutir los resultados de cuatro modelos, destacando sus principales aportes.

Previo a los resultados de los modelos, se explica brevemente el proceso de modelación, para entender la especificación final de cada modelo, junto con los atributos y características socioeconómicas incorporadas y no incluidas. Luego se presentan los resultados de cada modelo. Primero, se presentan los resultados de un modelo Logit Multinomial,. Luego se reportan los resultados para los moldeos tipo Logit Mixto, que captan correlación entre elecciones. Cada uno representa una de las estrategias de decisión descritas en la metodología: hogares priorizadores de localización residencial (PLR) y hogares priorizadores de establecimiento escolar (PEE). El último modelo estima mediante un enfoque de clases latentes, los resultados para ambas estrategias de elección y la probabilidad de adoptar una estrategia o la otra.

### **6.1 Especificación de variables del modelo**

El proceso de modelación consistió en iterar, con los atributos estudiados, probando especificaciones que se adecuaban a lo descrito por la literatura y fueran significativas para el modelo. Cabe mencionar que los modelos se estimaron desde el más simple al más complejo: Logit Multinomial, modelos Logit Mixtos con diferentes estrategias de elección y al final el modelo de clases latente que capta ambas estructuras. No obstante, y a modo de poder comparar de mejor manera las distintas estimaciones, el modelo más complejo (el de clases latentes), fue el que definió la especificación final a considerar para todos los modelos. La Tabla 6-1 muestra los atributos incorporados en los modelos, junto con el tipo de variable y valores posibles que puede tomar.

Tabla 6-1: Variables consideradas en los modelos, tipo de variable y rango.

<b>Variable</b>	<b>Tipo</b>	<b>Valores</b>
<b>Distancia euclidiana entre residencia y colegio (en km)</b>	Continua	$\geq 0$
<b>Atributos asociados a la elección de establecimiento escolar</b>		
Mensualidad (\$ CLP)	Discreta	Ver nota*
Establecimiento con educación Científico – Humanista	Binaria	{0,1}
Índice de Vulnerabilidad (IVE)	Continua	[0,1]
Proporción de jefes de hogar en el barrio con educación universitaria	Continua	[0,1]
Accesibilidad a oficinas	Continua	[0,1]
<b>Atributos asociados a la elección de localización residencial</b>		
Valor medio de la renta de vivienda (\$ CLP)	Continua	$\geq 0$
Accesibilidad a comercio	Continua	[0,1]
Accesibilidad a industria	Continua	[0,1]
Accesibilidad a oficinas	Continua	[0,1]
Proporción de jefes de hogar en el barrio con educación universitaria	Continua	[0,1]
<b>Características de los hogares y sus viajes al establecimiento</b>		
Ingreso per cápita	Continua	$\geq 0$
Tasa de Motorización (vehículos/hogar)	Discreta	{1,2,3, ...}
Posee casa propia	Binaria	[0,1]
Ingreso per cápita (i.p.c.) < 100 mil CLP	Binaria	[0,1]
Ingreso per cápita (i.p.c.) entre 100-400 mil	Binaria	[0,1]
Ingreso per cápita (i.p.c.) > 400 mil CLP	Binaria	[0,1]
Viaje al colegio en modo no motorizado	Binaria	[0,1]
Viaje al colegio en modo Otros	Binaria	[0,1]

\*Debido a que entregó un mayor ajuste, se optó por utilizar las medias de los intervalos de pago: {0, 5.000, 17.500, 37.500, 75.000, 300.000}.

Atributos de la localización residencial, como medidas de accesibilidad a salud, al igual que la presencia de estaciones de metro cerca fueron incorporadas a los modelos sin lograr significancia estadística. De la misma manera, atributos del entorno como la calidad de la edificación fueron incorporados sin éxito, por lo que se descartaron.

Por el lado de la elección de establecimiento escolar se consideraron pruebas estandarizadas como resultados PSU y SIMCE de los establecimientos. Sin embargo, estas no fueron significativas al momento de ser incorporadas en los modelos. De todas formas, esto puede

ser consistente con que los hogares, a pesar de declarar valorar estos aspectos, tienden a no considerarlos de manera efectiva en la elección (Bellei et al. 2020; Elacqua et al., 2006; Gómez et al, 2012). Atributos de número de docentes por estudiante al igual que la subvención por matrícula, que pueden ser indicadores de la calidad del establecimiento, fueron incluidas sin obtener resultados adecuados. Aspectos valóricos, como por ejemplo la orientación religiosa del establecimiento también fueron incorporados sin éxito, incluyendo variaciones sistemáticas según nivel de ingreso del hogar. Cabe mencionar que con respecto al atributo mensualidad, el último rango posee una gran heterogeneidad, por lo que es importante considerar la limitación del atributo. Ese valor se obtuvo mediante la revisión de medios de comunicación respecto de establecimientos privados el año 2012.

Por último, el único atributo observado que comparten ambas elecciones corresponde a la distancia euclidiana entre la casa y el establecimiento. Esta variable se incorpora de manera distinta para cada modelo, según lo explicado en la Sección 3. Para el MNL, se incorpora como un atributo compartido en cada par de alternativa de localización residencial y establecimiento (400 distancias diferentes, una por función). Mientras, en los modelos Logit Mixto la distancia se incorpora a sólo una elección según la estrategia definida. Por ejemplo, para la estrategia PLR, la distancia se incorpora como un atributo del establecimiento escolar, dada la localización escogida (20 distancias en la función de elección de colegio). Mientras, para la estrategia PEE, este atributo está en la función de utilidad de las alternativas de localización residencial, condicionada al establecimiento escogido (20 distancias en la función de elección de localización).

## **6.2 Resultados para el modelo Logit Multinomial (MNL)**

El modelo MNL busca demostrar que es razonable modelar ambas elecciones en conjunto, desde la simplicidad establecida por la estructura de correlación del modelo Logit Multinomial. Además, permite fijar una base para comparar con los resultados de modelos más sofisticados. Los resultados se presentan en la Tabla 6-2.

Para los atributos con interacciones, estas se consideran como una componente adicional a la base. Por lo tanto, para comprender su verdadero impacto es necesario analizar la suma

del coeficiente interactuado más el base. Además, tanto para este, como para los modelos siguientes se consideró un nivel de significancia al 90% (test-t mayor o igual a 1,64). Como se mencionó, aquellos parámetros con poca significancia fueron considerados igual, pues cobran relevancia en los modelos posteriores.

Tabla 6-2: Resultados modelo MNL.

<b>Especificación</b>	<b>MNL</b>	
	<b>Value</b>	<b>t-test</b>
<b>Atributo</b>		
<b><i>Distancia euclidiana entre residencia y colegio (en km)</i></b>	-0,47	-39,20
Interacción para hogares que viajan en modo no motorizado	-2,35	-23,10
Interacción para hogares que viajen en transporte escolar	-0,52	-13,90
<b><i>Coefficientes de elección de establecimiento escolar</i></b>		
Mensualidad (en MM CLP)	-4,39	-9,17
Colegio con educación Científico - Humanista		
Interacción para hogares con i.p.c. > 400 mil CLP	2,59	7,04
Índice de Vulnerabilidad (valor entre 0 – 1)		
Interacción para hogares con i.p.c. ≥ 100 mil CLP	-1,80	-8,73
Proporción de jefes de hogar en el barrio con educación universitaria aprobada (valor entre 0 - 1)	1,25	7,13
Interacción para hogares con i.p.c. > 400 mil CLP	0,34	0,66
Accesibilidad a oficinas	0,64	3,88
<b><i>Coefficientes de elección de localización residencial</i></b>		
Valor promedio de la renta de vivienda (en MM CLP)	-0,89	-2,94
Accesibilidad a comercio	0,06	0,18
Accesibilidad a industria	-1,54	-6,97
Accesibilidad a oficinas	-0,03	-0,08
Proporción de jefes de hogar en el barrio con educación universitaria aprobada (valor entre 0 - 1)	-4,69	-12,30
Interacción para hogares con i.p.c. entre 100 - 400 mil CLP	3,10	8,11
Interacción para hogares con i.p.c. > 400 mil CLP	7,57	12,50
Número de parámetros	16	
Log-Verosimilitud	-8.779	
Rho <sup>2</sup> ajustado	0,389	
AIC	17.590	

Cómo es de esperar, el atributo de la distancia entre la alternativa de localización residencial y la alternativa de establecimiento es significativo y negativo. Es decir, a mayor lejanía entre

ambas localizaciones, menor es la utilidad presente para el tomador de decisión. Esto, según los resultados se intensifica en aquellos casos donde el viaje que se realiza es en modos no motorizados como la caminata o la bicicleta, lo que es razonable debido a que los individuos tienden a estar menos dispuestos a recorrer distancias largas en estos modos. Algo similar, aunque en menor magnitud, ocurre con los hogares con estudiantes que realizan su viaje en transporte escolar. El atributo de la distancia junto con las interacciones estimadas son significativas. Se probaron interacciones con viajes en transporte público y auto, pero no resultaron significativas.

Por el lado de los atributos de la elección de establecimiento escolar, se encontraron comportamientos similares a los descritos por la literatura. El valor mensual del colegio resultó tener un coeficiente negativo, lo que era esperable. Además, los hogares con ingreso per cápita superior a 400 mil CLP valoran de manera positiva una educación científico humanista. Esto puede tener sentido, desde el punto de vista que la mayoría de los colegios del universo muestral imparten este programa de enseñanza. Además, este tipo de enseñanza entrega a futuro oportunidades en carreras universitarias, a las cuales accede con mayor facilidad los quintiles de mayor ingreso, lo que puede estar ligado con el éxito futuro del estudiante en la sociedad (Zimmerman, 2019). Se probó la interacción de esta variable con otros niveles de ingreso al igual que rangos etarios, pero no se logró significancia estadística y en algunos casos aparecieron resultados poco intuitivos. Se optó por eliminar aquellas interacciones no apropiadas.

Para hogares de ingreso per cápita superior a 100 mil CLP, existe una aversión a establecimientos de mayor vulnerabilidad escolar. Esto tiene sentido según la literatura científica, que asegura la intención de escoger colegio como un mecanismo de pertenencia a la comunidad y reproducción social (Bellei et al., 2020; Elacqua et al., 2006; Méndez y Gayo, 2019). Esto, visto desde el punto de vista que el IVE es una aproximación del nivel socioeconómico del alumnado.

Similar es la percepción de los hogares al momento de evaluar el entorno sobre el cual se sitúa el colegio, donde hay una tendencia a preferir establecimientos ubicados en barrios de mejor nivel educacional. Esto puede deberse tanto a lo explicado por la literatura de elección

escolar, como también a que las localizaciones residenciales de menor vulnerabilidad tienden a ser preferidas por los hogares (Weisbrod, Ben-Akiva y Lerman, 1979), lo que puede replicarse en el contexto escolar. La interacción para hogares con ingreso per cápita mayor a 400 mil CLP no fue significativa para este modelo, pero se mantuvo pues está definida por el modelo de clases latentes explicado en la Sección 6.4. Para finalizar, se observa una preferencia de los hogares por escoger establecimientos con buena accesibilidad a oportunidades de oficina. Esto puede verse como una medida razonable, pensando en que pueden existir viajes encadenados donde ir al establecimiento es una etapa de este, y el destino final es el lugar de trabajo.

Para los atributos de la elección de localización residencial, los resultados tienden a ser los esperados en algunos atributos. Primero, el valor medio de la renta para la localización posee una percepción negativa por parte de los tomadores de decisión, lo que es razonable desde la teoría económica. También, la accesibilidad a industrias posee una percepción negativa, lo que es razonable considerando que los hogares tienden a no querer vivir cerca de esta (Hurtubia et al., 2010). No obstante, accesibilidades como la de comercio y oficina no obtuvieron significancia en el modelo MNL. Esto puede deberse a la heterogeneidad presente en torno a estos atributos. Por un lado, puede ser razonable debido a que los hogares quieren estar próximos a oficinas y comercio, reduciendo sus costos de viaje a ellos. Pero, aglomeraciones de estos usos generan posibles externalidades negativas como ruido o congestión (Hurtubia et al., 2010). Se consideran de todas formas debido a la relevancia en los próximos modelos.

Para finalizar, el modelo explica que los hogares son aversos a escoger localizaciones con barrios de mejor nivel educacional. Se esperaba que, al igual que para la elección de colegio, este parámetro fuese positivo. No obstante, puede deberse a problemas de endogeneidad en las variables, en particular para los estratos de menor ingreso. Es difícil dilucidar hasta qué punto los hogares escogieron una localización debido al nivel socioeconómico, o dado que viven en un entorno con bajo nivel educacional aparenta que prefieren vivir en este. Es más, a medida que aumenta el nivel de ingreso (asociado al nivel educacional), se observa cómo la sensibilidad se acerca a cero, e incluso es positiva para hogares con nivel de ingreso superior a 400 mil CLP. Esto tiene sentido, ya que los hogares de mayor ingreso tienen

mayor capacidad de elección sobre el barrio en el cuál vivir. En Santiago esta situación es más extrema, donde se observó en la Sección 5 que la ciudad posee una distribución segregada en relación con el nivel educacional. Otro aspecto que podría causar problemas de signo en estas variables puede deberse a que los valores del arriendo utilizados poseen una mala calidad, por lo que otros coeficientes absorben efectos que la variable correspondiente no capta adecuadamente.

En general, el modelo MNL logra captar de manera adecuada las preferencias de los hogares en torno a ambas elecciones. La distancia que vincula a cada par de alternativas es significativa y valorada por los hogares. Además, atributos monetarios y socioeconómicos de los hogares tienden a ser relevantes en la decisión, mientras que aspectos valóricos y académicos pierden significancia para todo tipo de hogar. Para la elección residencial estos aspectos son igual de relevantes, aunque se esperaba mejor significancia de atributos asociados a la accesibilidad a diferentes oportunidades distribuidas en la ciudad. El modelo es un buen punto de partida para comprender la modelación conjunta de ambas elecciones.

No obstante, las limitaciones de este son evidentes. Por un lado, asume que ambas decisiones se eligen en a la par, sin establecer condicionalidades o dependencias entre una y la otra. Méndez y Gayo (2019) comentan de manera breve que existen diferentes estrategias de decisión, donde los hogares priorizan una elección, condicionando a la otra. Además, la literatura muestra que existe heterogeneidad en las preferencia de los hogares (Bellei et al., 2020; Elacqua et al., 2006 ; Méndez y Gayo, 2019). Esta sólo es captada por variaciones sistemáticas, mientras que pueden existir atributos latentes asociados a las percepciones de la pertenencia a la comunidad, o a estrategias de movilidad, que no sean captadas por estas. Dado esto, es razonable plantear los modelos Logit Mixtos, que estiman estrategias mixtas para los hogares tomadores de decisión y capturan correlación entre elecciones.

### **6.3 Modelos Logit Mixto capturando correlación entre elecciones**

Se modelaron dos estrategias de elección diferentes utilizando este enfoque. La primera corresponde a una priorización de la elección localización residencial (PLR). En esta, los hogares eligen localización residencial y luego, condicionado a esta elección eligen

establecimiento escolar. La segunda corresponde a una priorización de la elección de establecimiento escolar (PEE), donde la localización residencial es elegida condicionada a la elección de establecimiento realizada.

Al modelar ambos modelos, se consideraron los mismos parámetros. No obstante, la diferencia radica en que el atributo de la distancia determina la condicionalidad en las elecciones. En una estrategia PLR, este atributo determina la elección de colegio (dada la localización residencial escogida). Mientras, para la estrategia PEE, el atributo de la distancia determina la elección de localización residencial (dado el colegio escogido).

Además, se incorporaron dos términos de error aleatorios, según lo explicado en la Sección 3.2, cuyo objetivo es captar factores no observados que determinan ambas elecciones. Para estimar estos modelos es necesario integrar sobre la función de densidad de la distribución de los términos de error adicionales, lo que se realizó mediante Integración de Montecarlo, con 2000 *draws*. Debido a que pueden existir una o más observaciones para el mismo hogar (producto de la presencia de múltiples estudiantes y, en consecuencia, múltiples vínculos hogar-establecimiento), se implementó una corrección por efecto panel. Es decir, observaciones del mismo hogar, poseen el mismo término aleatorio asociado a los factores no observados previamente mencionados.

A grandes rasgos, cada modelo estima más parámetros que el modelo MNL. No obstante, poseen un  $\rho^2$  ajustado menor que el modelo previo. Sin embargo, es de esperar que al incorporar términos de error aleatorios y definir una estrategia de decisión que condiciona una elección sobre la otra, se estimen modelos que logren comprender mejor el fenómeno estudiado, aunque pierdan ajuste. Cabe mencionar que, entre ambos modelos, aquél que posee una estrategia PEE posee mejor ajuste. Esto puede ser indicio que, al modelar una estrategia similar para todos los hogares, estos tienden a adecuarse más a esta estructura que a la estrategia PLR. La Tabla 6-3 expone los resultados. Se explicitan también los parámetros de las ecuaciones de la Sección 3.

Tabla 6-3: Resultados modelos Logit Mixto con estrategias mixtas.

Especificación	Estrategia 1: PLR		Estrategia 2: PEE	
	valor	test-t	valor	test-t
<b>Atributo</b>				
<b><i>Distancia entre residencia y colegio (en km) (<math>\phi</math>)</i></b>	-0,52	-36,2	-0,55	-38,2
Interacción para viaje en modo no motorizado	-1,81	-11,0	-1,64	-14,4
Interacción para viaje en transporte escolar	-0,41	-9,06	-0,43	-10,7
<b><i>Coefficientes de elección de establecimiento escolar (<math>\beta</math>)</i></b>				
Mensualidad (en MM CLP)	-6,08	-9,41	-5,46	-13,1
Colegio con educación Científico - Humanista				
Hogares con i.p.c. > 400 mil CLP	1,71	3,45	1,50	4,23
Índice de Vulnerabilidad				
Hogares con i.p.c. $\geq$ 100 mil CLP	-2,36	-7,90	-2,45	-13,7
Proporción de jefes de hogar en el barrio con educación universitaria aprobada (valor entre 0 - 1)	1,62	5,75	-1,33	-7,73
Interacción para hogares con i.p.c. > 400 mil CLP	1,66	1,86	5,89	10,8
Desv. Estándar % jefes de hogar con Ed. Universitaria ( $\gamma$ )	3,60	17,70	3,18	17,8
Accesibilidad a oficinas	0,55	1,42	0,32	1,46
Desv. Estándar accesibilidad a oficinas ( $\gamma$ )	10,60	27,3	6,50	24,3
<b><i>Coefficientes de elección de localización residencial (<math>\theta</math>)</i></b>				
Valor promedio de la renta del barrio (en MM CLP)	0,01	0,06	-1,26	-3,84
Accesibilidad a comercio	-5,80	-12,6	-2,34	-4,82
Accesibilidad a educación	7,39	19,5	-	-
Accesibilidad a industria	-3,21	3,25	-2,08	-1,01
Accesibilidad a oficinas	1,45	0,06	-0,57	-3,84
Desv. Estándar accesibilidad a oficinas ( $\gamma$ )	10,60	27,3	6,50	24,3
Proporción de jefes de hogar en el barrio con educación universitaria aprobada (valor entre 0 - 1)	-6,87	-6,87	-5,61	-18,4
Interacción para hogares con i.p.c. entre 100 - 400 mil CLP	4,13	4,13	3,35	10,9
Interacción para hogares con i.p.c. > 400 mil CLP	11,90	11,9	9,52	17,4
Desv. Estándar % jefes de hogar con Ed. Superior ( $\gamma$ )	3,60	17,7	3,18	17,8
Número de parámetros	19		18	
Log-Verosimilitud	-12.064		-11.915	
Rho2 ajustado	0,346		0,354	
AIC	24.167		23.865	
Número de realizaciones	2.000		2.000	

### 6.3.1 Análisis de los coeficientes del modelo

En miras de no repetir las interpretaciones realizadas en la sección previa, se discute brevemente las similitudes que posee este modelo con las del modelo MNL. Por un lado, los parámetros de distancia en ambas estrategias poseen un comportamiento similar, siendo negativa la sensibilidad de este atributo sobre las elecciones de establecimiento (para la estrategia PLR) y de localización residencial (para la estrategia PEE). En la elección de localización residencial, otros atributos que se comportan de igual forma que en el modelo MNL corresponden a la mensualidad escolar, enseñanza científico humanista e índice de vulnerabilidad. Estos comportamientos son razonables según lo explicado previamente.

El atributo asociado al nivel educacional del barrio del colegio se mantiene positivo para el modelo con estrategia PLR mientras que, para la estrategia PEE, se torna negativo. A pesar de no ser el signo esperado, es explicable por posibles factores endógenos presentes en la elección, como se explicó previamente. Para los hogares de mayor ingreso, el efecto neto sí es positivo y significativo. Esto va de la mano con la potencial mayor capacidad de elección que tienen para ubicarse en barrios con mejor nivel educacional.

Además, se estimaron factores no observados que afectan la sensibilidad de los hogares con respecto al nivel educacional del barrio del colegio, por lo que existe aleatoriedad con respecto a las preferencias en este atributo. Esto implica que, con un nivel de significancia de un 90%<sup>5</sup>, el parámetro aleatorio posee una variación de  $\pm 5,90$  para la estrategia PLR y  $\pm 5,22$  para la estrategia PEE<sup>6</sup>, con respecto al coeficiente medio. De esta forma, los parámetros y sus interacciones poseen heterogeneidad no caracterizable, lo que puede invertir el sentido de las preferencias. Las probabilidades de que el parámetro sea negativo para un hogar, según cada modelo se presentan en la Tabla 6-4. La interpretación de estos factores no observados se analizará en la siguiente subsección.

---

<sup>5</sup> Se optó por utilizar una significancia del 90% para ser consistente con el análisis de los parámetros.

<sup>6</sup> Un parámetro aleatorio que distribuye Normal  $(\mu, \sigma)$  posee un intervalo de confianza, al 90%, de  $\mu \pm 1,64 \cdot \sigma$ , siendo  $\mu$  el coeficiente asociado al atributo y  $\sigma$  la desviación estándar estimada.

Para finalizar, el coeficiente asociado a la accesibilidad a oficinas se mantiene positivo para ambas estrategias. Sin embargo, no es significativo al 90%. Además, se captaron factores no observados a través de un término de error aleatorio. Dada la magnitud del parámetro, el coeficiente medio estimado para el atributo oscila con una variación de  $\pm 17,38$  para la estrategia PLR y  $\pm 10,66$  para la estrategia PEE. Esto implica que la heterogeneidad en el parámetro es alta. Existen hogares con una percepción negativa y otros positiva del parámetro. Las probabilidades de que el parámetro sea negativo, según cada modelo se presentan en la Tabla 6-4.

Tabla 6-4: Coeficientes aleatorios en elección de establecimiento escolar y probabilidad de ser negativos, según estrategia modelada.

<b>Atributo</b>	<b>Estrategia 1: PLR</b>	<b>Estrategia 2: PEE</b>
Proporción de jefes de hogar en el barrio con educación universitaria aprobada (valor entre 0 - 1)	32,6%	66,2%
Interacción para hogares con i.p.c. > 400 mil CLP*	18,1%	7,6%
Accesibilidad a Oficina	48,0%	48,0%

\*La interacción considera el coeficiente base sumado al interactuado.

Para la elección de localización residencial, es interesante analizar sus coeficientes. Primero, el coeficiente de valor de renta de vivienda se mantiene negativo para el modelo con estrategia PEE. En cambio, para la estrategia PLR el parámetro se vuelve positivo, pero con una significancia nula. Este resultado no es esperado. No obstante, la construcción del atributo posee limitaciones serias. Por un lado, representa un valor agregado que en ocasiones considera valores estimados de precio imputado por los encuestados en la EOD. Por otro, no se ajusta a un valor por m<sup>2</sup> de suelo. Esto genera distorsiones sobre el valor real de este. Por esto, y a pesar de su muy baja significancia, se optó por mantener el parámetro.

Los coeficientes de accesibilidad a comercio ganan significancia al modelar ambas estrategias por separado, con un valor negativo. Esto podría llegar a ser razonable si los hogares prefieren no vivir cerca de barrios comerciales o, en su defecto, la muestra no posee hogares que vivan cerca de él. Por otra parte, la accesibilidad a industria posee coeficientes razonables e iguales a lo explicado en el modelo MNL. Además, debido a que la estrategia

PLR escoge localización sin depender del establecimiento, se incorporó un parámetro de accesibilidad a educación, significativo y positivo.

El parámetro de accesibilidad a oficinas presenta diferencias respecto al modelo MNL. Para el modelo PEE, se vuelve significativo y negativo. Esto, en conjunto con el análisis de accesibilidad a oficinas para la elección escolar, da indicios que un modelo con esta estrategia puede adecuarse a hogares que prefieren vivir lejos de la oficina, pero que esta se sitúe cerca del colegio. Mientras, la estrategia PEE posee un coeficiente positivo, pero no significativo. No obstante, hay que considerar el efecto que el término de error posee sobre esta interpretación. Dado que el término estimado es igual para ambas elecciones, la variabilidad presente al 90% de confianza es igual a la explicada previamente. La Tabla 6-5 explica en detalle las probabilidades de que el coeficiente sea negativo o positivo.

Con respecto al atributo asociado al nivel educacional del barrio de la localización residencial, ambos modelos se comportan de igual manera al MNL, donde existe una percepción negativa sobre la elección (que puede responder a aspectos endógenos). Pero, a medida que aumenta el nivel de ingreso de los hogares, esta percepción se acerca a ser positiva. De todas formas, es necesario considerar el impacto que los factores no observados tienen sobre este parámetro, ya que las sensibilidades pueden variar según cada hogar. Los términos de error estimados entregan el mismo valor de variabilidad al 90%, según lo explicado previamente. Mientras, la Tabla 6-5 muestra las probabilidades de que los coeficientes sean positivos o negativos, según las interacciones asociadas al atributo. La interpretación de estos factores no observados estimados se explica a continuación.

Tabla 6-5: Coeficientes aleatorios en elección de localización residencial y probabilidad de ser negativos, según estrategia modelada.

<b>Atributo</b>	<b>Estrategia 1: PLR</b>	<b>Estrategia 2: PEE</b>
Accesibilidad a Oficina	44,6%	53,5%
Proporción de jefes de hogar en el barrio con educación universitaria aprobada (valor entre 0 - 1)	97,2%	96,1%
Interacción para hogares con i.p.c. entre 100 -400 mil CLP	60,2%	76,1%
Interacción para hogares con i.p.c. > 400 mil CLP	8,1%	10,9%

### **6.3.2 Factores no observados que capturan correlación entre elecciones**

Los modelos Logit Mixtos propuestos para ambas estrategias, PLR y PEE, buscan hacerse cargo de captar factores no observados que correlacionan las elecciones. Esto lo realizan a través de términos de error asociados a los atributos de iguales características de ambas alternativas. Los errores poseen una distribución Normal de media cero, y una desviación estándar estimada por el modelo, buscando captar heterogeneidad en las preferencias de los hogares con respecto a atributos asociados a ambas alternativas. Los resultados sobre estos factores, asociados a accesibilidad a oficina y el nivel educacional del barrio (de las localizaciones residenciales y de colegio) son similares en ambos modelos.

Para el atributo de accesibilidad a oficina, se estimó una desviación estándar significativa. Esto significa que existen factores no observados, asociados a las preferencias de los hogares con respecto al atributo de accesibilidad a oficinas, que determinan tanto la elección de establecimiento escolar como la elección de localización residencial. Estos factores pueden corresponder a estilos de vida o preferencias de movilidad. Por ejemplo, hogares que prefieren la vida en los suburbios ya que valoran la tranquilidad de esos barrios y por ende son aversos a vivir (o elegir un colegio) cerca del trabajo, u hogares con estrategias de movilidad sustentables que sólo se mueven en modos activos, y por ende valoran positivamente vivir cerca de mayores oportunidades de oficina. Trabajos como Waddell et al. (2007) o Waddell (1993) plantean lo interconectadas que pueden estar las elecciones de localización y de trabajo. Por lo tanto, pueden existir factores relevantes no considerados en este modelo que vinculen las oportunidades de oficinas con las preferencias de los hogares. Esto conlleva también a que para aquellas estrategias donde el parámetro de accesibilidad a oficina resultó poco significativo en alguna elección, aun así hay efecto del atributo sobre las preferencias. Por cierto, efectos heterogéneos según las preferencias de los hogares.

También, se estimaron términos de error sobre el atributo del nivel educacional del barrio de las elecciones. Esto significa que existen factores no observados asociados a las preferencias de los hogares respecto del nivel educacional del entorno de la residencia y del entorno del colegio, que determinan ambas elecciones. Esto va en línea directa con lo descrito por la literatura. Méndez y Gayo (2019) plantean cómo las expectativas en cuanto a éxito

profesional y cultural de los hogares, con respecto a estas elecciones, generan interacciones al momento de decidir. Estas expectativas podrían verse captadas por el término de error. Por otro lado, el sentido de pertenencia a la comunidad que valoran los hogares de la clase alta al momento de elegir un establecimiento, al igual que la heterogeneidad con respecto a las preferencias académicas entre diferentes hogares (Bellei et al., 2020) podrían ser factores incluidos dentro de la aleatoriedad del parámetro.

Estos parámetros logran captar heterogeneidad en las preferencias, que no se está detectando mediante variaciones sistemáticas. Además, dan luces de la fuerte correlación que puede existir entre ambas elecciones, que se vinculan por atributos observados y no observados.

### 6.3.3 Cálculo de elasticidades y comparación de estrategias

Debido a que los modelos Logit no son capaces de separar la estimación de los coeficientes de cada atributo con el parámetro de escala del modelo, comparar directamente los parámetros estimados puede llevar a conclusiones erradas respecto de la relevancia relativa de cada atributo en la elección. No obstante, es interesante comparar ambas estrategias de elección, definidas por cada modelo, y así comprender mejor las preferencias de los hogares.

Se compara la sensibilidad de los parámetros a través del cálculo de elasticidades de demanda del modelo. Se estima en qué porcentaje varía la probabilidad de elección de una alternativa, al variar en un porcentaje un atributo. La ecuación 6.1 explica el cálculo.

$$E_{P_{iq}, x_{ikq}} = \theta_{ik} \cdot x_{ikq} \cdot P_{iq} \quad (6.1)$$

Donde  $\theta_{ik}$  corresponde al parámetro asociado al  $k$ -ésimo atributo  $x_{ikq}$  de la alternativa  $i$  del individuo  $q$ . Mientras,  $P_{iq}$  representa la probabilidad de escoger la alternativa  $i$  para el individuo  $q$ . Para efectos del análisis, se estimó para todos los individuos la elasticidad de la demanda de los atributos que determinan la probabilidad de elección de la alternativa de establecimiento escogida. Luego, se calculó la elasticidad de la demanda para los atributos que determinan la probabilidad de elección de la alternativa de localización residencial escogida. Para finalizar, se estimó el valor promedio de estas elasticidades, para cada

elección y para cada estrategia de decisión. Cabe mencionar que, por parsimonia, se optó por considerar sólo los coeficientes promedio en aquellos atributos con términos de error asociados. No obstante, es importante considerar en el análisis que aquellas elasticidades y probabilidades, poseen aleatoriedad producto de las desviaciones estándar no consideradas.

La Tabla 6-6 muestra las elasticidades para las elecciones de establecimiento escolar para los modelos PLR y PEE. Aquellas elasticidades de atributos interactuados con características del hogar o del viaje fueron calculadas con el valor neto del parámetro (i.e. sumando el valor base al interactuado). El orden presentado es según todos los parámetros del modelo, a pesar de que las interacciones se aplican solo para aquellos hogares con esas características.

Tabla 6-6: Elasticidades de demanda para elección de establecimiento, según estrategia.

<b>Modelo</b>	<b>Orden</b>	<b>Atributo</b>	<b>Elasticidad</b>
PLR	1	Distancia residencia - colegio para viajes en modo no motorizado	-1,45
	2	IVE para hogares con i.p.c. $\geq 100$ mil CLP	-1,10
	3	Distancia residencia - colegio para viajes en transporte escolar	-0,82
	4	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. $> 400$ mil CLP	0,68
	5	Colegio con educación científico-humanista para hogares con i.p.c. $> 400$ mil CLP	0,55
	6	Distancia residencia - colegio	-0,37
	7	Proporción de jefes de hogar en el barrio con educación universitaria	0,20
	8	Accesibilidad a oficina	0,17
	9	Valor de la mensualidad escolar	-0,13
<b>Elección</b>	<b>Orden</b>	<b>Atributo</b>	<b>Elasticidad</b>
PEE	1	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. $> 400$ mil CLP	0,22
	2	IVE para hogares con i.p.c. $\geq 100$ mil CLP	-0,11
	3	Colegio con educación científico-humanista para hogares con i.p.c. $> 400$ mil CLP	0,10
	4	Proporción de jefes de hogar en el barrio con educación universitaria	-0,02
	5	Accesibilidad a oficina	0,01
	6	Valor de la mensualidad escolar	-0,01

De la tabla se desprende que, para la estrategia PLR que incorpora la distancia como un parámetro de la elección de establecimiento escolar, es un parámetro que cobra importancia en la elección. Para aquellos viajes en modos no motorizados o en transporte escolar, la distancia al colegio es el primer o segundo atributo más relevante. Mientras que, para aquellos viajes que se realizan en otros modos motorizados (e.g. automóvil o transporte público), la probabilidad de elección pasa a ser menos sensible a variaciones en la distancia.

Por el otro lado, la elasticidad sobre los otros atributos es similar para ambas estrategias. Primero, a mayor nivel de ingreso, mayor tiende a ser la sensibilidad de los hogares con respecto a aspectos socioeconómicos del establecimiento y su entorno. Para ambas estrategias, la probabilidad de elección en hogares sobre los 400 mil CLP per cápita es muy elástica con respecto al nivel educacional del barrio, ubicándose entre los atributos más sensibles (sin considerar la distancia). Con respecto al IVE, se observa también como todos los hogares sobre 100 mil CLP de ingreso per cápita, consideran este atributo dentro de los dos más sensibles para elegir un establecimiento. Por otra parte, los hogares de ingresos por sobre los 400 mil CLP, también son más sensibles a que el colegio sea científico humanista, en comparación con aumentos en la mensualidad escolar o mejor accesibilidad a la oficina (considerando el valor medio y no los factores aleatorios).

Esto no quita que los hogares de menores niveles de ingreso no sean sensibles a estos atributos. Es más, también tienden a ser más sensibles a variaciones en el nivel educacional del barrio del establecimiento en comparación con aspectos como la accesibilidad a oficinas o mensualidad. Esto, para ambas estrategias.

La similitud con respecto a las elasticidades puede deberse a que, al especificar una sola estrategia en cada modelo, no se captan las diferentes preferencias que puedan poseer los hogares. Pues, se restringe a que todos se comporten de manera similar. En ese sentido, un enfoque de clases latentes que capte según las características del hogar las preferencias de elección podría generar resultados diferentes. La Tabla 6-7 capta las elasticidades con respecto a la elección de localización residencial, para ambas estrategias de elección.

Tabla 6-7: Elasticidades de demanda para elección de localización residencial, según estrategia.

<b>Modelo</b>	<b>Orden</b>	<b>Atributo</b>	<b>Elasticidad</b>
PLR	1	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. > 400 mil CLP	0,25
	2	Accesibilidad a educación	0,11
	3	Proporción de jefes de hogar en el barrio con educación universitaria	-0,04
	4	Accesibilidad a oficina	0,03
	5	Accesibilidad a comercio	-0,03
	6	Accesibilidad a industria	-0,02
	7	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. entre 100-400milCLP	-0,01
	8	Valor medio de la renta	0,0001
<b>Modelo</b>	<b>Orden</b>	<b>Atributo</b>	<b>Elasticidad</b>
PEE	1	Distancia residencia - colegio para viajes en modo no motorizado	-1,14
	2	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. > 400 mil CLP	0,74
	3	Distancia residencia - colegio para viajes en modo "otros"	-0,57
	4	Proporción de jefes de hogar en el barrio con educación universitaria	-0,41
	5	Distancia residencia - colegio	-0,27
	6	Accesibilidad a industria	-0,18
	7	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. entre 100-400milCLP	-0,17
	8	Accesibilidad a oficina	-0,13
	9	Accesibilidad a comercio	-0,13
	10	Valor medio de la renta	-0,09

Se observa un comportamiento similar a lo ocurrido en las elasticidades de demanda para la elección de establecimiento escolar. Por un lado, el parámetro de distancia (interactuado con el uso de modos no motorizados) es el más importante en la estrategia PEE). Las interacciones con otros modos para la distancia también son relevantes, con la interacción con transporte escolar en tercer lugar y el efecto base de distancia en quinta prioridad.

Se vuelve a observar que, para aquellos hogares con ingreso per cápita superior a 400 mil CLP, el nivel educacional del barrio es de los atributos más elásticos. Esto se repite también

para hogares que no poseen interacciones (i.e. hogares con ingreso menor a 100 mil CLP), pero de manera negativa. Sin embargo, esto en parte se debe a potenciales efectos endógenos del atributo explicados previamente. Por último, para hogares de ingresos per cápita entre 100 mil y 400 mil CLP, la probabilidad de elección no posee gran sensibilidad con respecto a este atributo, ubicándose en las elasticidades de menor sensibilidad. Este comportamiento, según las diferentes características de los hogares se repite en ambas estrategias modeladas. Cabe mencionar que no se considera el efecto que los factores no observados estimados pueden tener sobre estas elasticidades.

Con respecto a las accesibilidades, se observa cómo la accesibilidad a educación si es un atributo sensible para la elección de localización residencial, en el caso de la estrategia PLR. Es más, para aquellos hogares que no viajan en modos no motorizados, es el atributo más relevante. Esto puede reflejar la relevancia de comprender en conjunto ambas elecciones, aun cuando se escoge localización residencial sin condicionarse al establecimiento escogido. Por otra parte, accesibilidades de industria, oficina y comercio se sitúan en posiciones bajas de la tabla, para ambas estrategias de elección. Existen variaciones en el orden de estas, pero tienden a tener elasticidades similares en cada elección con pequeñas diferencias.

Para finalizar, en ambos modelos la elección es poco sensible al valor de la renta mensual de la vivienda. Esto puede deberse en parte a que el cálculo del atributo es poco confiable, lo que lleva a preferencias poco sensibles sobre este. De todas formas, para el modelo con la estrategia PLR, donde el parámetro estimado era positivo y poco significativo, la elasticidad es muy cercana a 0.

Modelar en conjunto la elección de establecimiento escolar y localización residencial a través de modelos Logit Mixtos es efectivamente útil, logrando captar de manera adecuada las preferencias de los hogares. Además, identifica factores no observados asociados a las preferencias de los hogares con respecto a ciertos atributos que determinan ambas elecciones. Esto, sumado a que establece según sus especificaciones dependencia en las elecciones, siendo una la que condiciona a la otra según los atributos observados de las alternativas elegidas.

## **6.4 Modelo Logit Mixto: enfoque de clases latentes**

Tomando en cuenta lo visto en las secciones previas, existe espacio para poder captar heterogeneidad en las preferencias, a través de un enfoque de clases latentes. Los modelos Logit Mixto explicados previamente, a pesar de que estiman preferencias a través de variaciones sistemáticas e identifican heterogeneidad en las preferencias no caracterizables a través de términos aleatorios, asumen que los hogares adoptan una estrategia o la otra de manera absoluta. No obstante, es posible que algunos hogares utilicen una estrategia o la otra, o incluso combinaciones de ellas, lo que podría venir determinado por sus características.

Se estimó un modelo que considera ambas estrategias y estima sus respectivos parámetros mediante un enfoque de clases latentes. El modelo se estimó utilizando Integración de Montecarlo, para estimar los coeficientes aleatorios, con 2000 realizaciones. Además, se aplicó un efecto panel sobre aquellas observaciones que pertenecen al mismo hogar. Primero, se expondrán los resultados de la clase latente estimada, luego se analizarán los resultados de los parámetros de cada estrategia y sus variaciones con respecto al análisis previo. Para finalizar, se analizarán las elasticidades y sus cambios respecto a los modelos vistos con anterioridad.

### **6.4.1 Clase latente**

Para estimar la probabilidad de pertenecer a cada clase, se utilizó una estructura tipo Logit. Esta función de utilidad fue asociada a la estrategia PEE, como se describe en la ecuación 3.10, mientras que la probabilidad de elección de la clase PLR corresponde a la de la ecuación 3.12. Los resultados de los coeficientes de la clase latente se presentan en la Tabla 6-8.

Tabla 6-8: Resultados de parámetros de clase latente.

Características del hogar	Coeficientes asociados a la estrategia PEE	
	Valor	test-t
Constante de pertenencia a la clase	-0,19	-0,78
Hogar posee casa propia	-0,68	-3,05
Ingreso per cápita (en MM CLP)	6,02	4,55
Número de vehículos	1,04	5,82

Se observa que los parámetros utilizados son significativos, a excepción de la constante de pertenencia a la clase. El signo de las variables parece adecuado. Por un lado, la constante es negativa, lo que puede ser razonable. Aunque, debido a que el ingreso per cápita es una variable positiva mayor a cero y tiene una magnitud relevante (aproximadamente 0,31 útiles por cada 50 mil pesos), el efecto neto para el caso más extremo tiende a ser positivo o cercano a cero. Con respecto a la variable de ingreso, las estimaciones demuestran que a mayor ingreso es más probable pertenecer a la clase que prioriza el establecimiento escolar. Esto puede tener sentido, desde la perspectiva que los hogares de mayor ingreso tienden a poseer más expectativas sobre el establecimiento escogido (Méndez y Gayo, 2019) y valoran aspectos como la pertenencia a la comunidad escolar (Bellei et al., 2020).

No obstante, a pesar del nivel de ingreso, si el hogar es propietario de la vivienda donde vive, disminuye su probabilidad de pertenecer a la clase PEE. Hogares que poseen una casa propia, están sujetos a mayores restricciones de localización y elección. Sería razonable pensar que un hogar que sí es propietario de su vivienda, *ceteris paribus*, determine sus elecciones de movilidad y localización de trabajo o colegio, según el lugar en donde se encuentra establecido. En cambio, hogares que son arrendatarios son más móviles al momento de tomar decisiones de localización.

Por último, hogares con mayor número de vehículos son más probables a pertenecer a esta clase. Esto podría tener un efecto sobre las preferencias de los hogares con respecto a las distancias entre colegio y casa. Así también, puede ser un indicador de los estilos de vida que poseen. A mayor número de vehículos, podría ser menos probable que viajen en modos activos o transporte público. Podrían también estar más dispuestos a elegir un

establecimiento más lejano, si trae consigo otros beneficios académicos o sociales. La Figura 6-1 presenta la probabilidad, para los hogares de la muestra, de pertenecer a la clase que escoge según estrategia de Priorización de la elección de Establecimiento Escolar (PEE).

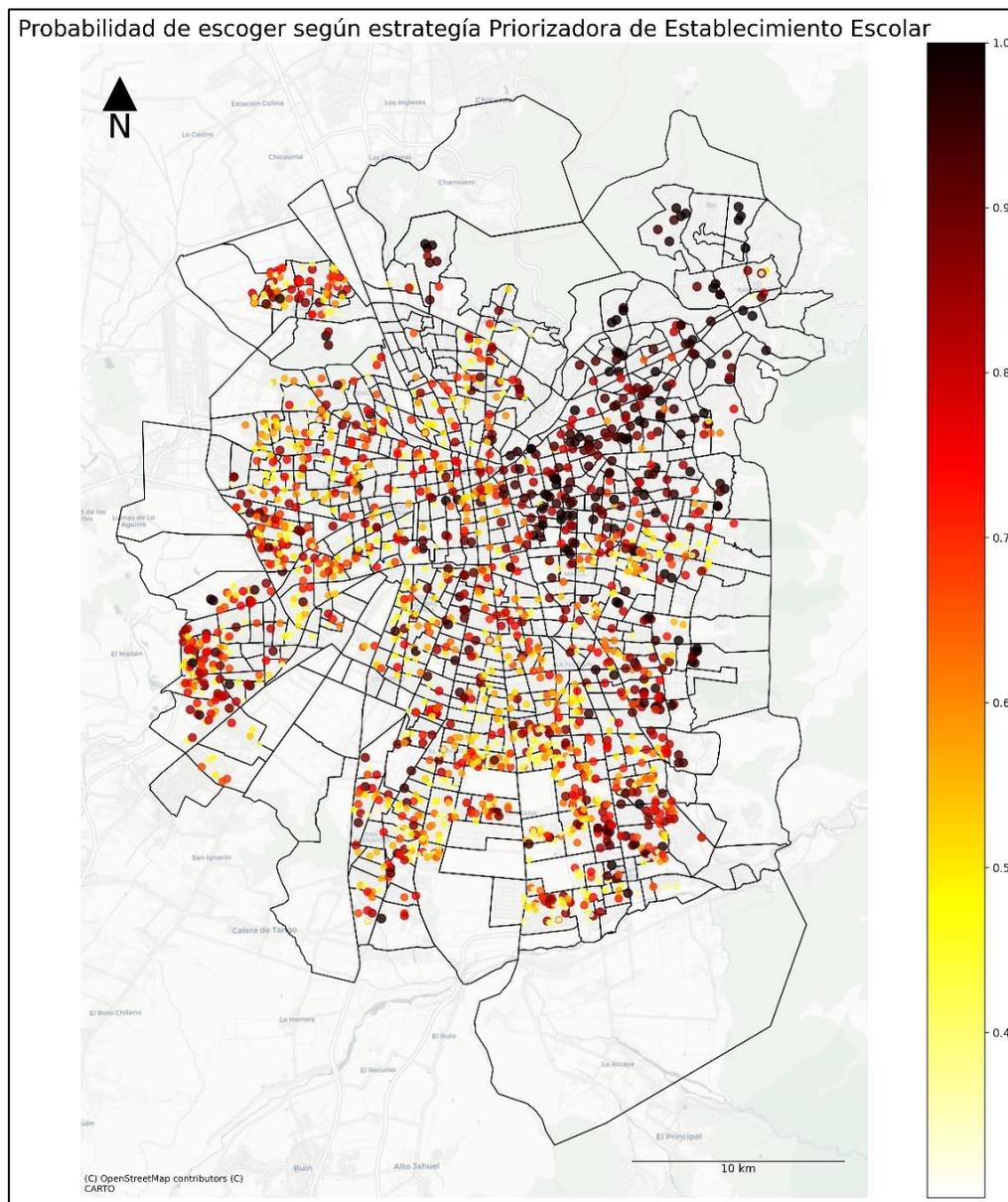


Figura 6-1: Probabilidad de escoger según estrategia PEE.

En la figura, se aprecia cómo existe una distribución cargada hacia el sector Oriente de la capital, donde la mayoría de los hogares tienden a elegir con alta probabilidad según la clase

PEE. Esto se debe en gran medida a que la clase latente capta con mayor importancia aspectos del nivel de ingreso y de la posesión vehicular, por lo que es razonable que, hacia sectores de mayor nivel socioeconómico, mayor sea la probabilidad de pertenecer a la clase. También, se observa de todas formas cómo existe alta probabilidad en otros sectores de la ciudad. Es más, la probabilidad mínima de pertenecer a esta clase es igual a un 30,01%. Esto puede significar en parte que, dada la especificación definida, los hogares tienden a ajustarse mejor hacia esta estrategia que hacia una estrategia PLR. En promedio, la probabilidad de pertenecer a la clase PEE corresponde a un 65,3%.

Se intentó incorporar más características de los hogares a la clase latente. En particular, se probaron especificaciones con consideraciones de posesión de bicicleta, género de los y las estudiantes o edad, entre otros. Se consideró utilizar también características del viaje entre el hogar y el colegio, que podrían ser una buena medida con respecto a las preferencias que poseen respecto a movilidad sustentable. Sin embargo, se optó por descartar esta opción debido a potenciales problemas de autoselección, pues no es posible dilucidar con claridad si el viaje determina la elección o viceversa.

#### **6.4.2 Análisis de los parámetros estimados**

Los resultados de los coeficientes estimados, para cada clase latente se presentan en la Tabla 6-9. El modelo presenta mejor ajuste que los modelos previos. Es esperable, pues se incorporaron nuevos parámetros de la función de utilidad de la clase latente.

Los resultados tienden a ser bien similares con respecto a los modelos expuestos en las secciones previas. Atributos de la distancia entre el hogar y el colegio, para ambas clases, mantienen las mismas interpretaciones previas.

Tabla 6-9: Resultado modelo con enfoque de clases latentes.

<b>Especificación</b>	<b>Clase 1: PLR</b>		<b>Clase 2: PEE</b>	
<b>Atributo</b>	<b>Value</b>	<b>t-test</b>	<b>Value</b>	<b>t-test</b>
<b><i>Distancia entre residencia y colegio (en km) (<math>\phi</math>)</i></b>	-0,69	-10,50	-0,60	-23,94
Interacción para viaje en modo no motorizado	-1,66	-6,68	-2,57	-8,03
Interacción para viaje en modo otros	-0,60	-3,63	-0,47	-6,31
<b><i>Coefficientes de elección de establecimiento escolar (<math>\beta</math>)</i></b>				
Mensualidad (en MM CLP)	-10,25	-5,52	-6,07	-10,44
Colegio con educación Científico - Humanista				
Hogares con i.p.c. > 400 mil CLP	13,95	0,47	1,44	3,74
Índice de Vulnerabilidad				
Hogares con i.p.c. $\geq$ 100 mil CLP	0,01	0,01	-3,39	-12,82
Proporción de jefes de hogar en el barrio con educación universitaria aprobada (valor entre 0 - 1)	5,41	7,11	-1,20	-4,76
Interacción para hogares con i.p.c. > 400 mil CLP	10,05	1,33	5,37	8,92
Desv. Estándar % jefes de hogar con Ed. Universitaria ( $\gamma$ )	-5,40	-7,64	3,10	13,01
Accesibilidad a oficinas	-2,98	-3,30	1,46	4,34
Desv. Estándar accesibilidad a oficinas ( $\gamma$ )	11,84	10,86	7,54	18,37
<b><i>Coefficientes de elección de localización residencial (<math>\theta</math>)</i></b>				
Valor promedio de la renta del barrio (en MM CLP)	-1,18	-1,01	-0,68	-1,92
Accesibilidad a comercio	-12,37	-6,17	-1,69	-2,73
Accesibilidad a educación	13,15	11,48	-	-
Accesibilidad a industria	-6,38	-7,84	-1,31	-3,48
Accesibilidad a oficinas	0,03	0,03	-1,63	-2,13
Desv. Estándar accesibilidad a oficinas ( $\gamma$ )	11,84	10,86	7,54	18,37
Proporción de jefes de hogar en el barrio con educación universitaria aprobada (valor entre 0 - 1)	-11,66	-11,66	-3,22	-8,85
Interacción para hogares con i.p.c. 100 - 400 mil CLP	2,69	2,69	2,60	2,05
Interacción para hogares con i.p.c. > 400 mil CLP	12,39	12,39	7,20	2,86
Desv. Estándar % jefes de hogar con Ed. Universitaria ( $\gamma$ )	-5,40	-7,64	3,10	13,01
Número de parámetros	43			
Log-Verosimilitud	-11.625			
Rho2 ajustado	0,371			
AIC	23.336			
Número de realizaciones	2000			

Por parte de los atributos vinculados a la elección escolar, las diferencias se focalizan en la clase con estrategia PLR. En esta, las interacciones con hogares de mayor nivel de ingreso se vuelven poco significativas. Esto tiene sentido porque es poco probable que hogares de

ingresos altos pertenezcan a esta clase. Por lo tanto, las preferencias tienden a ser menos relevantes al interactuarse con el hogar. Es más, el IVE estimado es positivo y poco significativo, a diferencia del parámetro negativo y con buena significancia del modelo PLR anterior. Cabe mencionar que el atributo Científico – Humanista del colegio posee un coeficiente positivo de gran magnitud (considerando que es una variable binaria), pero muy poco significativo.

Por otra parte, el parámetro estimado de accesibilidad a oficina para la estrategia PLR se vuelve negativo, pero gana significancia. Esto puede significar que hogares de menor ingreso tienden a valorar menos establecimientos cerca de sectores de oficinas, a diferencia de hogares de mayor ingreso (con más probabilidad ser de la clase PEE), que tienden a preferir establecimientos con buenas oportunidades de oficina cerca.

No obstante, y al igual que en el análisis anterior, es relevante considerar que es un parámetro con gran heterogeneidad no observable, por lo que es interesante analizar la probabilidad que sea positivo o negativo. Al mismo tiempo, se calcula la probabilidad para el atributo del nivel educacional del barrio del establecimiento. Las probabilidades que estos parámetros sean negativos, para cada clase, se presentan en la Tabla 6-10. De esta se observa un comportamiento similar en comparación con los resultados de los modelos previos, con una tendencia a mayor positividad en la clase PLR, con la consideración que la interacción con el nivel de ingreso alto es poco significativa.

Tabla 6-10: Coeficientes aleatorios de elección de establecimiento escolar y probabilidad de ser negativos según clase latente.

<b>Atributo</b>	<b>Clase 1: PLR</b>	<b>Clase 2: PEE</b>
Proporción de jefes de hogar en el barrio con educación universitaria aprobada (valor entre 0 - 1)	15,8%	65,1%
Interacción para hogares con i.p.c > 400 mil CLP	0,2%	8,9%
Accesibilidad a Oficina	59,9%	42,3%

A nivel de atributos asociados a la localización residencial, estos mantienen el sentido de sus parámetros según lo visto previamente. No obstante, existe una mejor significancia del

atributo asociado a la renta de vivienda de la localización junto con un término negativo. Esto es un resultado apropiado y con mayor sentido a lo visto previamente. De todas formas, hay que considerar las limitaciones de la variable debido a su cálculo. Los parámetros de accesibilidad mantienen su comportamiento, al igual que el atributo asociado al nivel educacional del barrio de la localización. Sin embargo, amerita de todas formas analizar cómo se comporta la aleatoriedad en este atributo, al igual que en el atributo de accesibilidad a oficina, según la Tabla 6-11.

Tabla 6-11: Coeficientes aleatorios de elección de localización residencial y probabilidad de ser negativos según clase latente.

<b>Atributo</b>	<b>Clase 1: PLR</b>	<b>Clase 2: PEE</b>
Accesibilidad a oficina	49,9%	58,6%
Proporción de jefes de hogar en el barrio con educación universitaria aprobada (valor entre 0 - 1)	98,5%	85,1%
Interacción para hogares con i.p.c entre 100 -400 mil CLP	95,2%	57,9%
Interacción para hogares con i.p.c > 400 mil CLP	44,6%	10,0%

De manera similar a los modelos anteriores, el atributo de accesibilidad a oficina posee una alta heterogeneidad con respecto a las preferencias de los hogares, ya que posee un 50% de probabilidad de ser positivo o negativo. Por otro lado, el atributo del nivel educacional del barrio se vuelve con alta probabilidad negativo, independiente del nivel de ingreso para la clase PLR. Esto puede deberse a la composición de la clase y las características que la determinan. Para la clase PEE, el comportamiento es similar a lo discutido previamente.

A grandes rasgos, la incorporación de un enfoque de clases latentes a la modelación permite caracterizar de manera más adecuada las preferencias de los hogares. En ese sentido, el comportamiento entre este modelo y los vistos con anterioridad es similar, con la diferencia que los parámetros se adecúan a las características de cada segmento latente. No obstante, para comparar de manera efectiva ambas clases, se pueden analizar las elasticidades de demanda con respecto a cada atributo en cada elección. Así, se podrá comparar con los modelos previos el efecto que las clases latentes tienen sobre la sensibilidad de ciertos atributos en probabilidad de elección.

### 6.4.3 Cálculo de elasticidades

Al igual que en el análisis realizado a los modelos previos, esta subsección analiza las elasticidades de demanda para ambas elecciones, según la estrategia de cada clase latente. Cabe mencionar, que se analizó la elasticidad con respecto a la probabilidad de la alternativa escogida, y se utilizaron los estimadores puntuales de los coeficientes. Por lo tanto, es importante considerar que puede existir variabilidad sobre los parámetros con términos de error. La Tabla 6-12 presenta las elasticidades para la probabilidad de elección de establecimiento escolar, según cada clase latente.

Tabla 6-12: Elasticidades de demanda para elección de establecimiento escolar, según clase latente.

Clase	Orden	Atributo	Elasticidad
PLR	1	Colegio con educación científico-humanista para hogares con i.p.c. > 400 mil CLP	5,21
	2	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. > 400 mil CLP	3,74
	3	Distancia residencia-colegio para viajes en modo no motorizado	-1,45
	4	Distancia residencia-colegio para viajes en modo "otros"	-1,24
	5	Accesibilidad a oficina	-0,95
	6	Proporción de jefes de hogar en el barrio con educación universitaria	0,72
	7	Distancia residencia-colegio	-0,54
	8	Valor de la mensualidad escolar	-0,22
	9	IVE para hogares con i.p.c. >=100 mil CLP	0,004
Clase	Orden	Atributo	Elasticidad
PEE	1	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. > 400 mil CLP	0,20
	2	IVE para hogares con i.p.c. >=100 mil CLP	-0,16
	3	Colegio con educación científico-humanista para hogares con i.p.c. > 400 mil CLP	0,10
	4	Accesibilidad a oficina	0,05
	5	Proporción de jefes de hogar en el barrio con educación universitaria	-0,01
	6	Valor de la mensualidad escolar	-0,01

Para la mayoría de los atributos, los hogares de mayor nivel de ingreso poseen una mayor elasticidad. Por ejemplo, para la clase que escoge según una estrategia PLR, aquellos hogares con nivel de ingreso sobre los 400 mil CLP priorizan que los establecimientos impartan enseñanza científico humanista y que el nivel educacional del barrio del colegio sea elevado. Es más, valoran estos atributos por sobre la distancia. Por otro lado, la clase PEE, también posee una mayor sensibilidad con respecto a estos atributos, lo que se refleja en que las elasticidades son las mayores en estas variables. Los valores de la elasticidad con respecto al valor de mensualidad y oficina son menos relevantes para ambas clases.

Mientras, para la clase PLR que considera la distancia entre residencia y colegio, valora este atributo de manera significativa para hogares que no poseen ingreso per cápita sobre los 400 mil CLP. Como es de esperar, este atributo es más sensible a medida que los viajes realizados son en modos activos, y menos sensible cuando la distancia se recorre en modos como auto o transporte público. Esto es razonable considerando que son hogares que escogen condicionándose a la localización residencial, de la cual son propietarios y tienen baja posesión vehicular.

Por último, la mayor diferencia se presenta con respecto al atributo del IVE, que se interactúa para aquellos hogares con ingreso mayor a 100 mil CLP. Se aprecia cómo, para la clase PLR que está caracterizada por hogares de menor ingreso, este atributo es poco relevante para la elección. Mientras, para la clase PEE, la probabilidad de elección es muy sensible a este atributo, siendo el segundo más relevante. El resultado va en línea con lo establecido por la literatura, donde existen preferencias con respecto al nivel socioeconómico del cuerpo estudiantil en particular para hogares de la clase media alta. Esto es distinto a lo que mostraban los modelos Logit Mixto que estimaban por separado cada estrategia, por lo que el modelo logra captar heterogeneidad caracterizada por la clase latente. Cabe mencionar de todas formas, que debido a la poca significancia del parámetro para la clase PLR, este análisis es parcial. Desde la perspectiva de la elección de localización residencial, la Tabla 6-13 presenta resultados de las elasticidades con respecto a los atributos que la determinan.

Tabla 6-13: Elasticidades de demanda para elección de localización residencial, según clase latente.

Clase	Orden	Atributo	Elasticidad
PLR	1	Accesibilidad a educación	0,23
	2	Accesibilidad a industria	-0,05
	3	Accesibilidad a comercio	-0,05
	4	Proporción de jefes de hogar en el barrio con educación universitaria	-0,04
	5	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. entre 100 - 400 mil CLP	-0,02
	6	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. > 400 mil CLP	0,01
	7	Valor medio de la renta	-0,01
	8	Accesibilidad a oficina	0,001
Clase	Orden	Atributo	Elasticidad
PEE	1	Distancia residencia-colegio para viajes en modo no motorizado	-1,71
	2	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. > 400 mil CLP	0,80
	3	Distancia residencia-colegio para viajes en modo "otros"	-0,61
	4	Accesibilidad a oficina	-0,40
	5	Distancia residencia-colegio	-0,29
	6	Proporción de jefes de hogar en el barrio con educación universitaria	-0,26
	7	Accesibilidad a industria	-0,12
	8	Accesibilidad a comercio	-0,10
	9	Proporción de jefes de hogar en el barrio con educación universitaria para hogares con i.p.c. entre 100 - 400 mil CLP	-0,05
	10	Valor medio de la renta	-0,05

Se aprecian diferencias significativas para la elección de localización residencial, entre cada segmento latente, lo que no ocurría al modelar las estrategias por separado. Por un lado, hogares que se comportan según una estrategia PLR, tienden a valorar más aspectos vinculados al acceso a oportunidades desde la localización residencial, en vez de aspectos socioeconómicos que ofrece la alternativa. En ese sentido, son más sensibles a oportunidades

de educación, industria o comercio. Después de eso, valoran aspectos vinculados al nivel socioeconómico del barrio. Cabe mencionar que, debido a los valores negativos no esperados de este parámetro, a mayor nivel de ingreso, la elasticidad es menor. No obstante, sería de esperar que los hogares de mayor ingreso tiendan a ser más sensibles frente a este atributo. Atributos de renta y accesibilidad a la oficina tienden a ser poco relevantes para la decisión, en comparación con otros aspectos. Aunque, hay que considerar que la construcción de la variable de precio es poco confiable desde la perspectiva del investigador, y la accesibilidad a oficina cuenta con una considerable heterogeneidad no evaluada en este análisis.

Mientras, los hogares que pertenecen a la clase PEE valoran de manera significativa la distancia entre la residencia y el colegio, particularmente aquellos que viajan en modos no motorizados al establecimiento. Pero, valoran también de manera significativa el nivel educacional del barrio de este, siendo de las elasticidades más importantes para hogares de ingreso per cápita mayor a 400 mil CLP. Mientras, para otras categorías de ingreso, valoran de manera mixta aspectos socioeconómicos como accesibilidades a oficina, industria o comercio. El atributo menos elástico es el valor de la renta.

Bajo este análisis, se puede destacar el aporte que las clases latentes realizan en la caracterización de las preferencias. Al estimar las estrategias, como modelos independientes, se observó cómo el orden de elasticidades comportaba de manera similar en cada estrategia. No obstante, al incorporar características de los hogares que determinaron la probabilidad de elegir según cada estrategia, se apreciaron diferencias.

A grandes rasgos, las principales conclusiones de estas diferencias radican en que los hogares de mayor nivel de ingreso tienden a valorar más aspectos socioeconómicos de las elecciones en comparación con los hogares de menor ingreso, independiente de la elección. Al mismo tiempo, la clase PLR tiende a ser más sensible a las oportunidades, y el acceso a ellas, desde su localización residencial. Mientras que la clase PEE es más elástica a variaciones en el nivel socioeconómico del cuerpo estudiantil de las alternativas escolares. No obstante, un atributo transversal a ambas estrategias de decisión es que independiente del orden en que se prioricen las elecciones, la distancia que vincula ambas alternativas es relevante al momento de elegir.

## 6.5 Comparación de modelos

A partir de los análisis descritos, se pueden comparar brevemente los modelos expuestos en cuanto a la calidad de los modelos, y el aporte estadístico que tiene entregar mayor complejidad. Dado que el modelo MNL posee diferencias significativas, con los modelos mixtos y de clases latentes propuestos, se aparta del análisis. Esto es razonable desde la perspectiva que su estimación tenía un propósito explicativo, de comprender en primera instancia el fenómeno, sin necesariamente captar correlación entre elección o definir priorizaciones a través de estrategias de elección.

Un test de razón de verosimilitud logra explicar si la mayor complejidad del modelo de clases latentes aporta de manera significativa a la explicación del fenómeno. Comparando este modelo con el mejor modelo, en términos de log-verosimilitud y significancia de los parámetros, del enfoque Logit Mixto (i.e. aquel que adopta una estrategia PEE), se observa que el modelo de clases latentes es una versión no restringida del otro. En particular, hay 23 parámetros no restringidos (19 correspondientes al modelo de elección y 4 al modelo de pertenencia de clase). Aun así, el test LR es igual a 580, muy por sobre un test  $\chi^2_{90\%}$ . Por lo tanto, es razonable concluir que el modelo de clases latentes es un modelo más complejo, y estadísticamente más explicativo que los modelos Logit Mixto con una sola estrategia de elección.

No obstante, y más allá de los resultados estadísticos. Este modelo posee mayores aportes en materia de poder caracterizar a los hogares en torno a sus preferencias y sus estructuras. Logra poder definir, de manera significativa, qué hogares tienden a priorizar una elección por sobre otra. En ese sentido, posee un aporte significativo a comprender el fenómeno de elección de establecimiento escolar y localización residencial, apuntando así a lograr los objetivos propuestos previamente.

## 7. CONCLUSIONES

La literatura de elección escolar en Chile es abundante y se ha dedicado de manera extensa a comprender las preferencias de los hogares, en un contexto de elección marcado por estratificación social y heterogeneidad en las preferencias. Esta ha estudiado cómo los hogares compensan aspectos académicos y de costo de los colegios con atributos socioeconómicos de sus cuerpos estudiantiles y/o de los barrios en que se encuentran localizados. La literatura ha identificado también el rol que juega la distancia entre la vivienda y el colegio en la probabilidad de elección. Por último, ha entregado evidencia respecto al vínculo existente entre esta elección según las características de los hogares y sus expectativas. Todo esto, en un contexto de gran heterogeneidad en las preferencias. Muchas veces caracterizable, otras veces sólo capaz de ser detectada.

Por otro lado, los modelos de elección conjunta han sido estudiados desde hace décadas por la literatura de transporte. En particular, se han estudiado variadas elecciones de transporte y sus interacciones, demostrando los vínculos existentes entre ellas. Por nombrar algunas, se ha estudiado el vínculo entre elección modal, posesión vehicular o lugar de trabajo. Estas, muchas veces se han vinculado con la localización residencial y se ha analizado el rol que esta elección cumple sobre el comportamiento de los viajeros en la ciudad. No obstante, a la luz de la revisión realizada, la literatura de modelos de elección discreta no ha profundizado en el vínculo existente entre la elección de localización residencial y la elección de establecimiento escolar. Esta interacción es en particular atractiva, debido a que los hogares viajan cada día para que sus hijos e hijas estudien. Por lo tanto, se pierde una gran oportunidad de comprender cómo las preferencias de los hogares en torno a estas elecciones pueden determinar la estructura de la ciudad. Además, la singularidad del sistema educacional chileno entrega una oportunidad única de entender el comportamiento de los hogares frente a la elección a la que se enfrentan.

Esta investigación propone llenar ese espacio de investigación estimando en conjunto ambas elecciones y captando factores observados y no observados que las determinan. Además, incorpora un enfoque de clases latentes, estimando la probabilidad de aplicar distintas estrategias de elección, que dependen de las características de los hogares. A través de los

modelos estimados, desde una estructura Logit Multinomial hasta un modelo Logit Mixto con clases latentes se concluyen resultados que confirman las discusiones ya esgrimidas por otros autores, al igual que otros resultados que aportan a la comprensión de estas elecciones.

Primero, se observa cómo para todos los modelos estimados, el parámetro de distancia entre el colegio y la localización residencial resultó ser sistemáticamente significativo y negativo. Es más, al analizar las elasticidades de demanda con respecto a ambas elecciones, se corroboró que en aquellos casos donde la distancia era un atributo de la función de utilidad, los hogares tenían una alta elasticidad respecto del atributo, tanto para la elección de localización como de colegio. Estos resultados son confirmativos por lo expuesto por estudios como el del Chumacero et al. (2011).

Segundo, se muestra que los hogares tienden a compensar, y en casos a priorizar, aspectos socioeconómicos de los colegios, por sobre aspectos académicos. No se logró estimar de manera significativa coeficientes asociados a resultados de pruebas estandarizadas, niveles de subvención, docentes por alumno, u otra variable asociada a la calidad escolar. En cambio, sí resultaron significativos atributos respecto del nivel socioeconómico del colegio y de su entorno o barrio, en particular para hogares de ingreso alto. Esto va en línea con lo expuesto por autores como Elacqua et al. (2006) o Bellei et al. (2020).

Sumado a estos factores, los modelos estimados logran captar factores no observados que determinan las preferencias sobre ciertos atributos tanto de la elección de establecimiento escolar como de localización residencial. A través de términos de error aleatorios asociados a atributos de las dos elecciones, se lograron captar los efectos de estos factores. En particular, estos términos están asociados al nivel educacional del barrio tanto de la localización residencial como del colegio, y vinculados a atributos de accesibilidad a la oficina desde ambas localizaciones. El primer factor se relaciona con lo discutido por Bellei et al. (2020) y Méndez y Gayo (2019), donde se plantea cómo variables socioeconómicas de los colegios y de las residencias están asociadas al sentido de pertenencia a una comunidad y a las expectativas que se tienen sobre los beneficios sociales, culturales o económicos que entregan ambas elecciones. Esto, en un contexto de alta heterogeneidad en las preferencias y con un importante enfoque hacia la clase media alta chilena, que tiende a poseer altas

expectativas respecto de los beneficios sociales, profesionales y culturales que le pueden entregar ambas elecciones. Mientras, el segundo factor es razonable debido a que autores como Waddell et al. (2007) han estudiado en profundidad la elección de localización residencial y de trabajo, sumado a que es razonable esperar que el viaje al colegio esté encadenado con actividades laborales de los apoderados. Así, el modelo estimado capta factores no observados que vincula a estas elecciones con una tercera, de lugar de trabajo. En ese sentido, este estudio logra medir e identificar estos aspectos que generan correlación entre elecciones.

Por último, otra conclusión relevante de este estudio corresponde al aporte que hace en caracterizar a los hogares según sus preferencias y las estructuras de elección que poseen. El modelo de clases latentes modela dos estrategias de elección. Aquellos hogares que priorizan la localización residencial por sobre la de establecimiento (PLR) y aquellos hogares que priorizan el establecimiento escolar por sobre la elección de localización (PEE), donde existió una tendencia a pertenecer más al segundo segmento según la muestra. Los resultados demuestran que una estrategia PEE se adecúa a hogares de mayor ingreso, con mayor posesión vehicular (y que por ende están dispuestos a viajar más lejos al colegio), y mayor movilidad residencial dado que no son propietarios de sus casas. Así, a través del cálculo de las elasticidades de demanda de ambas elecciones, los resultados exponen que los hogares PEE tienden a valorar con mayor importancia aspectos socioeconómicos del establecimiento, en comparación con los hogares PLR. También, los hogares PLR valoran la capacidad del sistema de transporte con respecto a las oportunidades presentes en torno a la localización residencial. También, los hogares de mayor ingreso tienden a ser más sensibles a atributos socioeconómicos, para ambas estructuras de decisión. Además, en ambas clases los factores no observados que correlacionan las elecciones son significativos.

En este sentido, es razonable asumir que la hipótesis de trabajo planteada es válida. A través de términos de error aleatorios, se estimaron factores no observados que vinculan a ambas elecciones, demostrando que existe correlación entre ellas. Además, se validó que existe heterogeneidad en las preferencias de los hogares con respecto a estas elecciones. Por un lado, existe heterogeneidad no caracterizable, medida a través de estos factores no observados. Por el otro, existe heterogeneidad caracterizable, medida tanto a través de las

variaciones sistemáticas presentes en las funciones de utilidad, como a través de un enfoque de clases latentes para diferentes estructuras de decisión. Se demostró también, mediante un análisis de las elasticidades, que esta heterogeneidad se refleja en la sensibilidad de la probabilidad de elección con respecto a los atributos, según las preferencias de los hogares. Así, es razonable concluir que los objetivos propuestos por el estudio de modelar en conjunto ambas elecciones para comprender las preferencias de los hogares fue cumplido con éxito.

De todas formas, y a pesar de los aportes que la investigación entrega, es importante reconocer las limitaciones que esta puede tener. En particular se destacan dos. La primera refiere a la metodología que georreferencia las coordenadas imputadas en la encuesta con las de las bases de datos de los establecimientos escolares. Esta elige el establecimiento que se ajusta mejor a la coordenada imputada. Sin embargo, esto puede inducir a errores en la elección determinada como la observada, ya que existen casos donde se puede escoger un establecimiento equivocado, entre varios otros que se encuentren cercanos entre sí. A futuro, disponer de una base de datos clara que indique la elección de establecimiento escolar observada por el código identificador de este, por ejemplo, podría ser útil para una estimación más precisa de los datos.

La segunda limitación está vinculada a cuestionar la verdadera capacidad de elección de los hogares respecto de los establecimientos escolar o la localización residencial. Gómez et al. (2012) menciona cómo pueden existir falta de información por parte de los hogares respecto a la elección de colegios, mientras que Méndez y Gayo (2019) recolectan información que indica que, en la mayoría de las comunas de Santiago, alrededor de la mitad de los hogares no tuvo capacidad de elección sobre la residencia donde vive ni el establecimiento educacional. Poder considerar este aspecto en la elección puede ser trascendental para captar de manera adecuada las preferencias de los hogares. Es más, no tomar en cuenta estas restricciones puede ser la causa de los efectos endógenos sobre algunos atributos, estimando parámetros con signos errados. Considerar a futuro estas limitaciones puede ser útil para obtener conclusiones valiosas y más confiables.

Para finalizar, cabe reflexionar sobre la contribución que esta investigación puede hacer a la elaboración de políticas públicas relacionadas con el mercado inmobiliario y el sistema

escolar chileno. El sistema de educación en Chile ha sido blanco de críticas desde hace décadas, donde tanto la literatura como la ciudadanía ha levantado la voz respecto a los problemas que presenta y las desigualdades que genera. En ese sentido, el nuevo Sistema de Admisión Escolar (MINEDUC, 2021) pretende acortar brechas y entregar mayor capacidad de elección a los hogares. Correa et al. (2019) explica en detalle su funcionamiento. Este recibe en orden de preferencias, los establecimientos elegidos por cada hogar y estudiante. En caso de tener un cupo disponible, el estudiante es asignado. Mientras, en caso de que la demanda sea mayor a la oferta de matrículas en el establecimiento, rigen los siguientes criterios de prioridad: que el o la aplicante tengan un hermano o hermana estudiando en el colegio, que al menos 1 apoderado trabaje en el colegio, que el estudiante sea ex alumno del establecimiento sin haber sido expulsado. Además, existen otros criterios asociados a cuotas mínimas que priorizan a estudiantes vulnerables, con buen rendimiento, o con necesidades especiales. Sin embargo, la única instancia donde la localización residencial juega un rol en la elección es en la situación que el aplicante no haya sido asignado a ningún establecimiento. En ese caso, se le asigna el colegio más cercano a su hogar.

Al observar la literatura al igual que los resultados de esta investigación, la distancia del hogar a la residencia se presenta como una variable determinante en la elección de los hogares. No obstante, esta no es considerada en el sistema que define las admisiones de los estudiantes a cada colegio. Similar situación ocurre con aspectos de inclusión y diversidad socioeconómica. Cabe entonces reflexionar sobre cómo a partir de las políticas públicas que se implementan, se logran reflejar las preferencias de los hogares y se mejora la distribución socioespacial de la ciudad. Desde una perspectiva, es cierto que estas preferencias deberían verse reflejadas a través de la lista de establecimientos declarada por cada familia. Si el estudiante queda matriculado dentro de sus primeras preferencias, se esperaría que el hogar haya compensado algunos atributos por sobre otros. Sin embargo, en aquellas ocasiones donde la elección no es ejercida de manera efectiva, es interesante cuestionar de qué manera pueden existir mecanismos que reflejen de mejor forma estas preferencias. Así, comprender la elección de establecimiento escolar y el vínculo que tiene con la elección de localización residencial es un primer paso para generar una mejor distribución de la ciudad e incluso reducir las desigualdades que la aquejan.

## BIBLIOGRAFIA

Agencia de Calidad de la Educación (2006). Base de Datos SIMCE 2006.

Alonso, W. (1964). *Location and Land Use* Harvard University Press. *Cambridge, Mass.*

Astroza, S., Garikapati, V. M., Pendyala, R. M., Bhat, C. R., y Mokhtarian, P. L. (2019). Representing heterogeneity in structural relationships among multiple choice variables using a latent segmentation approach. *Transportation*, 46(5), 1755-1784.

BCN (2015). Ley 20845 de inclusión escolar que regula la admisión de los y las estudiantes, elimina el financiamiento compartido y prohíbe el lucro en establecimientos educacionales que reciben aportes del Estado. URL: <http://bcn.cl/2f8t4>

BCN (2020). Índice de Vulnerabilidad Escolar en la comuna de Cerro Navia. URL: [https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/10221/28517/2/BCN\\_IVE\\_Cerro\\_Navia.pdf](https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/10221/28517/2/BCN_IVE_Cerro_Navia.pdf)

Bellei, C., Orellana, V., y Canales Cerón, M. (2020). Elección de escuela en la clase alta chilena. Comunidad, identidad y cierre social.

Ben-Akiva, M. (1973). *Structure of Passenger Travel Demand Models*. PhD Thesis, Massachusetts Institute of Technology. <https://dspace.mit.edu/handle/1721.1/14790>

Ben-Akiva, M. y Stephen R. Lerman. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand* (3rd ed.). The MIT Press.

Chumacero, R. A., Gómez, D., y Paredes, R. D. (2011). I would walk 500 miles (if it paid): Vouchers and school choice in Chile. *Economics of Education Review*, 30(5), 1103-1114.

Correa, J., Epstein, R., Escobar, J., Rios, I., Bahamondes, B., Bonet, C. y Epstein, B. (2019). School choice in Chile. En *Proceedings of the 2019 ACM Conference on Economics and Computation* (pp. 325-343).

DEMRE (2006). Base de Datos Resultados PSU 2006.

Elacqua, G. (2012). The impact of school choice and public policy on segregation: Evidence from Chile. *International Journal of Educational Development*, 32(3), 444-453.

Elacqua, G., Schneider, M., y Buckley, J. (2006). School choice in Chile: Is it class or the classroom?. *Journal of Policy Analysis and Management*, 577-601.

Gallego, F., y Sapelli, C. (2007). El financiamiento de la educación en Chile: una evaluación. *Pensamiento Educativo, Revista de Investigación Latinoamericana (PEL)*, 40(1), 263-284.

- Gayo, M., Otero, G., y Méndez, M.L. (2019). Elección escolar y selección de familias: reproducción de la clase media alta en Santiago de Chile. *Revista Internacional de Sociología*, 77(1):e120. <https://doi.org/10.3989/ris.2019.77.1.17.310>
- Gómez, D., Chumacero, R. A., y Paredes, R. D. (2012). School choice and information. *Estudios de Economía*, 39(2), 143-157.
- Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of planners*, 25(2), 73-76.
- Hurtubia, R., Gallay, O., y Bierlaire, M. (2010). Attributes of households, locations and real estate markets for land use modelling. *SustainCity Deliverable*, 2, 1-27.
- INE. (2018). Resultados Censo 2017.
- Lerman, S. R. (1976). Location, housing, automobile ownership, and mode to work: a joint choice model. *Transportation Research Record*, 610, 6-11.
- McFadden, D. (1974). The measurement of urban travel demand. *Journal of public economics*, 3(4), 303-328.
- Méndez, M. L., y Gayo, M. (2018). *Upper middle class social reproduction: Wealth, schooling, and residential choice in Chile*. Springer.
- Miller, E. J., y Lerman, S. R. (1981). Disaggregate modelling and decisions of retail firms: a case study of clothing retailers. *Environment and Planning A*, 13(6), 729-746.
- MINEDUC (2006). Bases de Datos de Docentes, Matrículas y Subvención de Establecimientos Escolares 2006.
- MINEDUC (2010). Bases de Datos Índice de Vulnerabilidad Escolar 2010.
- MINEDUC. (2014). Directorio Establecimientos Educativos 2014. <http://datos.mineduc.cl/dataviews/235565/vista-directorio-de-establecimientos-2014/>
- MINEDUC (2021). Sistema de Admisión Escolar. Sección de la web Qué es. Recuperado el 28 de febrero de 2021 de: <https://www.sistemadeadmisionescolar.cl/>
- Paleti, R., Bhat, C. R., y Pendyala, R. M. (2013). Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. *Transportation research record*, 2382(1), 162-172.
- Pinjari, A. R., Pendyala, R. M., Bhat, C. R., y Waddell, P. A. (2011). Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation*, 38(6), 933-958.

SECTRA. (2014). Actualización y recolección de información del sistema de transporte urbano, Etapa IX: Encuesta Origen Destino Santiago 2012.

SII. (2014). Catastro de Bienes Raíces del Servicio de impuestos Internos.  
[https://zeus.sii.cl/cvc\\_cgi/dfmun/dfmun\\_repGobierno.cgi](https://zeus.sii.cl/cvc_cgi/dfmun/dfmun_repGobierno.cgi)

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Valenzuela, J. P., Bellei, C., y Ríos, D. D. L. (2014). Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy*, 29(2), 217-241.

Waddell, P. (1993). Exogenous workplace choice in residential location models: Is the assumption valid?. *Geographical Analysis*, 25(1), 65-82.

Waddell, P., Bhat, C., Eluru, N., Wang, L., y Pendyala, R. M. (2007). Modeling interdependence in household residence and workplace choices. *Transportation Research Record*, 2003(1), 84-92.

Weisbrod, G., Lerman, S. R., y Ben-Akiva, M. (1980). Tradeoffs in residential location decisions: Transportation versus other factors. *Transport Policy and Decision Making*, 1(1), 13-26.

Zimmerman, S. D. (2019). Elite colleges and upward mobility to top jobs and top incomes. *American Economic Review*, 109(1), 1-47.

**ANEXOS**

## ANEXO A : MODELO LOGIT JERÁRQUICO

El objetivo de este anexo es profundizar respecto al modelo Logit Jerárquico utilizado, la estructura planteada y las razones por las cuales se descartó. Se modelaron dos estructuras jerárquicas, una para cada estrategia de elección: hogares que priorizan la localización residencial y luego escogen establecimiento escolar condicionado a la localización, y hogares que priorizan la elección de establecimiento escolar por sobre la de localización residencial. La Figura A-1 presenta un árbol de decisiones con la primera estrategia. Cada hogar  $q$  posee 20 alternativas  $r$  y 20 alternativas  $c$  sobre las cuales se estima el modelo. Estas, al igual que el modelo multinomial componen pares de alternativas (i.e. 400 alternativas evaluadas).

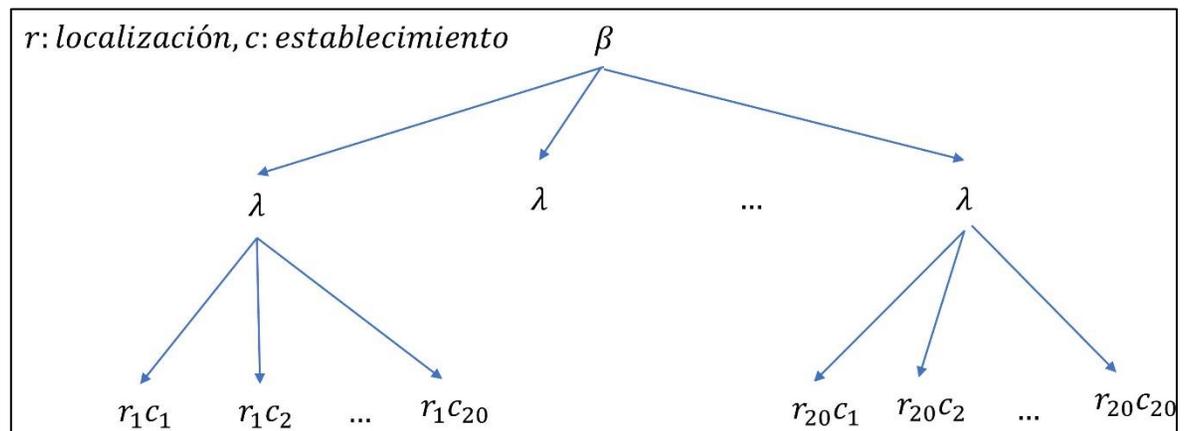


Figura A-1: Modelo Logit Jerárquico con una estrategia de decisión priorizadora de localización residencial.

De la figura se desprende cómo, para una elección que jerarquiza según la alternativa de localización residencial, estas se agrupan en cada nido según la alternativa de localización que comparten. Es razonable pensar que, debido a que las alternativas no son etiquetadas, los parámetros de cada nido poseen igual valor.

Producto de los términos de error que incorpora un modelo Logit Jerárquico y las varianzas definidas por las distribuciones Gumbel, los parámetros de escala de cada nido,  $\lambda$ , deben ser mayores al parámetro de escala  $\beta$ , definida por los términos de error i.i.d Gumbel(0,  $\beta$ )

presentes en las funciones de utilidad de cada alternativa evaluada. En Biogeme, esto se evalúa comparando significancia estadística del parámetro con respecto a  $\beta = 1$ . Sin embargo, al estimar este modelo, para ambas estructuras de decisión por separado y en conjunto a través de clases latentes, se obtuvieron valores estimados menores a 1 y significativamente distintos. Esto indica que, frente a esta definición de estructura, las varianzas son negativas. Por lo tanto, la especificación es incorrecta y no debería utilizarse.