



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

CARACTERIZACIÓN DEL CONJUNTO DE CONSIDERACIÓN EN ELECCIÓN DE
RUTA

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL

GUILLERMO NICOLÁS VILLALOBOS ZAID

PROFESOR GUÍA:
ANGELO GUEVARA CUE

MIEMBROS DE LA COMISIÓN:
MARCELA MUNIZAGA MUÑOZ
RICARDO HURTUBIA GONZÁLEZ

Este trabajo ha sido parcialmente financiado por el proyecto Fondecyt-Chile No. 1150590

SANTIAGO DE CHILE
2018

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL
POR: GUILLERMO NICOLÁS VILLALOBOS ZAID
FECHA: 2018
PROF. GUÍA: ANGELO GUEVARA CUE

CARACTERIZACIÓN DEL CONJUNTO DE CONSIDERACIÓN EN ELECCIÓN DE RUTA

En elecciones discretas, el conjunto de consideración contiene las alternativas entre las cuales un tomador de decisión realiza una elección. Uno de los supuestos fundamentales al modelar elecciones discretas es que el analista conoce dicho conjunto. Este supuesto es fácilmente cuestionable, sobre todo cuando el conjunto universal de alternativas es muy grande, como es en elección de rutas. Manski (1977) dio una solución a este problema desde el punto de vista teórico, proponiendo un enfoque donde se utiliza el conjunto de consideración como latente. Prato (2009) revisa una serie de métodos que han intentado resolver este problema de forma práctica. Entre estos enfoques, destacan aquellos en los que se eligen funciones de costo generalizado y se calculan las mejores K rutas para generar el conjunto. Sin embargo, independiente del método utilizado, si se viola algún supuesto el modelo no será capaz de representar la realidad, entregando estimadores inconsistentes y errores en la predicción.

Esta memoria realiza tres contribuciones orientadas a entender el proceso de generación del conjunto de consideración. La primera se basa en la utilización de una simulación de Monte Carlo para explorar la robustez de varios métodos prácticos utilizados para construir el conjunto de consideración. La segunda consiste en el análisis de tres métodos para recolectar datos sobre el conjunto de consideración: datos pasivos, encuesta online y encuesta sobre mapa. Finalmente, la tercera contribución desarrolla un experimento práctico de preferencias declaradas donde los participantes se ven enfrentados a una imitación del proceso de generación del conjunto de consideración. Esto permite estudiar características del conjunto y probar la robustez de algunos posibles métodos utilizados para construirlo.

En el capítulo basado en simulación de Monte Carlo se muestra primero que los métodos tradicionales generan grandes sesgos en la estimación. Además, se prueba un nuevo método generador del conjunto que utiliza una serie de elecciones experimentadas, resultando en buenas estimaciones y mostrando cierto potencial práctico, dada la disponibilidad de datos de hoy en día mediante datos pasivos. En el capítulo referido a la recolección de datos, se muestra evidencia preliminar que sugiere que los datos pasivos, si bien son atractivos debido a que muestran las preferencias reveladas de las personas, tienen la limitación de requerir de mucho procesamiento para dar cuenta de errores de medición. Por otro lado, los datos obtenidos a partir de encuestas mostraron ser factibles para utilizarse en un contexto exploratorio. Estos datos permitieron describir el tamaño del conjunto de consideración, estudiar las heurísticas que la gente dice utilizar para generar el conjunto, modelar el proceso de consideración e identificar las variables que afectan dicho proceso. Finalmente, se utilizan estos datos para estudiar el sesgo que causan los distintos métodos para generar el conjunto de consideración, corroborando que los métodos deterministas generan sesgos muy grandes, estimando incluso parámetros con signo contrario a los verdaderos.

Agradecimientos

Quisiera agradecer en primer lugar a mi familia, quienes me han apoyado para poder estudiar y avanzar durante toda mi vida. Un especial agradecimiento a mi hermana Maca, quien ha sido mi compañera de estudio, cursos, juegos, viajes y mucho más. Además, aunque no sepa leer, agradezco a nuestra perrita Amy que nos llena de alegría y ha estado con nosotros desde los primeros años de universidad. También quisiera agradecer a todos mis amigos, compañeros del colegio y de la universidad, especialmente a aquellos que me han acompañado por un largo tiempo y han sido muy buena compañía.

Agradezco además, a todos los profesores que han aportado en mi formación, sobre todo a aquellos que han ido un poco más allá de las clases y que han hecho del proceso de aprendizaje una buena experiencia. En particular le doy las gracias a todos los profesores del departamento de Transporte de la Universidad de Chile, especialmente a mi profesor guía, con quien llevamos a cabo este gran proyecto para mi titulación. También agradezco a todos los funcionarios de la universidad, cuyo trabajo es fundamental para que todos trabajemos en armonía.

Tabla de contenido

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	2
1.3. Metodología	3
1.4. Estructura de la memoria	3
2. Revisión bibliográfica	5
2.1. Introducción	5
2.2. Modelación de elecciones discretas incorporando el conjunto de consideración	5
2.3. Estudios sobre el conjunto de consideración en la literatura de marketing . .	7
2.4. Modelación de elección de ruta	8
3. Estudio del impacto de utilizar distintos conjuntos de consideración mediante método de Monte Carlo	9
3.1. Introducción	9
3.2. Generación de datos	10
3.3. Métodos generadores del conjunto de consideración verdadero	13
3.4. Métodos generadores del conjunto de consideración supuesto	19
3.5. Análisis y resultados de la simulación de Monte Carlo	21
4. Análisis crítico de tres métodos para recolectar datos sobre el conjunto de consideración	28
4.1. Introducción	28
4.2. Datos pasivos	28
4.3. Encuesta de preferencias declaradas online	37
4.3.1. Descripción de la encuesta	37
4.3.2. Análisis estadístico descriptivo de los datos	42
4.3.3. Discusión y extensiones	44
4.4. Encuesta de preferencias declaradas sobre mapa	45
5. Caracterización del proceso de consideración mediante datos de preferencias declaradas	48
5.1. Introducción	48
5.2. Tamaño del conjunto de consideración	49
5.2.1. Utilizando encuesta de preferencias declaradas online	49
5.2.2. Utilizando encuesta de preferencias declaradas sobre mapa	53
5.3. Heurísticas declaradas	53

5.4. Modelo binario de consideración	57
5.5. Modelo de elección	62
5.6. Análisis comparativo de métodos generadores del conjunto de consideración .	66
5.6.1. Modelo utilizando el conjunto de consideración verdadero	66
5.6.2. Modelos utilizando otros conjuntos de consideración	67
5.6.3. Comparación de modelos	72
6. Conclusiones	77
6.1. Introducción	77
6.2. Conclusiones generales	78
6.3. Recomendaciones metodológicas	78
6.4. Extensiones	79
Bibliografía	80
A. Tablas adicionales	82

Índice de tablas

3.1. Valores de los parámetros utilizados para generar las elecciones en la simulación de Monte Carlo.	10
3.2. Parámetros de modelo de consideración.	14
3.3. Valores de los parámetros a priori.	20
3.4. Listado de métodos generadores del conjunto de consideración supuesto.	23
3.5. Listado de métodos generadores del conjunto de consideración verdadero.	23
3.6. Análisis de Monte Carlo de los métodos generadores del conjunto de consideración supuesto, cuando la heurística original es “ Aleatorio ”.	24
3.7. Análisis de Monte Carlo de los métodos generadores del conjunto de consideración supuesto, cuando la heurística original es “ Todas las alternativas ”.	25
3.8. Análisis de Monte Carlo de los métodos generadores del conjunto de consideración supuesto, cuando la heurística original es “ Eliminación por aspecto ”.	26
3.9. Análisis de Monte Carlo de los métodos generadores del conjunto de consideración supuesto, cuando la heurística original es “ Modelo logit binario de consideración ”.	27
5.1. Resumen de Anova para explicar las diferencias en el tamaño del conjunto de consideración declarado, valor p de cada variable.	52
5.2. Heurísticas mostradas.	55
5.3. Resumen de Anova para explicar la representatividad de las heurísticas por grupo, valor p de cada variable.	56
5.4. Modelo de consideración con todas las variables.	60
5.5. Modelo de consideración definitivo.	61
5.6. Porcentaje de la utilidad sistemática que aporta cada variable sin contar la constante específica utilizada para dar cuenta de la utilidad base con la que se compara en un logit binario.	62
5.7. Modelo de elección utilizando el ranking de alternativas.	64
5.8. Modelo de combinado de consideración y elección sin restringir los parámetros comunes.	65
5.9. Modelo de combinado de consideración y elección restringiendo los parámetros comunes.	66
5.10. Modelo de elección sin utilizar el ranking de alternativas, conjunto de consideración utilizado corresponde a CV.	67
5.11. Parámetros a priori para método KRM_E.	68

5.12. Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante CC a partir de las alternativas mostradas.	69
5.13. Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante KRM_V a partir de las alternativas mostradas.	70
5.14. Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante KRM_E a partir de las alternativas mostradas.	70
5.15. Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante CLB a partir de las alternativas mostradas.	70
5.16. Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante CC a partir del universo de alternativas.	71
5.17. Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante KRM_V a partir del universo de alternativas.	72
5.18. Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante KRM_E a partir del universo de alternativas.	72
5.19. Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante CLB a partir del universo de alternativas.	72
5.20. Valores de test de especificación de razón de verosimilitud, método CV contra el resto, cuando se utilizan las alternativas mostradas para generar el conjunto de consideración supuesto.	74
5.21. Valores de test de especificación de razón de verosimilitud, método CV contra el resto, cuando se utiliza el universo de alternativas para generar el conjunto de consideración supuesto.	75
A.1. Resumen de Estadísticas de Datos Generales de los Participantes.	83
A.2. Resumen de Estadísticas de Datos del Hogar de los Participantes.	84
A.3. Resultados porcentuales de las respuestas sobre las heurísticas utilizadas para generar el conjunto de consideración	85

Índice de ilustraciones

3.1.	Datos de tiempo generados para cada alternativa y observación en una iteración arbitraria durante la simulación de Monte Carlo.	11
3.2.	Datos de costo generados para cada alternativa y observación en una iteración arbitraria durante la simulación de Monte Carlo.	12
3.3.	Porcentajes de elecciones de cada alternativa de la simulación, cuando la heurística original es “ Aleatorio ”.	15
3.4.	Porcentajes de elecciones de cada alternativa de la simulación, cuando la heurística original es “ Todas las alternativas ”.	16
3.5.	Porcentajes de elecciones de cada alternativa de la simulación, cuando la heurística original es “ Eliminación por aspecto ”.	17
3.6.	Porcentajes de elecciones de cada alternativa de la simulación, cuando la heurística original es “ Modelo logit binario de consideración ”.	18
3.7.	Comparación en la estimación de parámetros utilizando distintos métodos para generar el conjunto de consideración supuesto, cuando la heurística original es “ Aleatorio ”.	24
3.8.	Comparación en la estimación de parámetros utilizando distintos métodos para generar el conjunto de consideración supuesto, cuando la heurística original es “ Todas las alternativas ”.	25
3.9.	Comparación en la estimación de parámetros utilizando distintos métodos para generar el conjunto de consideración supuesto, cuando la heurística original es “ Eliminación por aspecto ”.	26
3.10.	Comparación en la estimación de parámetros utilizando distintos métodos para generar el conjunto de consideración supuesto, cuando la heurística original es “ Modelo logit binario de consideración ”.	27
4.1.	Cómo acceder a la cronología en Google Maps.	30
4.2.	Cómo acceder al menú de herramientas para descargar archivo *.kml en Google Maps.	31
4.3.	Rutas Google Maps y rutas de datos pasivos del “Usuario Caminata”.	34
4.4.	Rutas Google Maps y rutas de datos pasivos del “Usuario Transporte Público”.	35
4.5.	Rutas Google Maps y rutas de datos pasivos del “Usuario Auto”.	36
4.6.	Ejemplo de pregunta en el ejercicio de consideración.	39
4.7.	Ejemplo de pregunta en el ejercicio de elección.	40
4.8.	Preguntas sobre heurísticas utilizadas para generar el conjunto de consideración.	41
4.9.	Comparación de sexo entre Encuesta PD y EOD 2012.	42
4.10.	Comparación de edad entre Encuesta PD y EOD 2012.	43

4.11. Comparación del tamaño del hogar entre Encuesta PD y EOD 2012.	43
4.12. Comparación de ingreso entre Encuesta PD y EOD 2012.	44
4.13. Mapa presentado para dibujar las rutas que el encuestado considera, 48x33 <i>cm</i> ²	46
4.14. Respuestas de encuesta sobre mapa recolectadas.	46
5.1. Porcentaje de personas representadas por cada heurística presentada.	55
5.2. Comparación de modelos de elección utilizando distintos métodos de generación del conjunto de consideración supuesto, utilizando las alternativas mostradas.	73
5.3. Comparación de modelos de elección utilizando distintos métodos de generación del conjunto de consideración supuesto, utilizando el universo de alternativas.	75

Capítulo 1

Introducción

1.1. Motivación

En elecciones discretas un individuo considera múltiples alternativas y se ve enfrentado a elegir una. Dichas alternativas son las que dan forma al conjunto de consideración, como se explica en Hauser (2014), la existencia de este conjunto se justifica debido que, a pesar de que existe un beneficio asociado a considerar un mayor número de alternativas (debido a que esto aumenta la máxima utilidad esperada), también existe un costo de considerar una alternativa extra, causado por varios factores como el agotamiento o el límite cognitivo de las personas.

Uno de los supuestos fundamentales al modelar elecciones discretas es que el analista conoce el conjunto de consideración de cada individuo. Pero este supuesto resulta fácilmente cuestionable, sobre todo en contextos donde el conjunto universal de alternativas es muy grande, tal como es en el caso de elección de rutas. El problema es que violar este supuesto puede causar sesgos en la estimación de los parámetros en modelos de elección.

A lo largo de los años se han estudiado diversas formas de lidiar con este problema. Por un lado, se tienen soluciones más bien analíticas, como muestra Manski (1977), modelando el conjunto de consideración como latente. Mientras que por otro lado, se han generados distintos métodos prácticos, como los revisados en Prato (2009), que se enfocan en elección de ruta. Cada uno de estos enfoques tiene ventajas y desventajas, que serán revisadas junto a más literatura en el capítulo siguiente.

En elección de rutas, la mayor parte de los trabajos realizados se enfocan en resolver el problema desde el punto de vista del analista, asumiendo que se ignora lo que realmente pasa. En marketing, existen diversos esfuerzos por caracterizar el verdadero proceso de consideración, por ejemplo, Brown y Wildt (1992) intentan identificar el tamaño del conjunto para distintos tipos de productos, explorando la mejor forma de preguntar el conjunto, por otro lado, Hauser (2014) estudia la heurística que utilizan las personas, resumiendo varias formas de identificar el conjunto de consideración.

1.2. Objetivos

El objetivo principal de esta memoria es caracterizar el conjunto de consideración en elección de ruta, con el propósito de dar cuenta con el problema que presenta su construcción y comprensión. Para ello se plantean una serie de objetivos que son presentados a continuación:

1. Estudiar el impacto en la estimación de parámetros al utilizar conjuntos de consideración distintos del verdadero.
2. Analizar críticamente tres métodos para recolectar datos sobre el conjunto de consideración.
3. Identificar posibles heurísticas que las personas utilizan para generar el conjunto de consideración.
4. Caracterizar el proceso de consideración de alternativas mediante un experimento de preferencias declaradas sobre elección de ruta.

El primer objetivo planteado se encuentra muy ligado a la motivación de este trabajo, donde se quiere estudiar el impacto en la estimación de parámetros al utilizar conjuntos de consideración distintos del verdadero. Hacer esto muestra la importancia de conocer el proceso de generación del conjunto verdadero y a la vez se da luces sobre qué métodos son más robustos al momento de ser utilizados de manera práctica para la estimación de modelos de elección.

El segundo objetivo, que nace de la necesidad de obtener información para la caracterización del conjunto de consideración, es analizar de forma crítica tres métodos para recolectar datos sobre este conjunto. Los métodos considerados pertenecen a las familias de las preferencias reveladas y declaradas. De parte de las preferencias reveladas, se recopilan datos pasivos mediante Google Maps y se analiza la calidad de estos datos, junto con el potencial que tengan para brindar información sobre el conjunto de consideración. Por otro lado, las preferencias reveladas se estudian mediante dos encuestas. La primera es una encuesta online, que recopila datos socio económicos y realiza un ejercicio de simulación del proceso de generación del conjunto de consideración, junto con un ejercicio de elección y la recopilación de otro datos de interés. La segunda es una encuesta presencial sobre un mapa, donde se preguntan los mismo datos socio-económicos de la encuesta online, pero se les pide a las personas que dibujen las rutas que consideran en un mapa en un contexto en específico.

Finalmente, con el tercer y cuarto objetivo se pretende caracterizar el conjunto de consideración desde distintos puntos, utilizando los datos de preferencias declaradas recopilados. Como primera meta, se quiere identificar el tamaño del conjunto de consideración, lo cual se logra utilizando los conjuntos declarados en ambas encuestas. Además, se quiere identificar las posibles heurísticas que las personas utilizan para generar el conjunto de consideración, mediante la declaración de las personas y de un modelo logit binario de consideración. Por último, se quiere comparar los parámetros entre un modelo logit multinomial de elección y el modelo logit binario de consideración, mediante la realización de un modelo conjunto.

1.3. Metodología

El primer objetivo se aborda mediante simulaciones de Monte Carlo utilizando el software R, donde se generan datos sobre un escenario de elección basado en atributos de tiempo y costo que son usados para construir un conjunto de consideración específico a cada individuo modelado. Luego se estudia la capacidad para recuperar los parámetros de un modelo logit de elección mediante la construcción de conjuntos de consideración supuestos utilizando distintos métodos.

Posteriormente, para dar cuenta del segundo objetivo, se recolectaron datos que potencialmente podrían servir para la caracterización del conjunto de consideración en una aplicación práctica. El primer tipo de datos corresponde al itinerario georeferenciado de viajes recopilados a partir de las cuentas de usuarios de Google Maps y teléfonos inteligentes Android, para luego ser manipulados utilizando el software QGIS de forma de poder analizarlos. El segundo tipo de datos corresponde a una encuesta online desarrollada mediante la plataforma web Qualtrics, donde se recopilan datos socio-económicos y se realiza un experimento de consideración y elección de alternativas. Finalmente se recopilan datos mediante una encuesta presencial, donde se obtienen datos socio-económicos y se realiza un ejercicio de consideración de alternativas sobre un mapa impreso. Todos estos datos son evaluados, mostrando sus ventajas y desventajas al momento de ser utilizados para los objetivos planteados en esta memoria.

Finalmente, se utilizan datos de preferencias declaradas para realizar diversos aportes enfocados a la caracterización del conjunto. En primer lugar, para estudiar el tamaño del conjunto se obtienen los datos directamente de las encuestas online y sobre mapa, y se realizan test formales y análisis de varianza para detectar diferencias en el tamaño para distintos grupos socio-económicos. En segundo lugar se procesan las respuestas sobre las heurísticas que la gente declara que utiliza para construir el conjunto de consideración y, mediante un análisis de varianza, se identifica si existe alguna diferencia entre grupos socio-económicos. En tercer lugar se realiza un modelo logit binario para modelar el proceso de consideración. Además, se modelan las elecciones de las personas para comparar los parámetros de consideración y de elección. Finalmente, se realizan estimaciones de distintos modelos de elección que buscan recuperar los parámetros del modelo estimado con el conjunto de consideración declarado por las personas. Estos modelos, al igual que en la simulación de Monte Carlo, son estimados utilizando conjuntos de consideración generados por distintos métodos. Todos los análisis son desarrollados utilizando el software R.

1.4. Estructura de la memoria

En esta memoria, se tiene una estructura de 6 capítulos comenzando por el presente que contiene la introducción del trabajo de título presentado. El segundo capítulo contiene la revisión bibliográfica asociada al tema de interés, integrada por diversas áreas que utilizan modelos econométricos. En el tercer capítulo se presenta el estudio del impacto de utilizar distintos conjuntos de consideración supuestos por el modelador, para la estimación de pa-

rámetros de un modelo logit multinomial, mediante el método de Monte Carlo. En el cuarto capítulo se realiza un análisis crítico de tres métodos para recolectar datos sobre el conjunto de consideración, de donde se obtendrá la información que será utilizada para obtener resultados en el capítulo siguiente. De esta manera, en el capítulo cinco se realiza la caracterización del proceso de consideración utilizando datos de preferencias declaradas. Finalmente, en el capítulo seis se presentan las conclusiones del trabajo, entregando conclusiones generales, algunas recomendaciones metodológicas y extensiones que se pueden realizar a partir de este trabajo.

Capítulo 2

Revisión bibliográfica

2.1. Introducción

Existen tres áreas de la literatura que resultan de utilidad para lograr los objetivos propuestos. En primer lugar, se tienen los documentos asociados a elecciones discretas, en particular, destacan aquellos que incorporan el conjunto de consideración en sus análisis. En segundo lugar, están los documentos sobre el conjunto de consideración en el área de marketing, donde se exploran distintos aspectos sobre este conjunto, enfocados en las características que son importantes para que un producto se pueda comercializar. Por último, se encuentra la literatura del área de transporte, donde se tienen artículos sobre la modelación de elección de rutas, desde métodos que son utilizados para generar el conjunto de consideración, hasta las diversas maneras que se utilizan para realizar modelos de elección de rutas.

2.2. Modelación de elecciones discretas incorporando el conjunto de consideración

El marco teórico en el que se sustenta este trabajo es el de elecciones discretas, donde lo que se hace es modelar el proceso de elección de alternativas para entender y predecir el comportamiento de las personas. El primer supuesto que se debe realizar en este contexto es sobre el protocolo de decisión de las personas. El más utilizado es el de modelo de utilidad aleatoria, donde se asume que las personas realizan una elección cuando la utilidad que les reporta la alternativa seleccionada es la máxima de todas las existentes en el conjunto de alternativas consideradas. La utilidad desde el punto de vista del modelador, se divide en dos partes, una componente sistemática, que se puede calcular y una componente aleatoria, tal como se aprecia en la ecuación 2.1. Es en esta última, donde dependiendo del supuesto que se asuma, se generan distintos modelos de elección. El supuesto más utilizado, es el de asumir que la componente aleatoria, distribuye Gumbel o Valor Extremo Tipo 1, lo que da vida al conocido modelo logit, uno de los más utilizados gracias a su expresión cerrada para la probabilidad de elección como se muestra en la ecuación 2.2.

$$U_{in} = V_{in} + \varepsilon_{in} \quad (2.1)$$

$$P_n(i) = \frac{e^{\mu \cdot V_{in}}}{e^{\mu \cdot V_{in}} + \sum_{j \in C_n} e^{\mu \cdot V_{jn}}} \quad (2.2)$$

Donde los subíndices i,j representa a las alternativas i,j. El subíndice n representa al individuo n. Además, V_{in} es la utilidad sistemática de la alternativa i para el individuo n. μ es el parámetro de escala, que generalmente se fija en 1. Y finalmente, C_n corresponde al conjunto de alternativas consideradas por el individuo n.

Otro supuesto importante que se hace en elecciones discretas es que el modelador conoce el conjunto de consideración de las personas. Lo cual es sumamente cuestionable en un contexto donde las alternativas son muchas y no son claras de describir como es en elección de rutas. Para saber más sobre la base de elecciones discretas se recomienda ver el artículo de McFadden (1984) y los libros de Ben-Akiva y Lerman (1985) y Train (2009).

Por otro lado, existe una línea de trabajo originalmente desarrollada por Manski (1977), donde se utiliza un enfoque en que el conjunto de consideración es concebido como latente en la modelación, para poder hacerse cargo de los posibles errores que causa no conocer el verdadero conjunto de consideración de cada tomador de decisión. En la ecuación 2.3 se muestra la expresión desarrollada, que a simple vista puede parecer simple, pero poder conocer la probabilidad de cada posible conjunto de consideración para evaluar dicha expresión es impracticable.

$$P_n(i) = \sum_{C \in G_n} P_n(i | C) \cdot P_n(C) \quad (2.3)$$

Donde $P_n(i)$ es la probabilidad de que el individuo n elija la alternativa i. C es un conjunto de consideración y G_n son todos los conjuntos de posibles. Además, $P_n(i | C)$ es la probabilidad de que el individuo n elija la alternativa i dado que el conjunto de consideración es igual a C y $P_n(C)$ es la probabilidad de que el conjunto de consideración C sea el conjunto del individuo n.

En el trabajo de Swait y Ben-Akiva (1987), se estudian distintas restricciones, tanto deterministas como estocásticas, que sirven para formar expresiones para la probabilidad del conjunto de consideración. Un ejemplo de una restricción determinista en el contexto de elección de modo es la de posesión de licencia de conducir, de manera que si una persona no posee licencia entonces se descarta el modo de automóvil. En cambio, una restricción estocástica, puede ser el ingreso, donde una persona de ingreso alto tiene una probabilidad menor de incluir en su conjunto de consideración a la alternativa de transporte público. Con esto se genera una interpretación del proceso de generación del conjunto de consideración y se utiliza para entender distintos modelos de elección que incluyen a este proceso.

En la misma línea, Ben-Akiva y Boccara (1995) aplicaron una versión simplificada del método desarrollado por Manski, con la extensión realizada por Swait y Ben-Akiva (1987)

utilizando datos de preferencias declaradas. De manera que generan una expresión para la función de probabilidad de los posibles conjuntos de consideración utilizando información de encuestas sobre la disponibilidad de alternativas y la información de las elecciones observadas. En este artículo se muestra que esta metodología presenta una mejora muy pequeña con respecto a un modelo logit que considera todas las alternativas, lo cual no vale la pena para la capacidad computacional que se necesita.

Los trabajos aquí presentados, estudian maneras de mejorar la estimación de modelos de elección sin conocer en profundidad la verdadera forma del conjunto de consideración. En general estos métodos presentan un aumento en la complejidad computacional y es difícil encontrar una buena especificación para los modelos. Lo cual, sumado a la poca mejoría con respecto al enfoque clásico, hacen que se requiera más estudio en esta línea de trabajo.

2.3. Estudios sobre el conjunto de consideración en la literatura de marketing

En el área de marketing se estudia el conjunto de consideración en el contexto de la comercialización de productos de distinto tipo, debido a que para una marca es fundamental pertenecer al conjunto de consideración de las personas, o no tendrán ninguna posibilidad de ser escogidas.

En este contexto, Brown y Wildt (1992) realizaron un trabajo para determinar el tamaño del conjunto de consideración. En dicho trabajo se explican los distintos factores a tener en cuenta para poder preguntar sobre las alternativas consideradas, donde se presentan dos maneras de enunciar la pregunta que se debe realizar, así como también los diferentes métodos para hacer las encuestas necesarias. El primer método mencionado es el no asistido, que consiste en dejar que la gente piense en las alternativas existentes y diga las alternativas que considera. El segundo método es el asistido, donde se les muestra a las personas las alternativas existentes, y a partir de ellas se genera el conjunto de consideración. De esta manera logran obtener un tamaño de 4.54 en promedio de todas las definiciones utilizadas para conjuntos de consideración para comida rápida, refrescos y gasolina.

En trabajos más recientes como Hauser (2014), se menciona que para las compañías es fundamental que sus productos se encuentren en el conjunto de consideración de las personas, porque de no ser así, no existe la posibilidad de que sus productos sean elegidos en algún momento. En este mismo documento, se mencionan distintos tipos de heurísticas que los consumidores podrían utilizar para generar su conjunto de consideración. Además, se muestran algunas recomendaciones sobre cómo identificar las heurísticas utilizadas. Uno de los métodos mencionados, que inspira uno de los análisis realizados en esta memoria, es el de preguntar directamente a las personas sobre sus reglas de decisión. Así como también recolectar datos de preferencias declaradas, mostrando a las personas los productos existentes de algún rubro en específico y pedir que se declare si cada producto se considera o no, para posteriormente mediante algún método inferir la heurística utilizada.

Esta rama de la literatura presenta buenas ideas que no han sido aplicadas en elección

de ruta, y son parte importante de la motivación de esta memoria, brindando un enfoque de comprensión del conjunto de consideración en lugar de soluciones prácticas.

2.4. Modelación de elección de ruta

En elección de ruta, el número de alternativas disponibles es inconmensurable. Además, las opciones existentes suelen violar un supuesto fundamental del modelo logit multinomial que dice que las alternativas no deben estar correlacionadas, pero esto suele pasar cuando las rutas comparten alguna sección o se sobreponen.

Es común en este tópico, utilizar una estructura de red sobre las calles para obtener las distintas alternativas existentes. En esta sección se muestran los distintos trabajos que resumen los métodos existentes para generar el conjunto de consideración en elección de ruta, además de las distintas maneras de modelar la elección.

En el artículo de Bekhor et al. (2006) se estudian algunos algoritmos para generar el conjunto de consideración, basándose en el método de “Labelling Approach” propuesto por Ben-Akiva et al. (1984), el cual requieren funciones de costo generalizado para calcular rutas mínimas y así obtener el conjunto de consideración. Así mismo, Ramming (2001) y Prato (2009), muestran una revisión de diferentes maneras prácticas para generar el conjunto de consideración, donde la mayor parte de ellos se puede resumir en la realización de rutas mínimas, mediante algún algoritmo que permita rescatar un cierto número de rutas que cubra de buena manera el espacio. Además, se muestran las diferentes correcciones que se suelen utilizar al estimar los modelos logit con las alternativas en formato de red. Entre dichos métodos se encuentran, “Path Size Correction Logit”, distintas estructuras de la familia de modelos de valor extremo, estructuras no valor extremo como el probit, entre otros. En el mismo camino, Prato y Bekhor (2007) realizaron un análisis numérico de algunos de estos métodos, con la intención de probar un nuevo método basado en el algoritmo de “branch and bound”, sin embargo no se llegan a conclusiones contundentes.

Con el fin de mostrar la popularidad de los métodos deterministas, se presentan algunos artículos en que se han utilizado alguna de las variantes de estos métodos. Un primer ejemplo es el artículo de Raveau et al. (2014) donde se utiliza “Labelling Approach” para obtener el conjunto de alternativas disponibles en una red de Metro. Otro ejemplo encontrado, es la utilización de las k-ésimas rutas mínimas en el artículo de Papinski y Scott (2011). Además, se pueden encontrar aplicaciones de rutas mínimas con eliminación de arcos, como se presenta en Rieser-Schüssler et al. (2013).

En resumen, en esta rama de la literatura se presentan métodos prácticos para la generación del conjunto de consideración, que principalmente se hacen cargo del problema que causa el gran número de alternativas que existen en el contexto de elección de rutas.

Capítulo 3

Estudio del impacto de utilizar distintos conjuntos de consideración mediante método de Monte Carlo

3.1. Introducción

En este capítulo se analiza el impacto que tiene en la estimación de parámetros de un modelo de elección, que el modelador asuma diferentes heurísticas para la construcción del conjunto de consideración. El análisis se realiza mediante simulaciones de Monte Carlo implementadas en el software R (R Development Core Team, 2008). La metodología a seguir para llevar a cabo este experimento consiste en primer lugar de un proceso de generación de datos, con estos datos se generan cuatro escenarios de elección, donde en cada uno se genera el conjunto de consideración verdadero mediante distintos métodos o heurísticas. Luego, para cada uno de los escenarios se estiman modelos de elección utilizando distintos conjuntos de consideración supuestos por el modelador. Finalmente, se realiza un análisis de los sesgos causados por utilizar los distintos métodos de generación del conjunto de consideración supuesto para cada escenario presentado.

En la sección 3.2 se muestra como se generan los datos, tanto los atributos de las alternativas como las elecciones que dan vida a la simulación. Luego, en la sección 3.3 se explican las cuatro heurísticas que se utilizan para generar el conjunto de consideración verdadero, es decir, aquel que es utilizado para generar las elecciones. Además, en la sección 3.4 se muestran los seis métodos que se utilizan para generar el conjunto de consideración supuesto, es decir, aquel que será utilizado para estimar los parámetros del modelo de elección. Con esto, en la sección 3.5, se estudia el impacto que los métodos generadores del conjunto de consideración supuesto puede tener en la estimación de parámetros de un modelo logit multinomial, para poder identificar cuales son los métodos más robustos.

3.2. Generación de datos

En cada iteración de la simulación de Monte Carlo, se generan 2000 observaciones, donde mediante el pseudo-código presentado a continuación se crean los datos de tiempo y costo de 10 alternativas de manera que dentro de cierto rango aleatorio, cuando el tiempo aumenta el costo disminuye y viceversa.

1. Se establecen parámetros iniciales, inferiores y superiores para tiempo y costo como se indica a continuación:

$$t_{inferior} = c_{inferior} = 10$$

$$t_{superior} = c_{superior} = 30$$

2. Para cada alternativa se calcula la media que se utiliza para generar los datos de tiempo y costo según la siguientes formulas:

$$t_{prom} = t_{inferior} + (1 - id_{alternativa}) \cdot ((t_{inferior} - t_{superior})/10)$$

$$c_{prom} = c_{superior} + (1 - id_{alternativa}) \cdot ((c_{superior} - c_{inferior})/10)$$

3. Para cada alternativa se genera una desviación estándar de tiempo y costo según las siguientes fórmulas:

$$t_{desv} = 0,3 \cdot t_{inferior}$$

$$c_{desv} = 0,3 \cdot c_{inferior}$$

4. Utilizando las medias y desviaciones estándar para cada alternativa se generan 2000 observaciones mediante el valor absoluto de un número aleatorio generado a partir de una distribución normal.

En la Figuras 3.1 y Figura 3.2 se muestra la distribución de los datos de tiempo y costo respectivamente, generados para cada observación y alternativa en forma de gráfico de cajas y bigotes. Estos gráficos se presentan de manera de ilustrativa ya que son rescatados de una iteración arbitraria durante la simulación de Monte Carlo. La parte de abajo y arriba de las cajas corresponden al primer y tercer cuartil (25 % y 75 %). La barra horizontal en cada caja corresponde a la mediana y el diamante rojo corresponde a la media. Para más detalles se recomienda revisar la librería ggplot2 de Wickham (2009).

Una vez generados los datos de las alternativas para todas las observaciones, para cada caso presentado en la sección 3.3 se generan las elecciones mediante un modelo logit multinomial (ver sección 2.2). La especificación de la componente sistemática de la utilidad es la que se presenta en la ecuación 3.1 y los parámetros utilizados se presentan en la Tabla 3.1.

$$V_{in} = th_tiempo \cdot tiempo_{in} + th_costo \cdot costo_{in} \quad (3.1)$$

Tabla 3.1: Valores de los parámetros utilizados para generar las elecciones en la simulación de Monte Carlo.

Parámetro	Valor
th_tiempo	-0.25
th_costo	-0.20

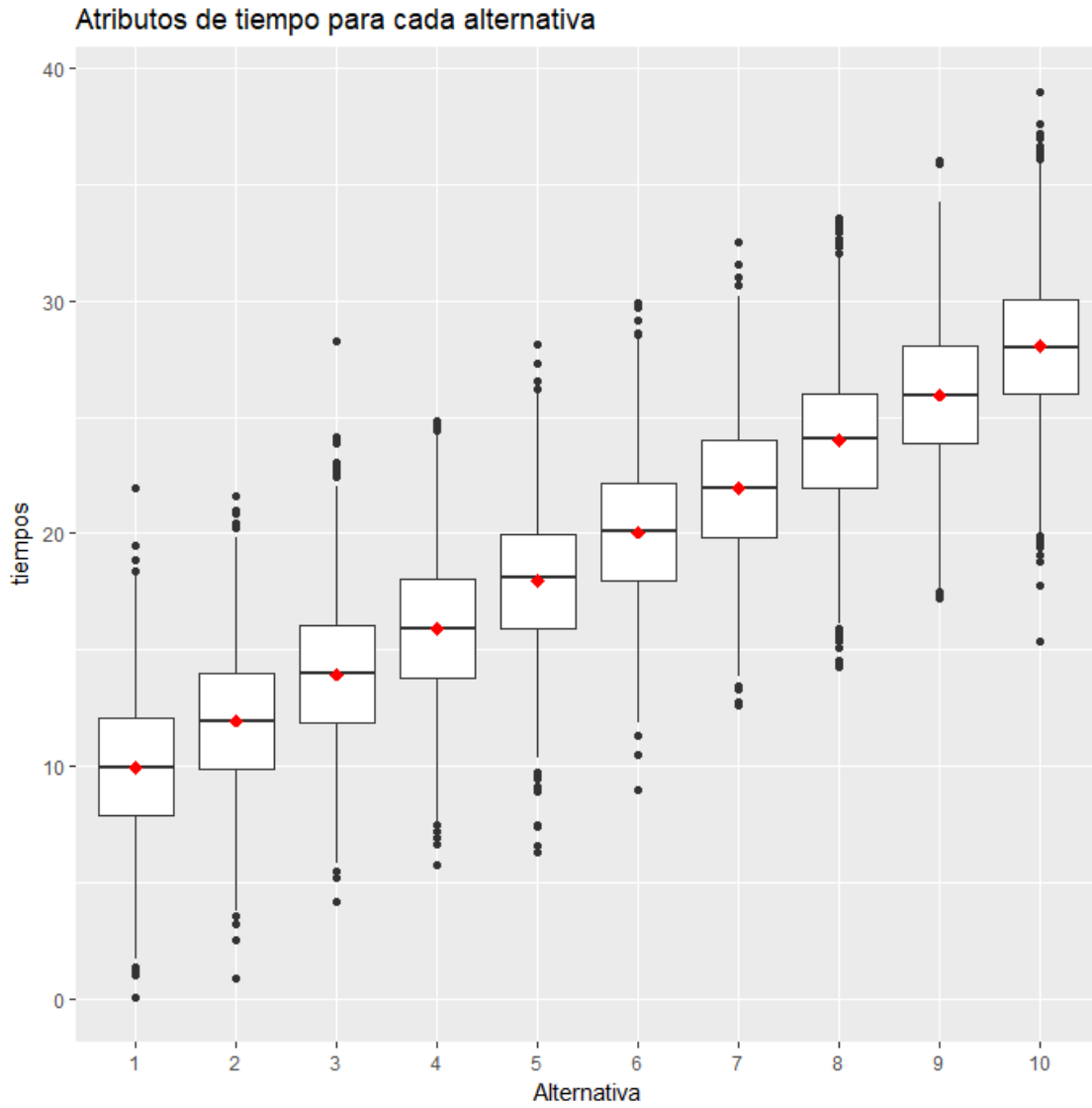


Figura 3.1: Datos de tiempo generados para cada alternativa y observación en una iteración arbitraria durante la simulación de Monte Carlo.

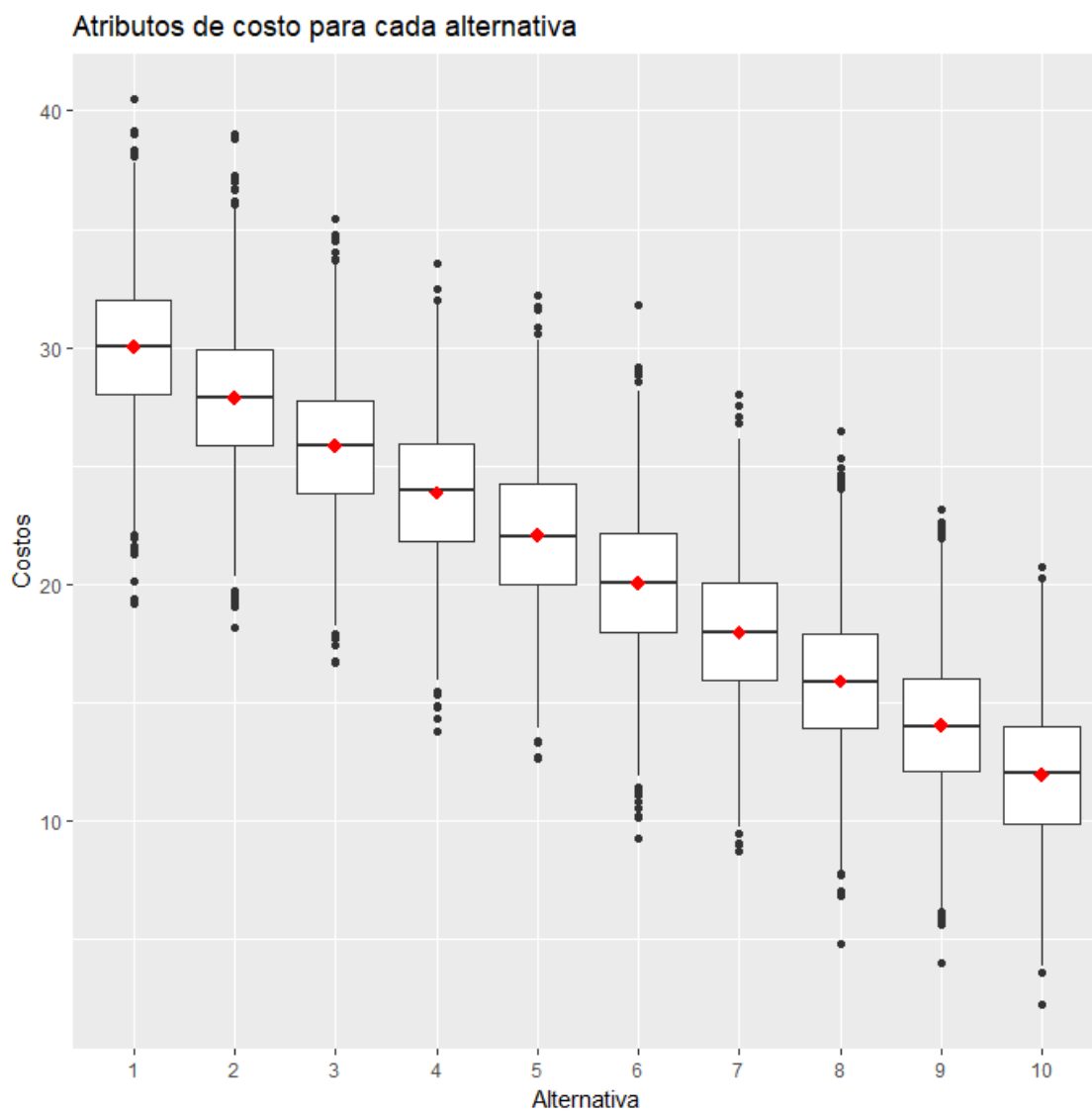


Figura 3.2: Datos de costo generados para cada alternativa y observación en una iteración arbitraria durante la simulación de Monte Carlo.

3.3. Métodos generadores del conjunto de consideración verdadero

Las cuatro heurísticas utilizadas para generar el conjunto de consideración verdadero C_n buscan cubrir distintas posibilidades de lo que realmente sucede en el proceso de generación del conjunto de consideración utilizado por las personas. Este conjunto es construido a partir del universo de 10 alternativas U_n . Estas reglas de decisión se presentan a continuación:

1. Aleatorio

Esta heurística genera el conjunto de consideración de manera aleatoria. Para ello, se crea una lista de todos los posibles conjuntos de consideración con al menos 3 alternativas. Luego, para cada una de las 2000 observaciones se elige de manera aleatoria uno de los conjuntos de dicha lista obteniendo así C_n . Este escenario es generado con el fin de cubrir todas las posibilidades, sin embargo, está probablemente muy alejado de la heurística que los individuos usan en la realidad.

2. Todas las alternativas

Esta regla de decisión simplemente hace que todas las alternativas existentes, en este caso 10, sean parte del conjunto de consideración verdadero, es decir, $C_n = U_n$. Al igual que el escenario anterior, este es poco realista, pero es utilizado para cubrir todas las posibilidades incluso en los extremos.

3. Eliminación por aspecto

En esta heurística se crea el conjunto de consideración verdadero mediante un algoritmo que se asemeja a eliminación por aspecto, eliminando las alternativas que no cumplan las siguientes restricciones:

$$\begin{aligned} \text{costo}_{in} &< \text{costo_promedio}_n + 0,8 \cdot \text{desv_costo}_n \\ \text{tiempo}_{in} &< \text{tiempo_promedio}_n + 0,8 \cdot \text{desv_tiempo}_n \end{aligned}$$

Este escenario es más realista que los dos anteriores, ya que es una de las posibles heurísticas que se encuentran en la literatura al referirse a la construcción del conjunto de consideración por parte de las personas.

4. Modelo logit binario de consideración

La última regla consiste en utilizar un modelo logit binario, la especificación de la utilidad para este modelo se presenta en la ecuación 3.2 y la probabilidad de considerar una alternativa se calcula siguiendo la expresión de la ecuación 3.3. Además, los parámetros para calcular la utilidad son dados y se presentan en la Tabla 3.2.

$$V_{in} = Cte + th_tiempo \cdot tiempo_{in} + th_costo \cdot costo_{in} \quad (3.2)$$

Donde V_{in} es la utilidad sistemática de la alternativa i para el individuo n . Cte es la constante que da cuenta de la utilidad de base que se utiliza en este tipo de modelos (véase literatura de la sección 2.2). Además, th_tiempo y th_costo son los parámetros de tiempo y costo respectivamente. Finalmente, $tiempo_{in}$ y $costo_{in}$ son los atributos de tiempo y costo de la alternativa i para la persona n .

$$P_n(i) = \frac{e^{\mu \cdot V_{in}}}{e^{\mu \cdot V_{in}} + 1} \quad (3.3)$$

Donde $P_n(i)$ es la probabilidad de que el individuo n considere la alternativa i . Además, μ es el parámetro de escala que se fija en 1. Finalmente, V_{in} es la utilidad sistemática de la alternativa i para el individuo n .

Tabla 3.2: Parámetros de modelo de consideración.

Parámetro	Valor
Cte	7.50
th_tiempo	-0.25
th_costo	-0.20

Luego, para cada alternativa se calcula la probabilidad de ser considerada y mediante la simulación de números aleatorios uniformes entre 0 y 1, se realiza la decisión de considerar o no cada alternativa (si la probabilidad calculada es menor al número aleatorio simulado), generando de esta manera el conjunto de consideración verdadero C_n . Este escenario puede llegar a representar bien la realidad ya que un modelo de consideración puede actuar en representación de alguna heurística utilizada por las personas.

En todos los casos, por temas prácticos, se filtran todas las observaciones con menos de 3 alternativas consideradas, ya que algunos de los métodos utilizados para generar el conjunto supuesto (ver sección 3.4) requieren un mínimo de alternativas para funcionar. De esta manera, con los conjuntos de consideración generados en cada uno de los casos, se utiliza un modelo logit multinomial, como se explica en la sección 3.2, para generar las elecciones.

Por último, en la Figura 3.3, Figura 3.4, Figura 3.5 y Figura 3.6, se presenta el porcentaje de veces que cada alternativa es escogida en la generación de datos, para cada uno de los escenarios de generación del conjunto de consideración verdadero.

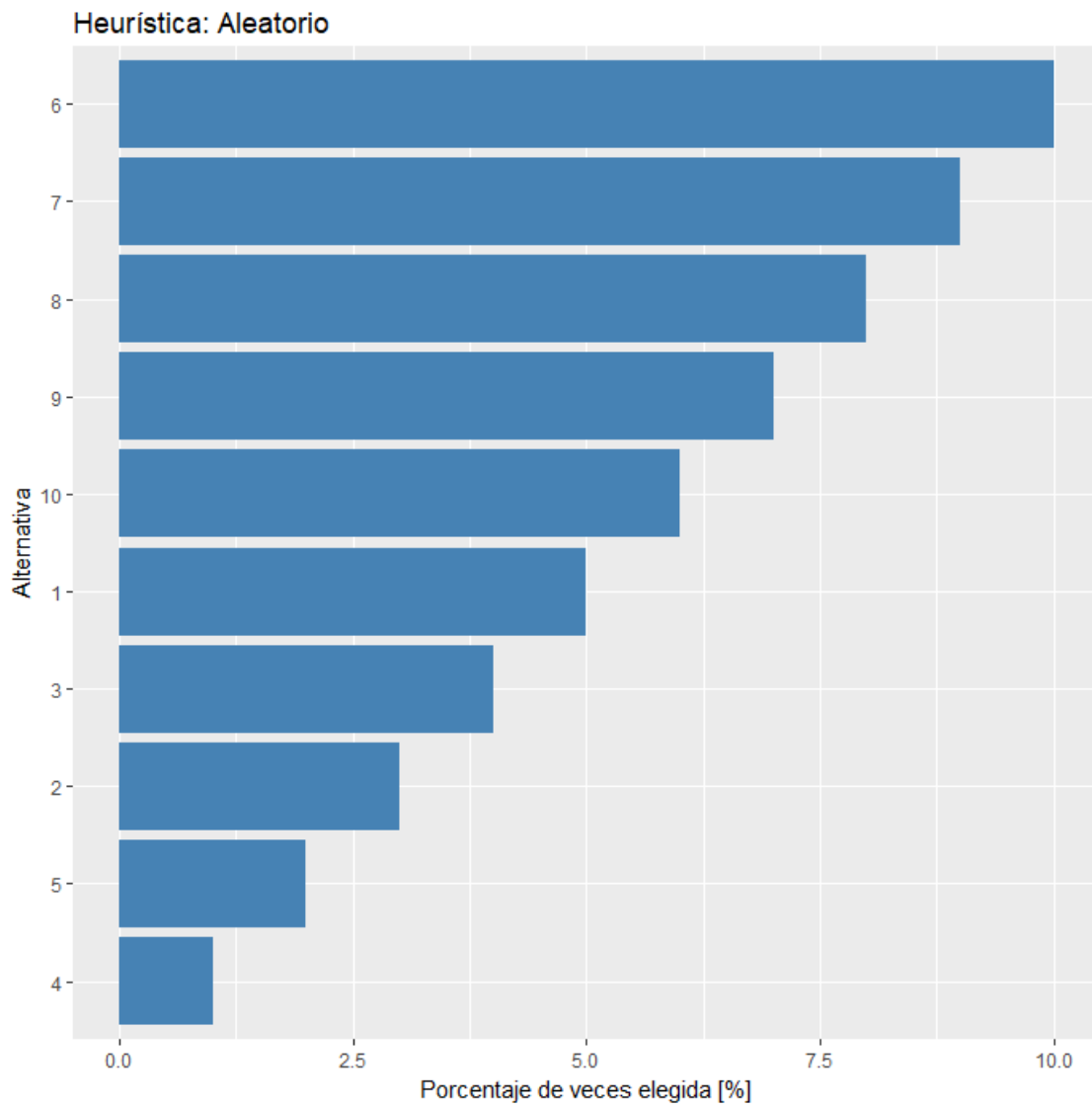


Figura 3.3: Porcentajes de elecciones de cada alternativa de la simulación, cuando la heurística original es “Aleatorio”.

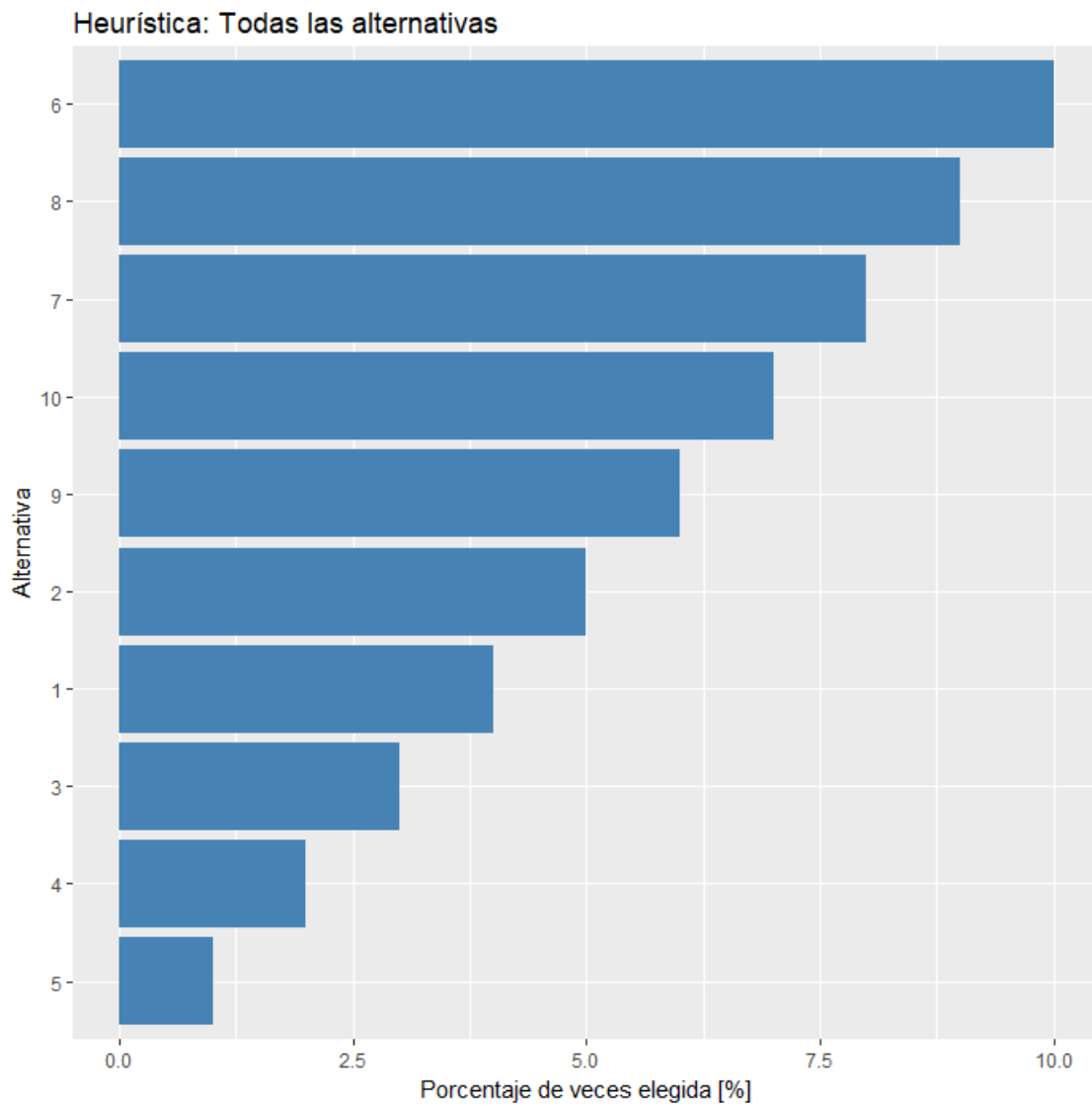


Figura 3.4: Porcentajes de elecciones de cada alternativa de la simulación, cuando la heurística original es “**Todas las alternativas**”.

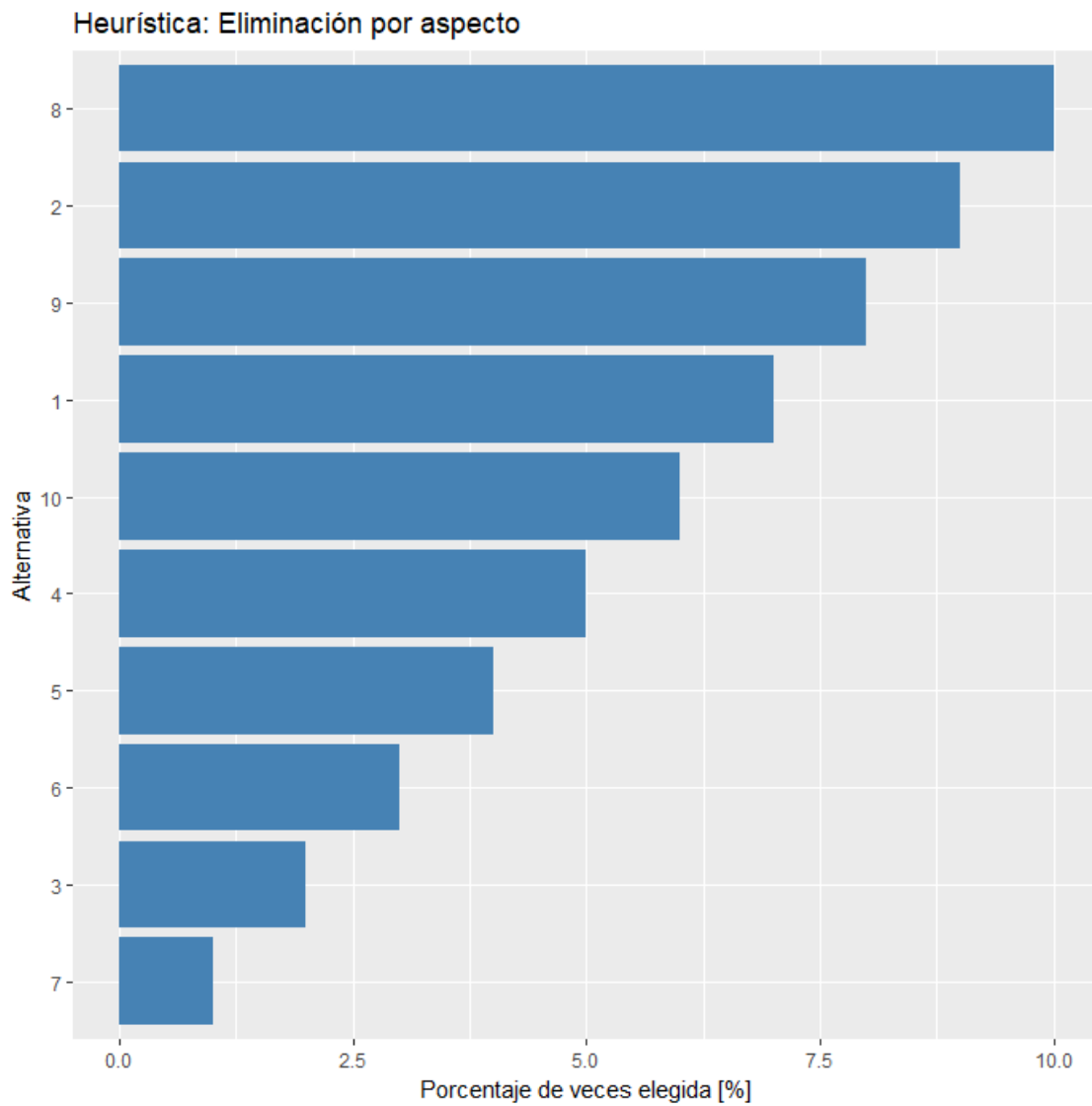


Figura 3.5: Porcentajes de elecciones de cada alternativa de la simulación, cuando la heurística original es “**Eliminación por aspecto**”.

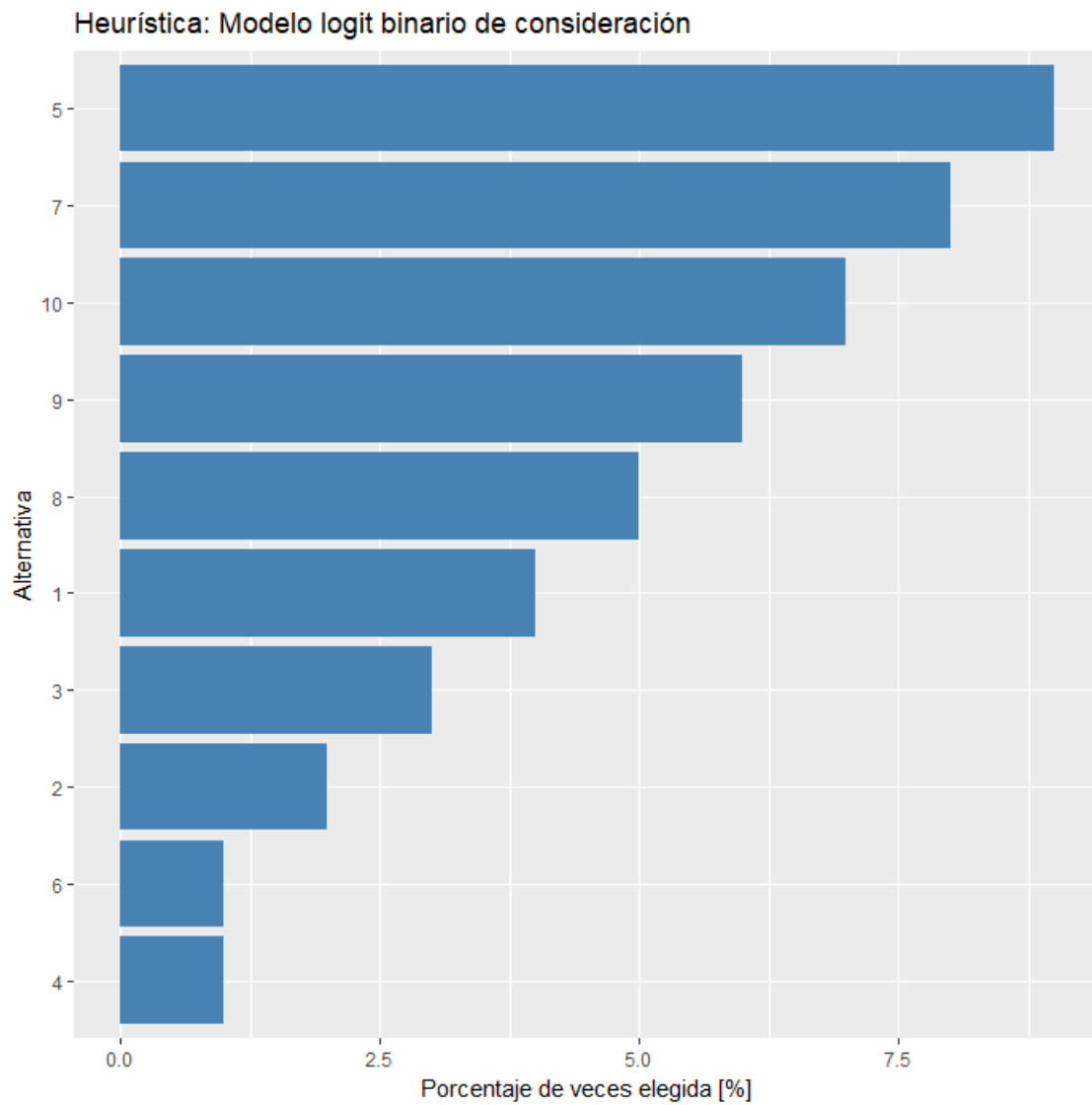


Figura 3.6: Porcentajes de elecciones de cada alternativa de la simulación, cuando la heurística original es “Modelo logit binario de consideración”.

3.4. Métodos generadores del conjunto de consideración supuesto

Al momento de modelar en elecciones discretas, el analista debe realizar ciertos supuestos sobre el proceso de generación del conjunto de consideración debido a su falta de conocimiento sobre este, lo cual de estar errado se traduce en un sesgo en la estimación de parámetros. En esta sección se presentan diversos métodos que pueden ser utilizados por los analistas para afrontar esta etapa del proceso.

A continuación se presentan seis métodos generadores del conjunto de consideración supuesto, los cuales son utilizados para evaluar el sesgo que causan en la estimación de parámetros de un modelo logit igual al utilizado para generar la elección (ver sección 3.2).

1. Conjunto de consideración verdadero (CV)

Este método consiste en utilizar en cada caso el conjunto de consideración verdadero que se usó para generar los datos, según se describe en la sección 3.3. Este ejercicio tiene como fin verificar que los parámetros se recuperan de manera correcta.

2. Conjunto de consideración experimentado (CE)

Este método consiste en simular suficientes elecciones de cada individuo hasta observar al menos 2 alternativas seleccionadas además de la original para cada uno. De esta manera el conjunto de consideración supuesto se compone por las alternativas que se observan como experimentadas, de manera similar a lo que sería una base de datos de serie de tiempo.

Este método muestra ventajas en cuanto a la factibilidad de su aplicación práctica. Las alternativas observadas pueden ser obtenidas mediante datos pasivos como GPS en el caso de elecciones de ruta (los cuales se estudian en la sección 4.2) o con el historial de compras de una tarjeta bancaria en el caso de cualquier producto.

3. Conjunto de consideración de las k-ésimas rutas mínimas, utilizando los parámetros verdaderos (KRM_V)

En este método se construye el conjunto de consideración utilizando la regla determinista de rutas mínimas, tratando de emular distintos métodos señalados en artículos como Prato (2009) y Ramming (2001). Este método consiste en calcular la componente sistemática de la utilidad, con la misma forma funcional mostrada en la ecuación 3.1 utilizada por el modelo logit multinomial con el que se generan las elecciones, utilizando parámetros a priori. En este caso particular, los parámetros a priori son los verdaderos utilizados en el modelo de elección original, mostrados en la Tabla 3.1. Luego, con la utilidad sistemática calculada se realiza un ranking y se eligen las 3 mayores para formar el conjunto de consideración supuesto, además si la elección original no se encuentra en este conjunto, entonces se agrega.

4. Conjunto de consideración de las k-ésimas rutas mínimas, utilizando parámetros erróneos (KRM_E)

Este método es equivalente al anterior, pero se utilizan parámetros a priori arbitrariamente distintos a los verdaderos del modelo de elección presentados en la Tabla 3.3.

Tabla 3.3: Valores de los parámetros a priori.

Parámetro	Valor
<i>th_tiempo</i>	-0.26
<i>th_costo</i>	-0.20

Con esto se calcula la utilidad sistemática utilizando la especificación del modelo de elección mostrada en la ecuación 3.1. Luego, se hace un ranking de las alternativas y se eligen las 3 mejores para formar el conjunto de consideración, además si la alternativa escogida no es una de las 3 mejores, se agrega al conjunto.

Este método es de los más prácticos por su simpleza y además presenta un caso realista ya que los parámetros a priori nunca son los reales, o el modelador no estaría tratando de estimarlos. Cabe destacar que los parámetros presentados en la Tabla 3.3 son elegidos basándose en los parámetros verdaderos con una ligera modificación, por lo que un análisis de sensibilidad al modificar estos parámetros podría ser realizado en análisis futuros.

5. Conjunto de consideración completo (CC)

Este método, es el más simple de modelar, ya que incluye a todas las alternativas en el conjunto de consideración supuesto. En el caso en que el conjunto de consideración verdadero también se genere con la misma regla, entonces este método será equivalente a aquel que recupera los parámetros del modelo original.

6. Conjunto de consideración con logit binario (CLB)

Por último, este método es el único que pertenece a la familia de los métodos estocásticos. Lo que se hace en este caso, es estimar un modelo logit binario, donde a partir del conjunto de consideración verdadero, se utiliza cada alternativa como una observación. Luego, cada elección consiste en si una alternativa es considerado o no. La especificación de este modelo es, al igual que en los otros casos, una combinación lineal de las variables de tiempo y costo, además de una constante específica que es necesaria en un logit binario, ya que la utilidad de una alternativa es comparada con una utilidad base que es desconocida y se fija en cero.

Una vez que se estima el modelo de consideración, se utiliza el mismo modelo para simular un conjunto de consideración, es decir, se calcula la probabilidad de cada alternativa de ser elegida y luego se simulan variables aleatorias uniformes entre 0 y 1 para

decidir si una alternativa es considerada o no. De esta manera, las alternativas cuya probabilidad es mayor que el número simulado son agregadas al conjunto de consideración supuesto.

3.5. Análisis y resultados de la simulación de Monte Carlo

A continuación se presentan resultados que muestran el sesgo que causa cada uno de los 6 métodos generadores del conjunto de consideración supuesto (ver Tabla 3.4) bajo los 4 escenarios de generación del conjunto de consideración verdadero (ver Tabla 3.5). En la Figura 3.7, Figura 3.8, Figura 3.9 y Figura 3.10, se presentan los estimadores de la razón entre el parámetro de costo con el parámetro de tiempo, para 100 instancias (o repeticiones) de estimación en formato de gráfico de cajas y bigotes. Estos valores se contrastan con el valor poblacional de 0.8, denotado con una línea segmentada horizontal. La parte de abajo y arriba de las cajas corresponden al primer y tercer cuartil (25 % y 75 %). La barra horizontal en cada caja corresponde a la mediana y el diamante rojo corresponde a la media. El límite de los "bigotes" de cada diagrama de cajas corresponde a la última observación disponible, siempre y cuando esta no se encuentre a más de 2 veces el largo de la caja de distancia de la mediana, los puntos fuera de este límite se consideran outliers. Todos los gráficos fueron desarrollados utilizando el software R, mediante la librería ggplot2 de Wickham (2009).

Por otro lado, en la Tabla 3.6, Tabla 3.7, Tabla 3.8 y Tabla 3.9, se presentan resultados numéricos, mostrando cuatro métricas para caracterizar el sesgo de cada método generador del conjunto de consideración supuesto. La primera métrica corresponde al sesgo promedio, el cual indica la diferencia entre el valor promedio de la razón de los parámetros estimados y el valor de la razón verdadera. La segunda corresponde al error de la raíz cuadrada de la media (RMSE), a modo de ejemplo, para un estimador insesgado, el RMSE es la raíz cuadrada de la varianza. La tercera corresponde a un test-t, donde se compara el valor de la razón de los parámetros estimados con el valor de la razón verdadera, el valor crítico para este test es 1.984 para 99 (iteraciones - 1) grados de libertad con 95 % de confianza, donde un estadístico menor a este valor, representa que se acepta la hipótesis nula de que el valor estimado es estadísticamente igual al valor real. Finalmente, se presenta la cobertura empírica con un 75 % de confianza, lo que muestra cuantas veces de las 100 iteraciones el valor de la razón de parámetros estimada es estadísticamente igual a la razón verdadera, a modo de ejemplo, para el método que recupera los parámetros utilizando el conjunto verdadero la cobertura empírica al 75 % debe ser de alrededor de 75.

A partir de los tablas y gráficos presentados, lo primero que destaca es que en los dos escenarios poco realistas, es decir, generando el conjunto de consideración verdadero mediante las heurísticas "Aleatorio" y "Todas las alternativas", todos los métodos para generar el conjunto supuesto se comportan bien excepto KRM_E. Esto causa preocupación debido a que la hipótesis que se sostiene en este análisis es que KRM_E es uno de los métodos más comúnmente utilizados. Luego, en los dos casos más realistas, es decir, cuando se genera el conjunto de consideración verdadero con las heurísticas de "Eliminación por aspecto" y "Modelo logit

binario de consideración”, el método KRM_V también falla, lo que entrega indicios sobre que este tipo de métodos no son apropiados para estimar los parámetros en este contexto. Se debe notar que incluso en el caso cuando se genera el conjunto de consideración verdadero con las heurísticas de “Eliminación por aspecto”, el método KRM_E presenta razones de los parámetros negativas, lo que indica que alguno de los signos de los parámetros se recuperan de forma incorrecta, lo que inmediatamente invalidaría la estimación. Además, en estos mismos escenarios realistas, el método de generación de conjunto supuesto CC también falla en la estimación de los parámetros. Adicionalmente, se observa que incluso el método CLB falla en el caso en que la heurística verdadera es “Eliminación por aspecto”. Por otro lado, el método generador del conjunto de consideración supuesto CE logra estimar los parámetros de forma correcta, con el único contra de tener más varianza que con el conjunto verdadero. Esto muestra que el método CE se presenta como un método robusto.

De lo anterior se desprende que los métodos KRM_V y KRM_E son los menos eficaces para enfrentar el problema planteado. Una de las posibles explicaciones para esto, es que estos métodos son los que simultáneamente dejan fuera alternativas consideradas e introducen al conjunto supuesto alternativas que no están en el conjunto verdadero. Por otro lado, el método CC falla a pesar de ser un escenario con un número relativamente bajo de alternativas en comparación a la realidad en el contexto de elección de rutas. También se puede observar que el método CLB falla en el escenario realista cuando la heurística es “Eliminación por aspecto”, esto a pesar de ser calibrado conociendo el conjunto de consideración verdadero de todas las observaciones. Se debe reconocer que en un escenario real se estimaría el modelo de consideración a partir de una muestra pequeña por lo que los resultados deberían empeorar. Finalmente, el método novedoso CE, muestra mucho potencial e incluso podría comportarse mejor si se pidieran más de 3 alternativas, pero esto requeriría mayor procesamiento de la simulación y sería menos práctico en la realidad. La principal intuición detrás del comportamiento robusto de este método es que todas las alternativas contenidas en el conjunto supuesto se encuentran en el conjunto verdadero, ya que son elegidas en algún momento por la persona. Además, este método tiene principal importancia en el contexto de elección de rutas, donde las elecciones pueden cambiar con facilidad.

Se debe tener en cuenta que hay que tener cuidado con generalizar los resultados en un experimento numérico. Sin embargo, estos hallazgos sirven de guía para un análisis más profundo sobre la modelación en elección de ruta y la generación del conjunto de consideración.

Finalmente, existen muchas extensiones posibles para realizar esta simulación. Por un lado, se pueden dejar de realizar supuestos sobre la independencia de las alternativas (asumido por el modelo logit multinomial), esto se debe realizar generando los datos con algún nivel de correlación, idealmente imitando lo que sería el escenario de elección de rutas y posteriormente modelando con métodos que se hagan cargo de dicha correlación, como el logit anidado, probit, entre otros. Por otro lado, los métodos generadores del conjunto supuesto se pueden mejorar, además de probar otros y realizar análisis de sensibilidad sobre los distintos parámetros utilizados, por ejemplo, el número de alternativas seleccionadas en el ranking de los métodos KRM_V y KRM_E o las alternativas observadas en el método CE.

Tabla 3.4: Listado de métodos generadores del conjunto de consideración supuesto.

Abreviación	Método
CV	Conjunto de consideración verdadero
CE	Conjunto de consideración experimentado
KRM_V	Conjunto de consideración de las k-ésimas rutas mínimas, utilizando los parámetros verdaderos
KRM_E	Conjunto de consideración de las k-ésimas rutas mínimas, utilizando parámetros erróneos
CC	Conjunto de consideración completo
CLB	Conjunto de consideración con logit binario

Tabla 3.5: Listado de métodos generadores del conjunto de consideración verdadero.

Heurística original (Escenario)
Aleatorio
Todas las alternativas
Eliminación por aspecto
Modelo logit binario de consideración

Tabla 3.6: Análisis de Monte Carlo de los métodos generadores del conjunto de consideración supuesto, cuando la heurística original es “Aleatorio”.

Método	Sesgo Promedio	RMSE	test-t ($\frac{th_costo}{th_tiempo} = 0,8$)	Cobertura Empírica 75 %
CV	0.00138	0.0174	0.0797	83
CE	0.00356	0.0273	0.132	58
KRM_V	-0.00339	0.0188	0.183	81
KRM_E	-0.0601	0.0630	3.16	5
CC	-0.000110	0.0186	0.00589	77
CLB	0.0000109	0.0200	0.000547	79

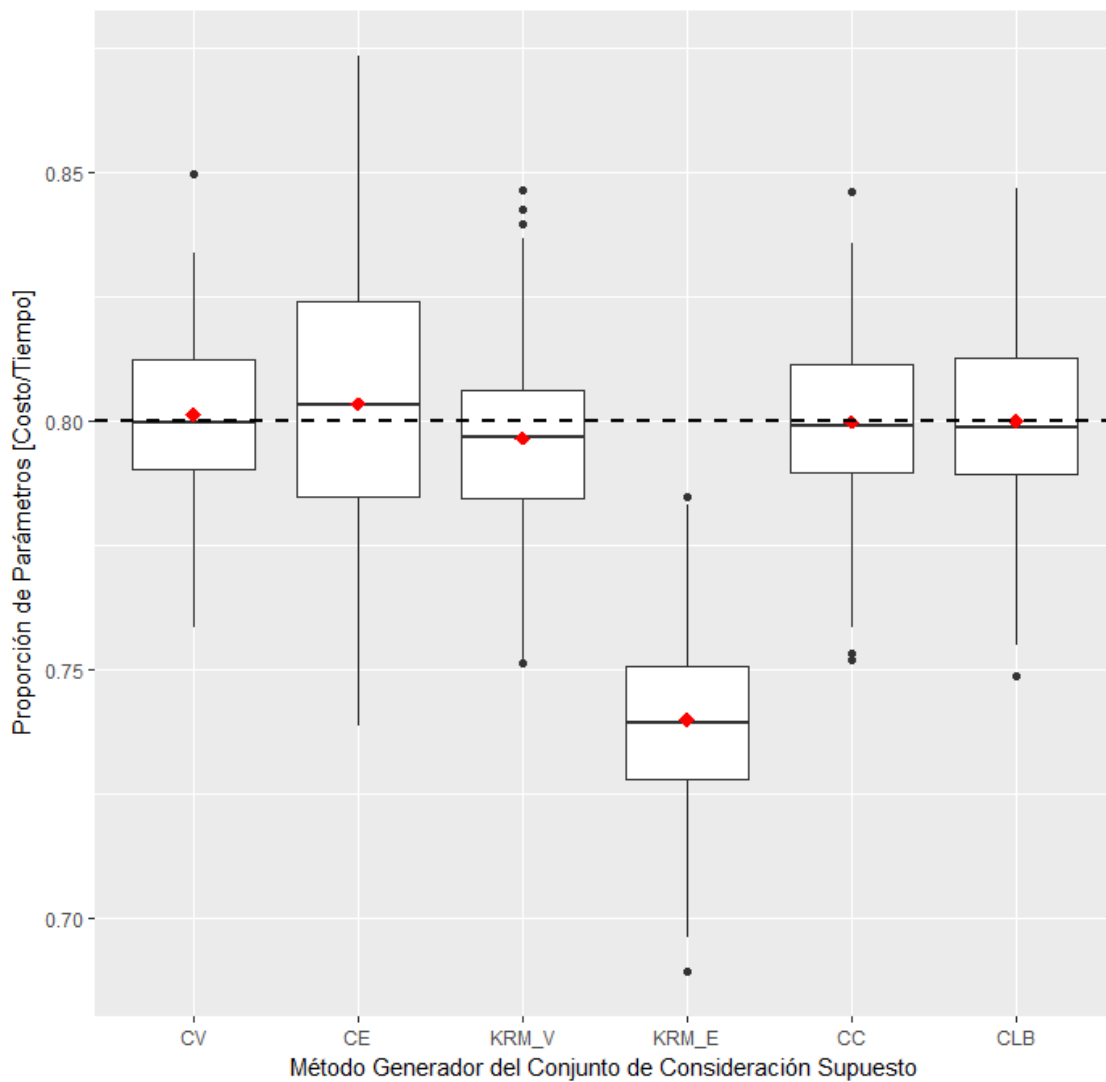


Figura 3.7: Comparación en la estimación de parámetros utilizando distintos métodos para generar el conjunto de consideración supuesto, cuando la heurística original es “Aleatorio”.

Tabla 3.7: Análisis de Monte Carlo de los métodos generadores del conjunto de consideración supuesto, cuando la heurística original es “**Todas las alternativas**”.

Método	Sesgo Promedio	RMSE	test-t ($\frac{th_costo}{th_tiempo} = 0,8$)	Cobertura Empírica 75 %
CV	0.00169	0.0145	0.117	87
CE	0.00755	0.0519	0.147	42
KRM_V	-0.0138	0.0332	0.456	57
KRM_E	-0.105	0.110	3.377	3
CC	0.00169	0.0145	0.117	87
CLB	0.00169	0.0145	0.117	89

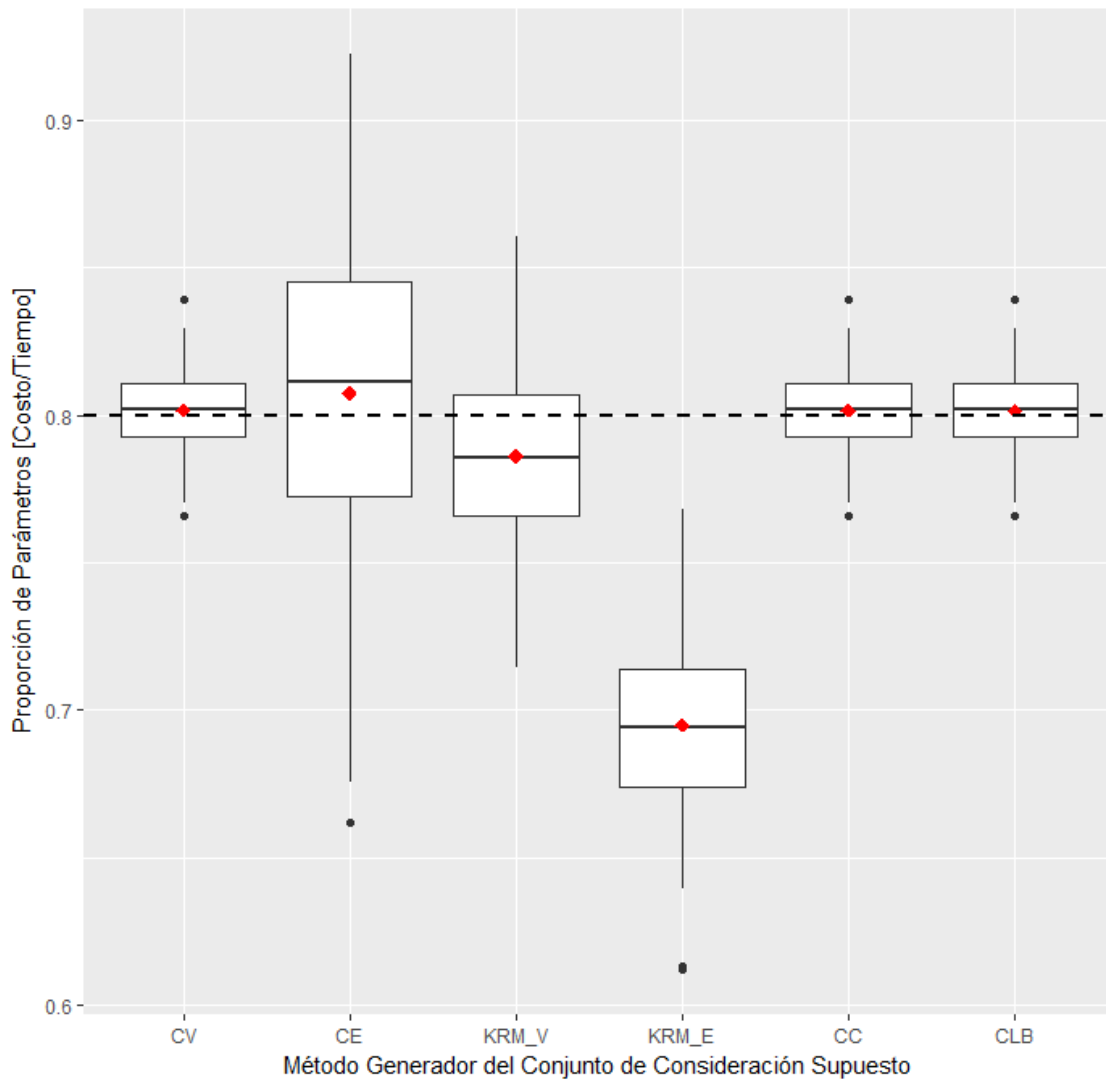


Figura 3.8: Comparación en la estimación de parámetros utilizando distintos métodos para generar el conjunto de consideración supuesto, cuando la heurística original es “**Todas las alternativas**”.

Tabla 3.8: Análisis de Monte Carlo de los métodos generadores del conjunto de consideración supuesto, cuando la heurística original es “**Eliminación por aspecto**”.

Método	Sesgo Promedio	RMSE	test-t ($\frac{th_costo}{th_tiempo} = 0,8$)	Cobertura Empírica 75 %
CV	0.00259	0.0218	0.119	78
CE	-0.00620	0.04191	0.150	58
KRM_V	-0.532	0.538	6.75	0
KRM_E	-0.654	0.660	7.65	0
CC	0.136	0.136	15.7	0
CLB	0.114	0.115	9.14	2

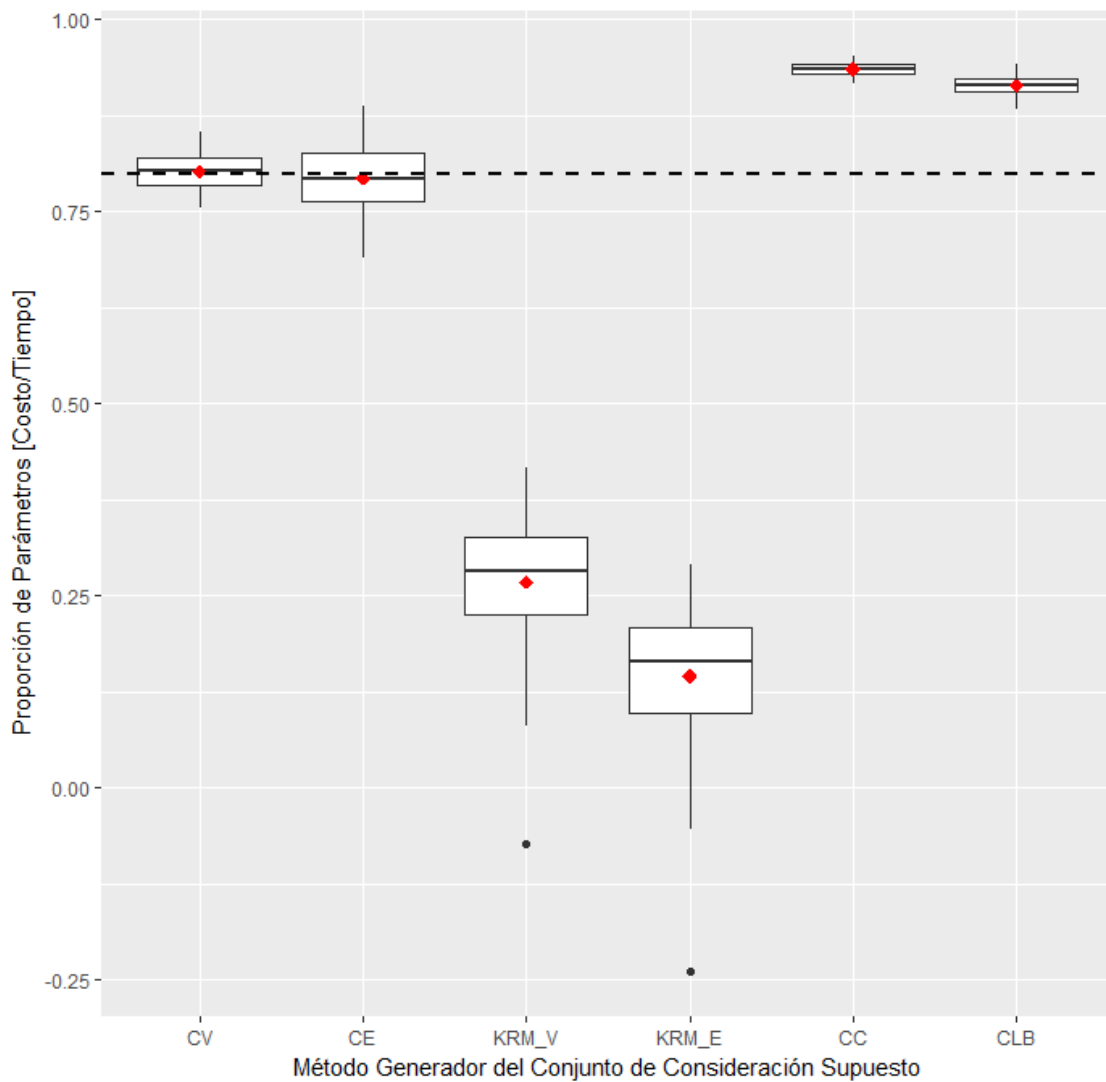


Figura 3.9: Comparación en la estimación de parámetros utilizando distintos métodos para generar el conjunto de consideración supuesto, cuando la heurística original es “**Eliminación por aspecto**”.

Tabla 3.9: Análisis de Monte Carlo de los métodos generadores del conjunto de consideración supuesto, cuando la heurística original es “**Modelo logit binario de consideración**”.

Método	Sesgo Promedio	RMSE	test-t ($\frac{th_costo}{th_tiempo} = 0,8$)	Cobertura Empírica 75 %
CV	0.00104	0.0153	0.0681	90
CE	0.00361	0.0284	0.128	68
KRM_V	-0.168	0.175	3.31	0
KRM_E	-0.313	0.319	5.28	0
CC	0.0450	0.0467	3.62	32
CLB	0.00963	0.0188	0.599	84

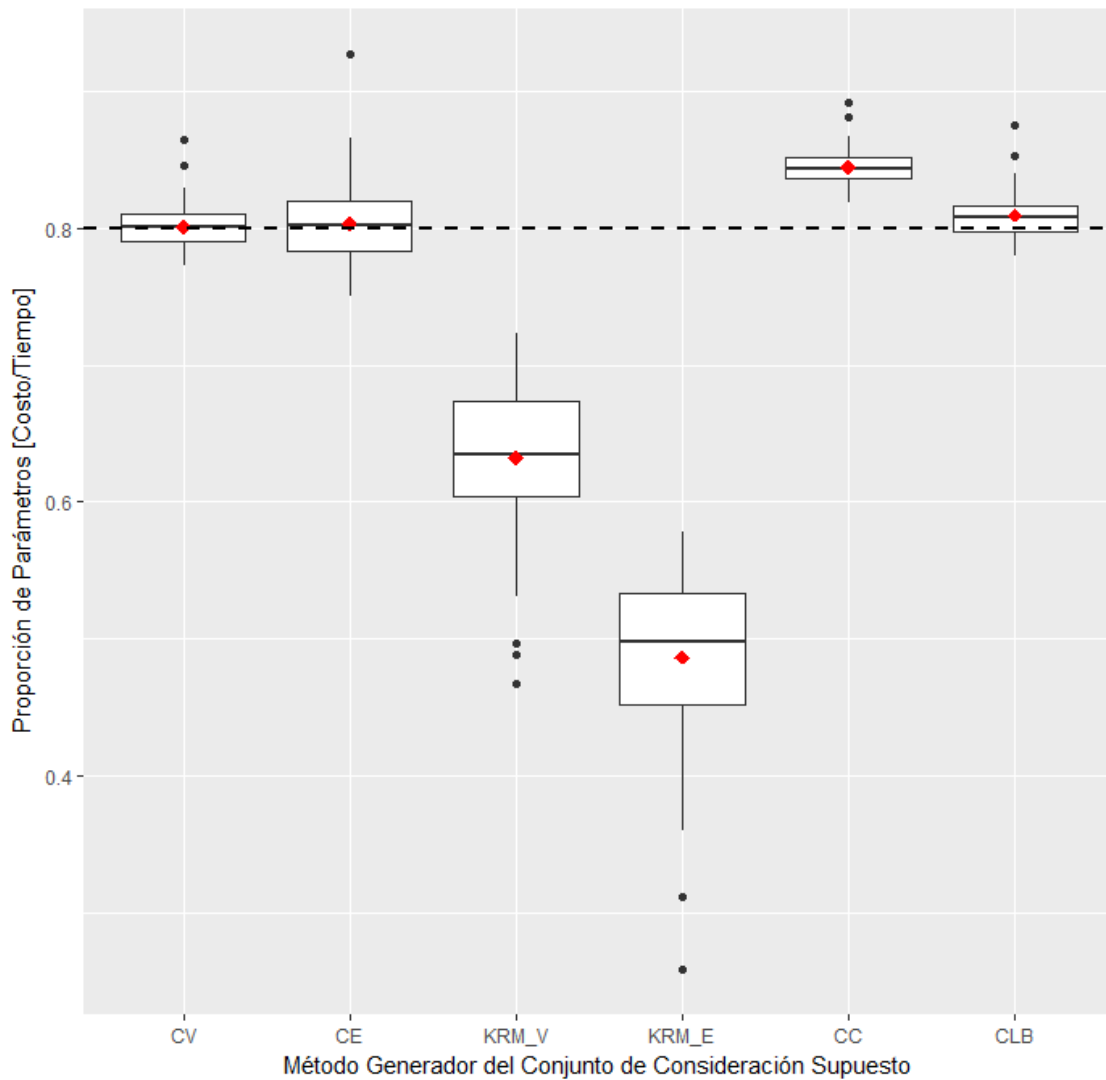


Figura 3.10: Comparación en la estimación de parámetros utilizando distintos métodos para generar el conjunto de consideración supuesto, cuando la heurística original es “**Modelo logit binario de consideración**”.

Capítulo 4

Análisis crítico de tres métodos para recolectar datos sobre el conjunto de consideración

4.1. Introducción

Con el fin de realizar un estudio con datos empíricos, en este capítulo se explora de manera preliminar la factibilidad de tres herramientas para la recolección de datos que potencialmente podrían brindar información sobre el conjunto de consideración. La primera herramienta corresponde a la recopilación de datos pasivos generados por Google Maps para usuarios de teléfonos inteligentes Android. La segunda herramienta es una encuesta de preferencias declaradas online. La última herramienta para la recolección de datos sobre el conjunto de consideración corresponde a una encuesta presencial utilizando un mapa para dibujar. En las siguientes secciones se presenta cómo funciona cada una de estas herramientas, además de mencionar las ventajas y desventajas de cada una, a la luz de su aplicación a nivel de prototipo a una muestra de tamaño reducido.

4.2. Datos pasivos

La herramienta basada en datos pasivos, se sustenta en la data recolectada por Google Maps de las personas que utilizan teléfonos celulares inteligentes con sistema operativo Android. Como se menciona en Cain (2016) lo que hace Google es tomar los datos de una variedad de fuentes como GPS, WiFi, torres de celular y sensores de dispositivos como los giroscopios y los acelerómetros, referenciándolos en espacio y tiempo, para luego mediante algún proceso interno unir los puntos que tengan sentido en forma de ruta, pudiendo incluso estimar el modo de transporte del viaje. Una primera limitación que se detectó en este tipo de datos fue que las rutas generadas eran solo líneas sobre un lienzo y no se encontraban sobre una estructura de red definida.

En este estudio se les solicitó a tres personas facilitar los datos de algunos de sus viajes a su lugar de estudio o trabajo. Los participantes son conocidos como “Usuario Caminata”, “Usuario Transporte Público” y “Usuario Auto” y se distinguen por los distintos modos de transporte que sus nombres representan. Como referencia, los tres participantes son estudiantes universitarios cuyas edades rondan los 25 años. Este proceso se llevó a cabo mediante una reunión con cada individuo para que ingresaran sus datos de Google de forma segura en un computador para descargar sus datos. A continuación se detallan los pasos seguidos para poder obtener este tipo de datos.

1. Asegurarse que la persona a quien se le solicitan los datos tengan un smartphone Android y hayan activado la opción de permitir a Google guardar sus datos. Esto se puede verificar en la aplicación de Google Maps viendo si se puede revisar la cronología.
2. Para poder extraer los datos se debe entrar a la cuenta de Google de cada individuo desde un computador, es por esto que se les solicitó a las personas que se presentaran para ingresar a su cuenta de forma segura mientras sus datos son extraídos.
3. Una vez que se ingresa la cuenta de Google de alguna persona en Google Maps, lo que se debe hacer es descargar los datos en formato KML, que corresponde a un archivo del estilo que utilizan los sistemas de información geográfica. Para esto, se debe seleccionar la pestaña de menú que se encuentra en la esquina superior izquierda de la pantalla y elegir la opción llamada “Tu cronología” (ver Figura 4.1, rectángulo rojo).
4. Luego, se selecciona el día que se quiere descargar y en la esquina inferior derecha aparecerá un símbolo de engranaje que representa un menú de opciones, se debe seleccionar y luego elegir “Exportar los datos correspondientes a este día en formato KML” (ver Figura 4.2, rectángulo rojo).
5. Repetir para cuantos días se quiera tener información.

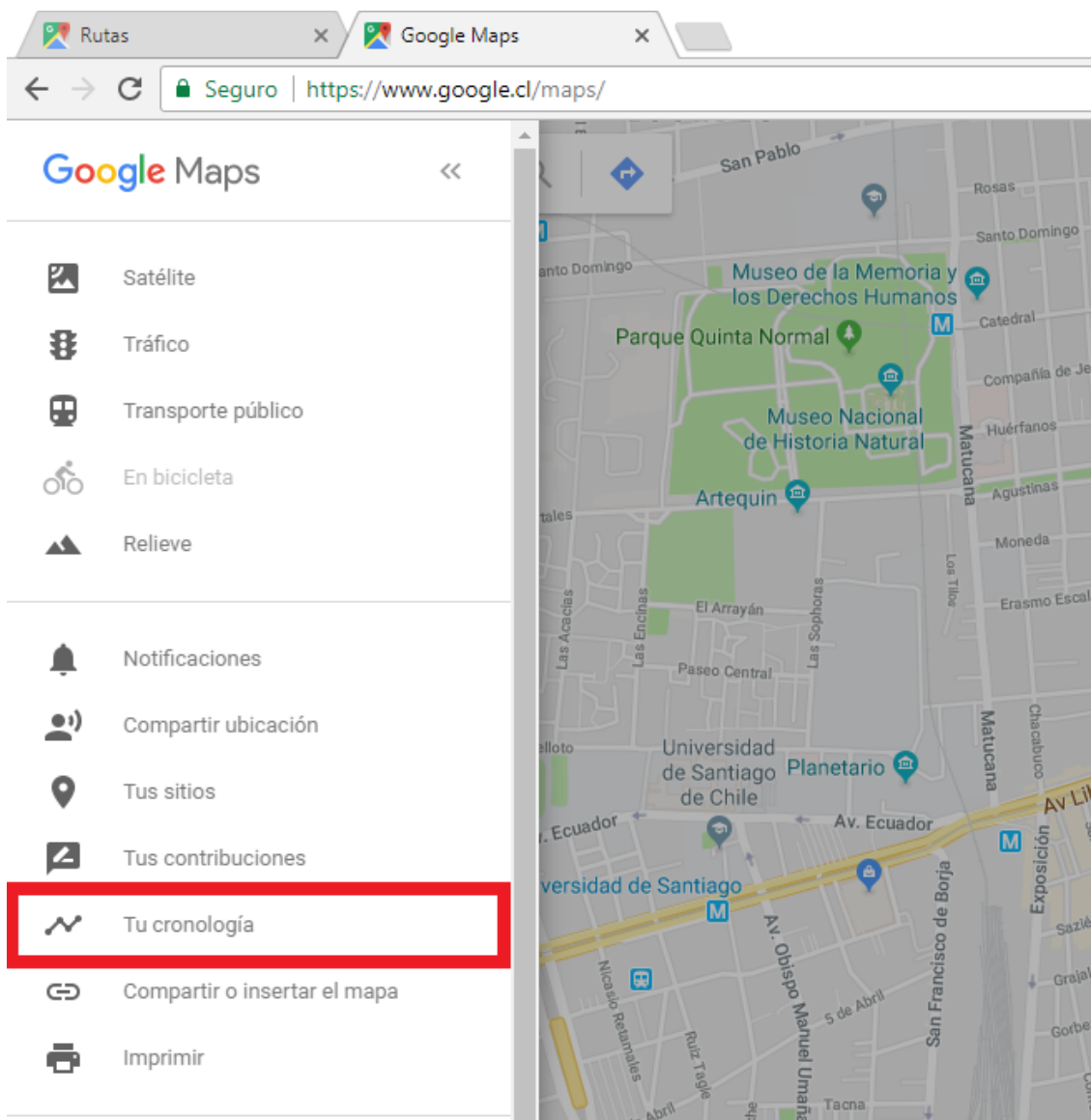


Figura 4.1: Cómo acceder a la cronología en Google Maps.

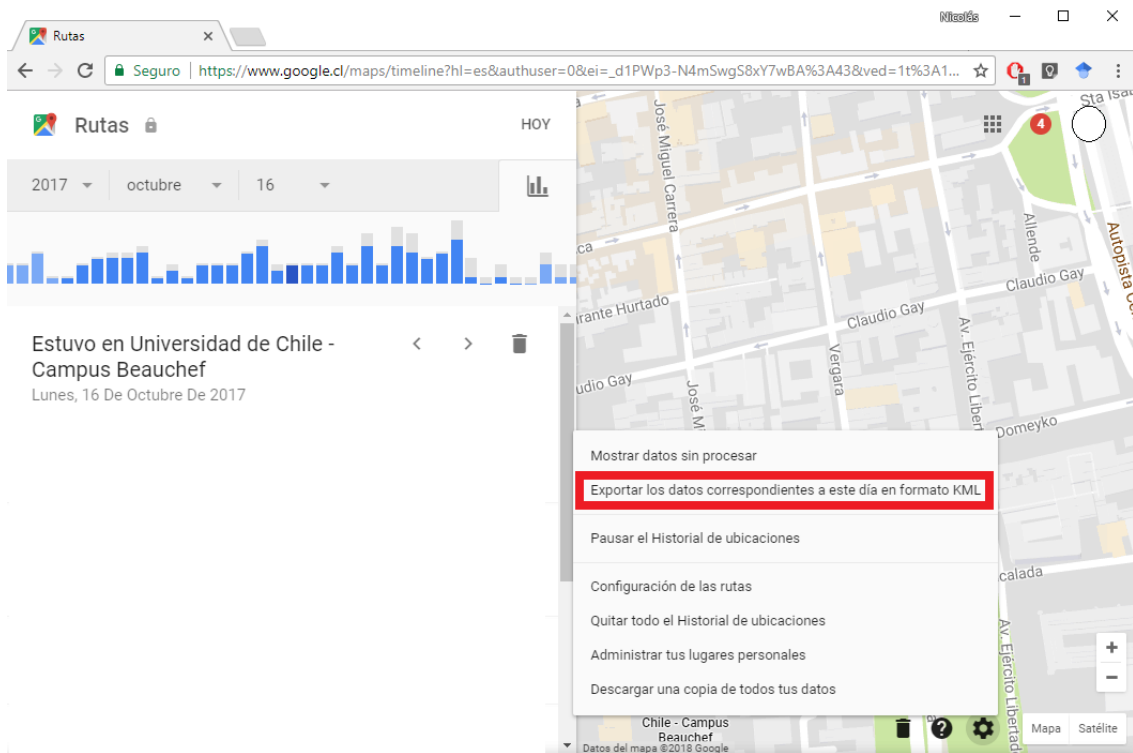


Figura 4.2: Cómo acceder al menú de herramientas para descargar archivo *.kml en Google Maps.

Una vez descargados los datos en el formato deseado, utilizando el software QGIS desarrollado por QGIS Development Team (2009), se deben seguir los siguientes pasos para obtener la información de la forma que se ve en la parte inferior de la Figura 4.3, Figura 4.4 y Figura 4.5.

1. Primero, se deben importar todos los archivos recolectados con extensión *.kml de una persona a QGIS.
2. Luego, se debe descargar el mapa proporcionado por OpenStreetMaps asociado al área de interés y se cargan únicamente las líneas para una mejor visualización (para más detalle ver Gandhi (2017)).
3. Posteriormente, se deben filtrar para cada día los viajes que son de interés. Es decir, que hay que filtrar las líneas de cada día que representen el viaje que se desea analizar, por lo que para esto hay que tener información declarada del viaje (origen y destino).
4. Además, si en algún día no se logra identificar el viaje, por una muy mala calidad de los datos, entonces simplemente se elimina del estudio.
5. Se deben dejar seleccionadas las líneas correspondientes al viaje de interés.
6. Con las líneas seleccionadas, se debe seleccionar la “herramienta vectorial” de la barra de herramientas de QGIS, en la lista desplegable que aparece se debe seleccionar “herramientas de geometría”, de la misma manera, seleccionar “multipartes a partes sencillas” y ejecutar como “proceso por lote”. Con esto las rutas se guardan en archivos independientes.
7. Finalmente, se deben borrar los archivos importados originalmente, es decir, solo se

mantiene el mapa de OpenStreetMaps. De esta manera se cargan las rutas obtenidas del paso anterior. El resto corresponde a detalles estéticos.

El paso siguiente consiste en utilizar el procedimiento anterior para recopilar datos de tres sujetos, que proveen información de tres modos de transporte distintos, siendo estos el transporte público, automóvil y caminata. En la Figura 4.3, Figura 4.4 y Figura 4.5 se muestran los resultados de esta recopilación y procesamiento de datos, en contraste con las alternativas que brinda Google Maps para realizar el viaje con el mismo origen y destino.

En la Figura 4.3 se presenta al “usuario caminata”, quien realiza un viaje corto, entre su casa y la universidad. Los datos recolectados corresponden a días laborales de Agosto del 2017. En primer lugar, se puede apreciar que todos los datos recolectados excepto uno pueden ser comparados directamente con su contraparte proporcionada por Google Maps, en particular se puede ver que la ruta más usada (izquierda) coincide con algo que se puede hacer utilizando las rutas sugeridas y la ruta sin identificar solo posee un trazo erróneo. En segundo lugar, se debe notar que aunque hayan rutas que aparentemente son iguales, ningún elemento vectorial es igual a otro, esto tiene que ver directamente con lo discutido anteriormente, donde se menciona que estos datos no vienen definidos sobre una estructura de red. Una de las cosas que muestra este individuo es que se puede obtener datos de buena calidad, lo cual se puede explicar por la baja velocidad de desplazamiento del sujeto y la poca interferencia de señal en la zona en que fue observado.

Por otro lado, en la Figura 4.4 se presenta al “usuario transporte público”, en este caso la mayor parte de las rutas se realizan en una combinación de metro y micro para ir desde el hogar al lugar de trabajo, en el periodo de práctica profesional realizada en Febrero del 2017. Una de las cosas interesantes sobre estos datos es que representan los viajes desde la primera vez que se realizan, por lo que podrían estar mostrando el aprendizaje realizado por el sujeto para trasladarse. Además, se puede apreciar que la mayor parte de los datos se parece a grandes rasgos a una de las alternativas que presenta Google Maps, con algunos tramos más distorsionados que otros, solo se destacan las dos rutas que se escapan a la izquierda y derecha de la figura, los cuales no se logran identificar. De esta manera, es una posibilidad lograr traducir estos datos a una estructura de red donde se identifique en su mayoría el trayecto de manera correcta. En comparación a los datos presentados en la Figura 4.3 se aprecia un deterioro en la calidad de los datos, a pesar de que este viaje es relativamente recto y por lo tanto no requiere demasiado detalle y las velocidades tampoco son muy altas.

Finalmente, en la Figura 4.5 se presenta al “usuario auto”, quien realiza un viaje relativamente largo de alrededor de 14 kilómetros entre su hogar y lugar de trabajo, en periodo de practica profesional realizada entre Enero y Febrero del 2017 . Se puede apreciar una gran dispersión entre las rutas, además de una baja precisión sobre la red, sin embargo, en comparación con el mapa de Google, se podría decir que las rutas son asignables a las dos rutas que utilizan Vespucio Norte. Además, se debe considerar que las rutas pasan por una de las condiciones geográficas más complejas de Santiago, el Parque Metropolitano de Santiago. Teniendo que pasar incluso por el túnel que lo atraviesa.

A modo de conclusión general, esta aplicación piloto sugiere que este tipo de datos muestra mucho potencial, debido a que son preferencias reveladas y podrían llegar a ser utilizados sobre una red real. La principal desventaja de este tipo de datos es que son difíciles de con-

seguir y filtrarlos para colocarlos sobre una red real sobrepasa el alcance de este trabajo. Además, dadas las características del conjunto de consideración, al ser un constructo, no puede ser observado en base a únicamente preferencias reveladas. Sin embargo, esta herramienta se presenta como una buena alternativa para utilizar el método que genera el conjunto de consideración supuesto en base a las alternativas experimentadas por las personas (CE), mencionado en el capítulo 3 sección 3.4. Donde por un lado se muestra que dicho método posee factibilidad práctica, ya que este tipo de datos logra identificar las elecciones realizadas en un periodo de tiempo tan largo como se desee y se tenga la capacidad de procesar, pudiendo de esta manera armar el conjunto de consideración supuesto para dicho método. Por otro lado, se refuerza el potencial de este tipo de datos, al ser capaces de implementar dicho método, el cual mostró ser capaz de recobrar los parámetros del modelo de elección presentado en el capítulo 3. A pesar de lo anterior, no es posible desarrollar esta herramienta en el alcance de esta memoria debido a la complejidad de procesar los datos.

Se debe hacer notar que los datos pasivos recolectados por Google Maps requieren que no se pierda la señal del teléfono inteligente, por lo que dependerá mucho del país y de las condiciones de cada ruta. Se recuerda también que la muestra obtenida es muy pequeña y es útil de forma exploratoria, sin embargo no se deben sacar conclusiones definitivas a partir de ella. Una posible forma de desarrollar este tipo de datos para el contexto deseado es generar una aplicación para teléfonos celulares inteligentes que recolecte los datos pasivos de sujetos seleccionados para utilizarla. Lo anterior puede generar una mejor calidad en los datos y permitir un mejor procesamiento para el contexto deseado, pudiendo incluso presentar los resultados sobre una estructura de red.

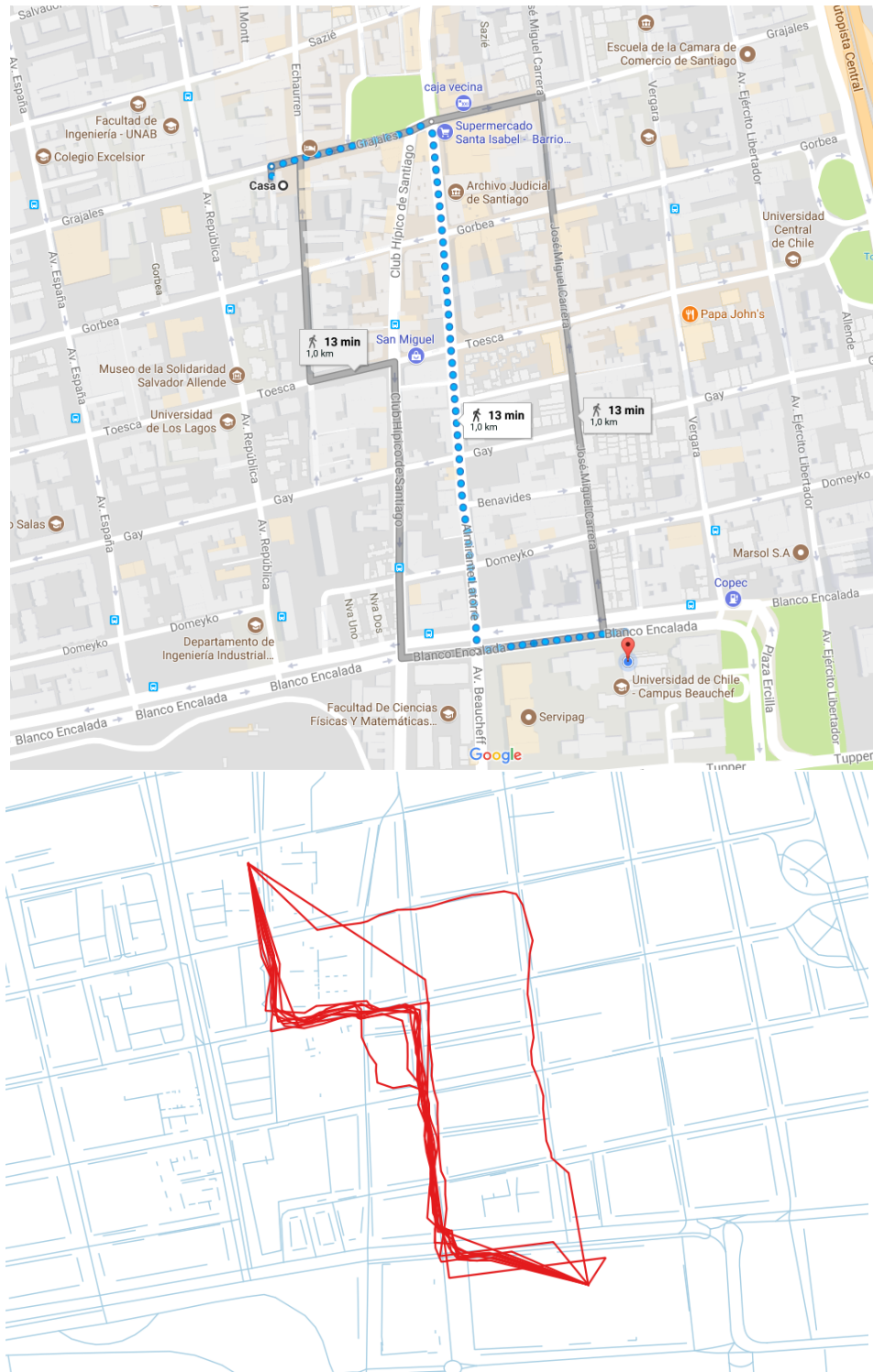


Figura 4.3: Rutas Google Maps y rutas de datos pasivos del “Usuario Caminata”.



Figura 4.4: Rutas Google Maps y rutas de datos pasivos del “Usuario Transporte Público”.

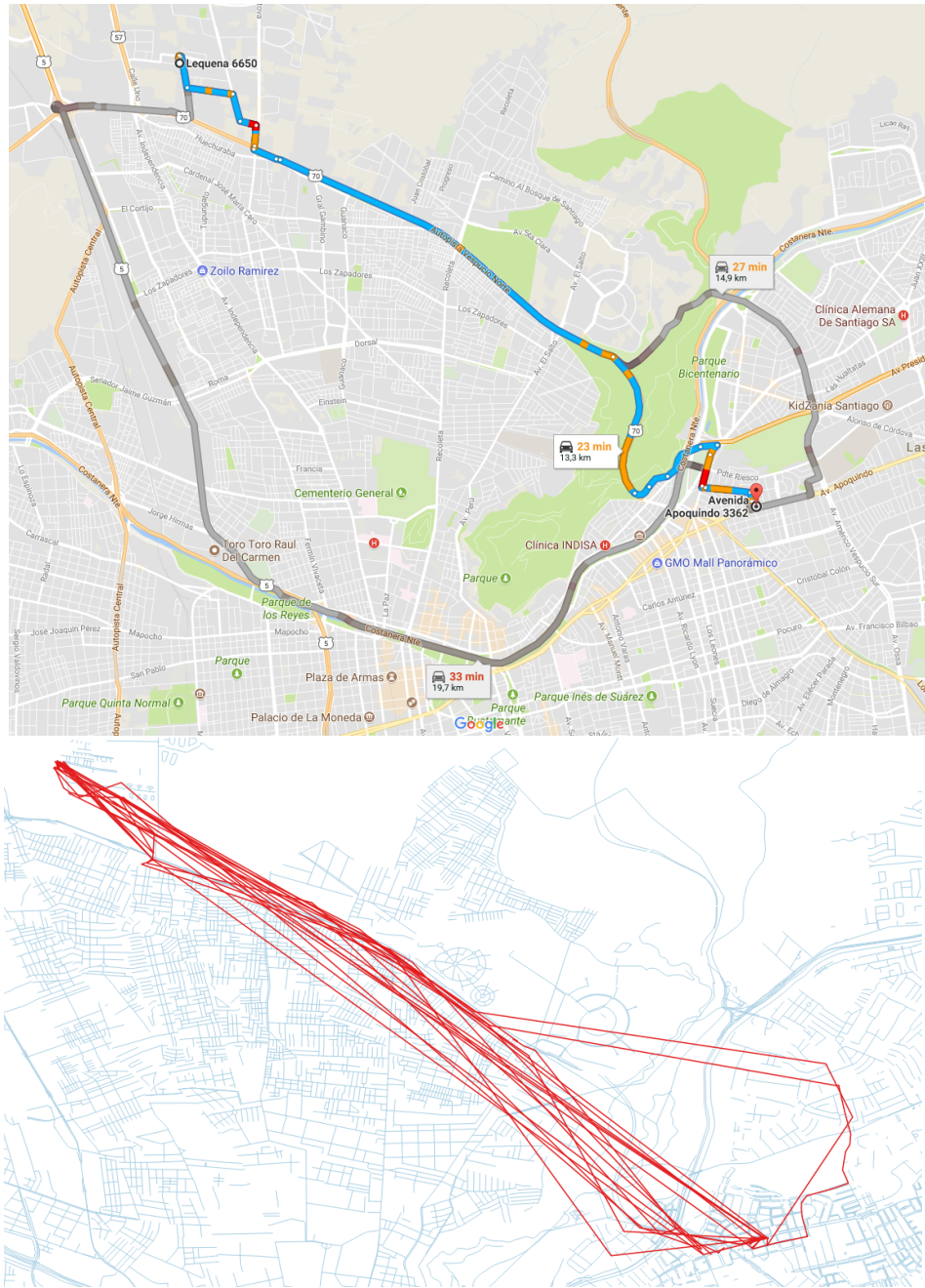


Figura 4.5: Rutas Google Maps y rutas de datos pasivos del “Usuario Auto”.

4.3. Encuesta de preferencias declaradas online

Como segundo método de recolección de datos, se estudio la implementación de una encuesta online de preferencias declaradas, inspirada en los trabajos del área de marketing como se muestra en Brown y Wildt (1992), donde se estudia el tamaño del conjunto de consideración y Hauser (2014), donde se analizan las posibles heurísticas que utilizan los usuarios, así como también las maneras de reconocerlas. Esta encuesta se divide en cuatro secciones. La primera consiste en una recolección de datos socio-económicos, la cual es seguida de un experimento de consideración de rutas para transporte privado. Posteriormente se presenta un experimento de elección basado en las alternativas consideradas. Finalmente, se realiza un ejercicio de preguntas sobre el nivel de representación de algunas reglas de decisión que pueden describir el proceso de generación del conjunto de consideración.

4.3.1. Descripción de la encuesta

A continuación, se presentan las distintas secciones de la encuesta, donde se describe cómo se presenta cada una de ellas en la encuesta y el por qué se hizo de esta manera.

Datos socio-económicos

Al igual que en la mayoría de las encuestas, con el fin de caracterizar a la población encuestada y utilizar variables socio-económicas para la modelación, se realizan una serie de preguntas generales y específicas para el tema tratado. La recolección de estos datos se dividió al principio y final de la encuesta para no abrumar a los participantes.

Los datos preguntados en la primera sección de datos socio-económicos se presentan a continuación:

- Trabaja o estudia en Beauchef
- Género
- Edad
- Nivel educacional
- Ocupación
- Tiempo vivido en Santiago
- Principal modo de transporte al trabajo o estudio

Se debe notar que la primera pregunta, “Trabaja o estudia en Beauchef” se refiere a si la persona pertenece de alguna manera a la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, lo cual puede resultar de importancia debido a que los experimentos están diseñados utilizando de origen de las rutas la facultad. Las preguntas de “Género”, “Edad”, “Nivel educacional” y “Ocupación” son formuladas comúnmente para dividir grupos socio-económicos que suelen realizar actividades similares. Por otro lado, la pregunta “Tiempo vivido en Santiago” puede ser relevante ya que puede servir como indicador del nivel de

conocimiento de la red presentada, el cual puede ser distinto según el tiempo que la persona ha vivido en Santiago. Finalmente, la pregunta sobre el “Principal modo de transporte al trabajo o estudio” busca distinguir a la gente que según el modo de transporte se vean más representados en un experimento sobre elección de rutas en transporte privado.

En la segunda parte se realizan preguntas enfocadas a obtener datos sobre el hogar. Los datos obtenidos a partir de estas preguntas se presentan a continuación:

- Posesión de licencia
- Número de personas en el hogar
- Número de licencias en el hogar
- Vehículos motorizados disponibles
- Días de uso de vehículo motorizado
- Ingreso mensual líquido del hogar

La pregunta “Posesión de licencia”, es similar al modo de transporte y espera dar cuenta de que el individuo esté más familiarizado con el entorno del transporte privado. Las preguntas “Número de personas en el hogar”, “Número de licencias en el hogar” y “Vehículos motorizados disponibles” buscan dar cuenta de la competencia por el automóvil que se puede generar en un hogar. Por otro lado, la pregunta “Días de uso de vehículo motorizado” también intenta dar cuenta del nivel de realismo con el que las personas toman los experimentos. Finalmente, la pregunta sobre el “Ingreso mensual líquido del hogar” busca distinguir posibles grupos socio-económicos.

Todos estos datos fueron pensados para poder realizar modelos y correcciones que sean necesarios. En este punto se pueden haber hecho otras preguntas que luego del periodo de lanzamiento han surgido como potencialmente importantes.

Experimento de consideración

El experimento de consideración realizado es el núcleo de lo que hace a esta encuesta útil para el tópico estudiado. En esta etapa de la encuesta se les presenta a las personas un contexto específico, el cual consiste en realizar un viaje en auto entre el campus Beauchef de la Universidad de Chile y Plaza Baquedano, un día jueves normal, a las 13:30 hrs con propósito trámite. La idea de preguntar un contexto en específico sobre una situación futura es que las personas no estén sesgadas por algún contexto particular, tal como se muestra en Brown y Wildt (1992).

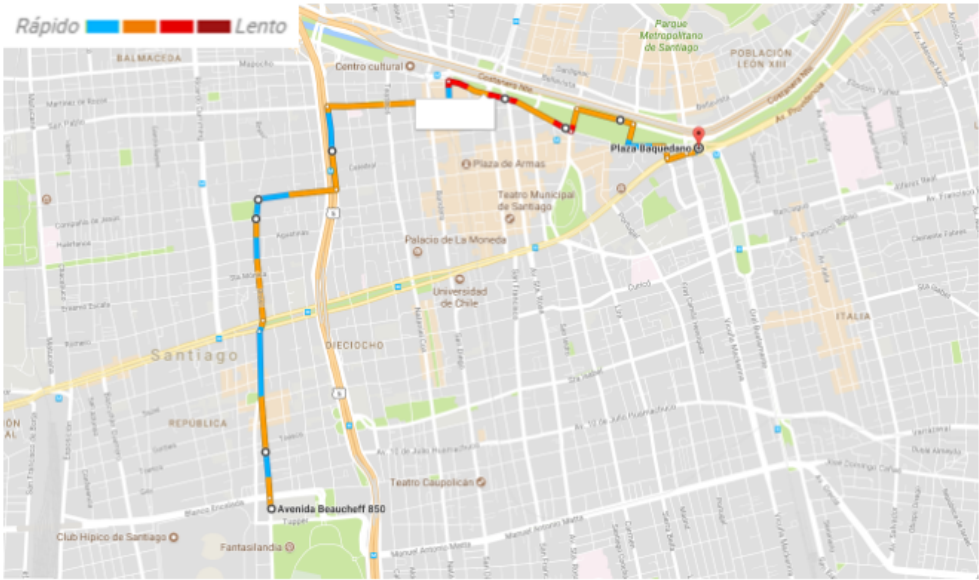
A cada persona se les presenta 10 o 15 alternativas aleatoriamente seleccionadas de un grupo de 20 existentes, esto se considera como un diseño asistido según Brown y Wildt (1992) ya que las alternativas son presentadas y no deben ser pensadas por las personas. En este caso, para cada alternativa se muestran las calles que utiliza la ruta, el tiempo mínimo y máximo, la distancia en ruta y un mapa con la ruta con una escala de congestión, de donde se desprenden variables como el número de virajes en cada sentido. Las 20 alternativas pueden aparecer con 3 valores distintos de tiempo cada una, es decir, existen 3 bloques distintos, donde para cada bloque existen tiempos mínimos y máximos diferentes. Esto se hace para

poder generar varianza en los atributos, sin embargo, las demás variables no se modifican entre bloques ya que son deducibles a partir del mapa. Para cada ruta presentada se pregunta si la consideraría para realizar dicho viaje. En la Figura 4.6 se presenta el ejemplo de una ruta preguntada.

Responda las siguientes preguntas respecto a las rutas que se presentan a continuación.

Suponga que debe realizar un viaje para hacer un trámite desde Beauchef 850 a la Plaza Baquedano en automóvil, un típico día jueves en semana laboral a la hora de almuerzo (13:30 hrs).

¿Consideraría utilizar la siguiente ruta para dicho viaje?
Almirante Latorre/Brasil - Compañía - Manuel Rodríguez - San Pablo - Ismael Valdés Vergara/Cardena José María Caro/Merced
Tiempo mínimo: 22 min
Tiempo máximo: 52 min
Distancia: 6.0 km



Sí

No

Figura 4.6: Ejemplo de pregunta en el ejercicio de consideración.

Experimento de elección

Posterior al experimento de consideración se encuentra un experimento de elección, donde se le pide a las personas que ordenen en un ranking las alternativas que consideraron en la sección anterior. Solo las opciones consideradas son mostradas en este segmento. Además, se presentan con la misma cantidad de información que antes. En la Figura 4.7 se puede

apreciar el formato en que este ejercicio es presentado.

A continuación, se muestran las alternativas que indicó que consideraría utilizar de la sección anterior.

Realice un ranking de las alternativas, donde 1 es la mejor alternativa.

1 2 3 4 5 6

Almirante Latorre - Toesca/Santa Isabel - Vicuña Mackenna
Tiempo mínimo: 16 min
Tiempo máximo: 37 min
Distancia: 4.8 km



Almirante Latorre - Toesca/Santa Isabel - Santa Rosa - Alameda
Tiempo mínimo: 13 min
Tiempo máximo: 29 min
Distancia: 4.5 km



Figura 4.7: Ejemplo de pregunta en el ejercicio de elección.

Preguntas sobre heurísticas

Una manera de conocer las heurísticas que las personas utilizan para construir su conjunto de consideración en elección de rutas, corresponde a preguntar directamente por la heurística declarada, tal como se menciona en Hauser (2014). Con este fin se solicita elegir el nivel de representación de 9 heurísticas que podrían haber utilizado las personas para construir su conjunto de consideración. Estas heurísticas son generadas a partir de funciones de costo

generalizado que se se utilizan para calcular rutas mínimas en el artículo de Bekhor et al. (2006). Como se puede ver en la Figura 4.8, los 5 niveles de representatividad varían desde “no me representa” hasta “me representa perfectamente”, destacando que no necesariamente estos niveles son equidistantes en cuanto a representatividad. Además, en la parte inferior de esta figura se muestra que las personas podían decir alguna otra regla que los representara.

¿Qué tan representado se siente con las siguientes frases que describen el proceso de consideración de rutas?

	Me representa perfectamente	Me representa muy bien	Me representa moderadamente bien	Me representa ligeramente bien	No me representa
"Considero las alternativas de menor tiempo promedio"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
"Considero las alternativas de menor tiempo en un escenario optimista"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
"Considero las alternativas de menor tiempo y distancia en promedio"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
"Considero las alternativas con menos virajes a la izquierda"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
"Considero las alternativas con más tramos en calles seguras"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
"Considero las alternativas que utilizan vías de alta capacidad (Ej. Alameda)"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
"Considero las alternativas que utilizan más autopistas"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
"Considero las alternativas de menor distancia"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
"Considero las alternativas con menor número de calles distintas"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Si existe algún otro criterio que no se haya mencionado, por favor indicar aquí (opcional)

Figura 4.8: Preguntas sobre heurísticas utilizadas para generar el conjunto de consideración.

4.3.2. Análisis estadístico descriptivo de los datos

Luego de un periodo de 2 semanas aproximadamente de recolección de datos comprendido entre el 29 de Agosto y el 8 de Septiembre del 2017, donde la encuesta se publicó en el portal web de la Universidad de Chile y en varios grupos de redes sociales como Facebook, ofreciendo un sorteo de una Gift Card de 20 mil pesos a los participantes, se obtuvo un total de 192 respuestas. Lo anterior causa que esta encuesta sea etiquetada como una muestra de conveniencia, con el fin de explorar distintos ámbitos del conjunto de consideración.

Al momento de revisar las respuestas, se obtienen 189 encuestas que se consideran válidas, omitiendo solo a las personas que consideraron 1 alternativa en el ejercicio de consideración. Adicionalmente, para evitar posibles sesgos por falta del interés en responder la encuesta, se estudió la posibilidad de descartar las encuestas que fueron respondidas muy rápidamente. Sin embargo, este filtro adicional fue descartado ya que las personas que se demoraron menos en responder la encuesta respondieron de manera lógica y por lo tanto no había otra forma de filtrar los datos.

Como ya se presentó en la sección de datos socio-económicos, se realizaron una serie de preguntas de interés. En la Figura 4.9, Figura 4.10, Figura 4.11 y Figura 4.12, se presenta la comparación de algunos datos recopilados por la encuesta, comparándolos con los datos equivalentes de la encuesta origen-destino 2012 de SECTRA (2014). De esta manera se compara el tipo de población encuestada con la que realmente existe en Chile.

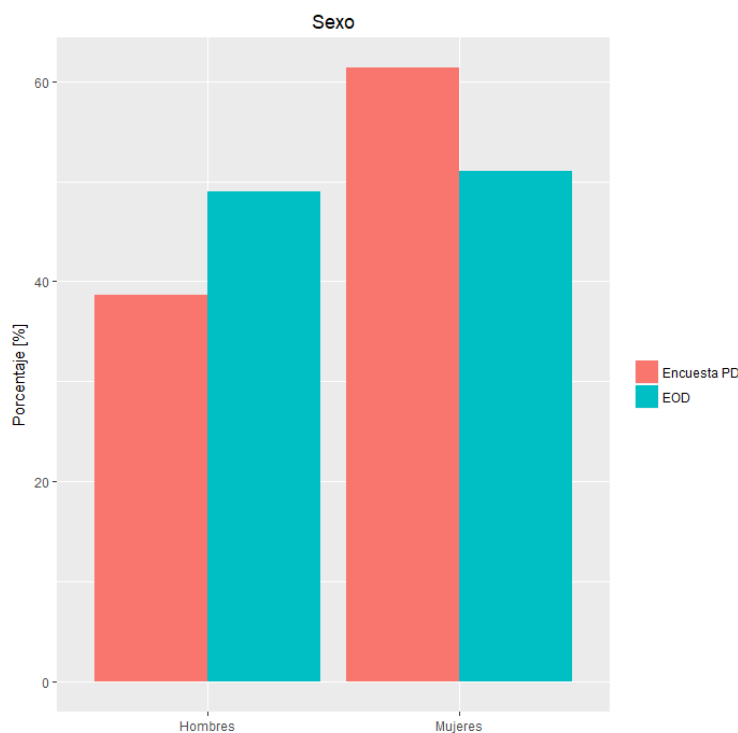


Figura 4.9: Comparación de sexo entre Encuesta PD y EOD 2012.

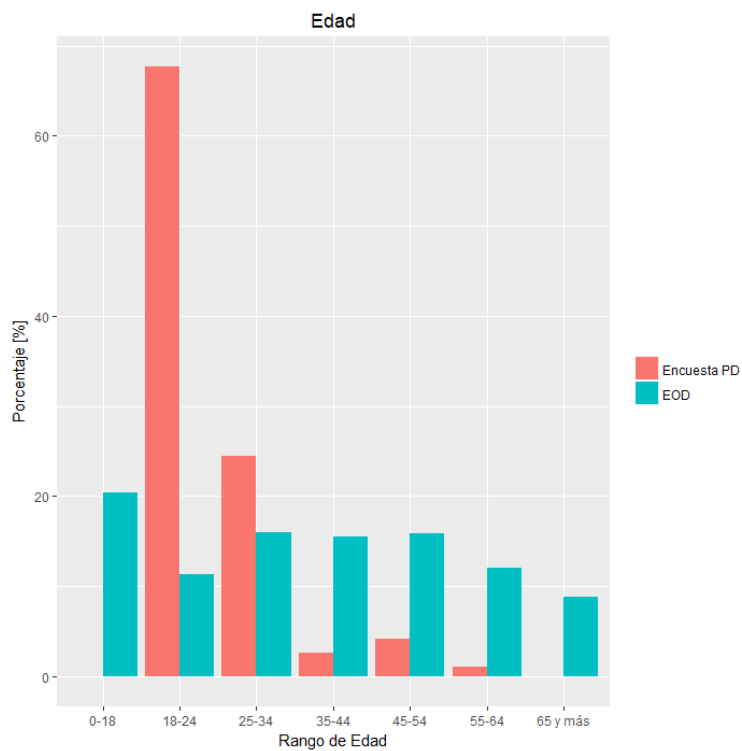


Figura 4.10: Comparación de edad entre Encuesta PD y EOD 2012.

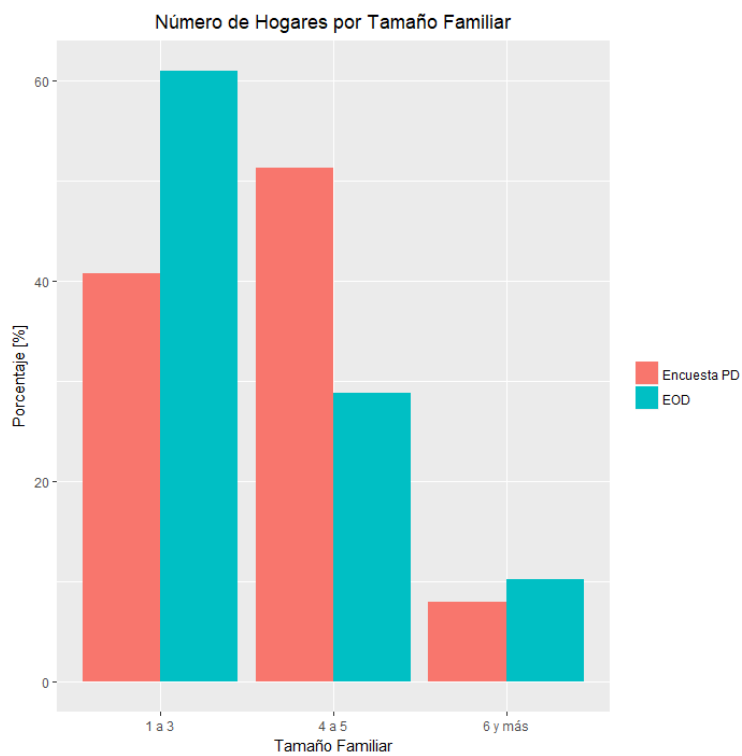


Figura 4.11: Comparación del tamaño del hogar entre Encuesta PD y EOD 2012.

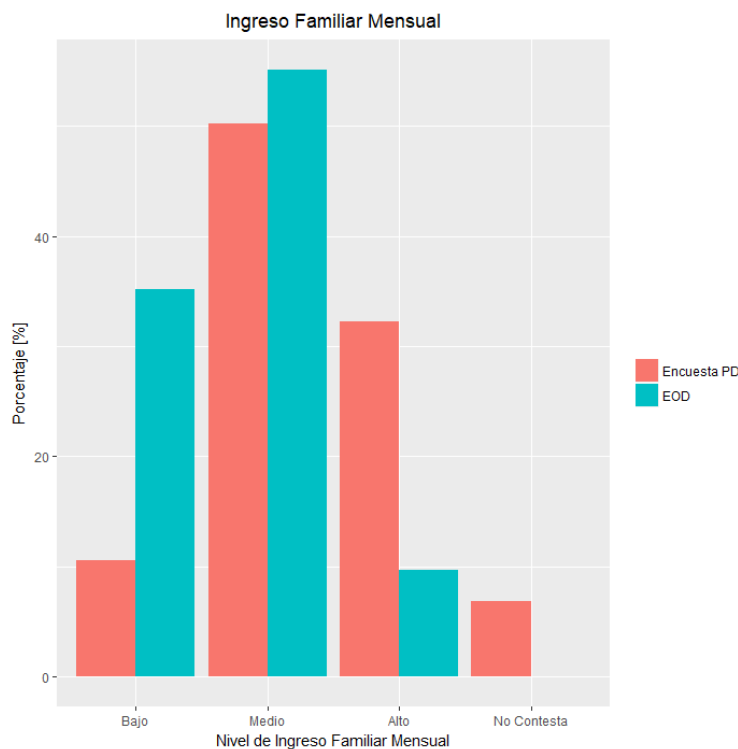


Figura 4.12: Comparación de ingreso entre Encuesta PD y EOD 2012.

Como se puede apreciar las principales diferencias se ven en la edad e ingreso, debido a que la encuesta fue mayormente difundida a la gente de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. En el caso de la edad se aprecia un sesgo hacia la población entre 18 y 24 años, mientras que en el ingreso, se observa un sesgo hacia los ingresos más altos, lo que también puede ser atribuido a la difusión web de la encuesta. Esta encuesta no pretende ser representativa de la ciudad de Santiago, pero las caracterización de la muestra permite dar luces sobre los posibles sesgos que pueden existir.

En el Anexo A, la Tabla A.1, Tabla A.2 y Tabla A.3, muestran las respuestas de la encuesta en detalle.

4.3.3. Discusión y extensiones

A partir de la sección de preguntas socio-económicas, surge una de las ideas sobre haber podido preguntar en este estudio si la persona utiliza Waze ,Google Maps o alguna otra aplicación que funcione como ayuda a la navegación, ya que potencialmente las personas que están acostumbradas a estas aplicaciones, pueden no pensar en la ruta que utilizan y por lo tanto su proceso de generación de rutas puede verse disminuido. En particular, este tipo de personas podrían haber dejado de generar su conjunto de consideración por largos periodos, causando que al responder el experimento sus respuestas se vean sesgadas por un proceso improvisado. Es por esto que la finalidad de esta pregunta podría haber sido a modo de comprobar esta hipótesis y a la vez corregir el sesgo causado en los modelos por este grupo

de personas.

Por otro lado, para los experimentos de consideración y elección es importante destacar que la gente puede no saber leer un mapa y su interpretación puede resultar ambigua causando sesgo en las respuestas. Si bien este fenómeno puede ser abordado en parte por el nivel educacional de las personas, no se sabe con certeza el daño que se puede causar. En este experimento en particular, hay muchas otras variables que son presentadas junto con el mapa, y no se necesita específicamente orientación para responder las preguntas, por lo que el sesgo causado podría ser menor de lo pensado.

En el caso del experimento de consideración, se debe notar que otra opción es mostrar todas las alternativas disponibles de una vez y no preguntarlas una a una, ya que de esta manera se representaría mejor la realidad. El método actual con el que se presentan los datos puede causar un sesgo de orden o lexicográfico, lo cual se puede ver en la sección 5.4. Este método alternativo no fue utilizado debido a que por temas de diseño web no fue factible.

4.4. Encuesta de preferencias declaradas sobre mapa

Como tercera herramienta se propuso realizar una encuesta con un ejercicio de consideración, donde se le pidió a los usuarios que dibujen las rutas que considerarían para realizar un viaje en un contexto dado. Este tipo de encuesta se encuentra bajo la metodología no asistida, como se habla en el artículo de Brown y Wildt (1992), para obtener el conjunto de consideración declarado por las personas. La hipótesis detrás de este experimento, es que al ser de carácter no asistido, el tamaño del conjunto de consideración declarado será menor.

En esta herramienta se presenta un mapa como el de la Figura 4.13, y se pide a los participantes que dibujen las rutas que considerarían bajo el mismo contexto de la encuesta online, el cual consiste en realizar un viaje en auto entre el campus Beauchef de la Universidad de Chile y Plaza Baquedano, un día jueves normal, a las 13:30 hrs. con propósito trámite. Además, se realizan las mismas preguntas de la sección socio-económica y de preguntas sobre heurísticas, de manera que solo se reemplaza el ejercicio de consideración y de elección por el mapa.

Al ser de carácter presencial, se han logrado obtener 5 respuestas, las cuales son suficiente para realizar un análisis comparativo con la encuesta online, pero no lo son para concluir. Las respuestas recolectadas se muestran en la Figura 4.14. A manera de análisis preliminar, se puede afirmar que se ve un buen potencial en este tipo de respuestas debido a que las personas se enfrentan a un escenario más realista, de hecho, si se contrastan estos resultados con los presentados en la sección 4.2 de datos pasivos se puede ver que el número de rutas distintas es similar, lo que indica que la principal ventaja de esta herramienta es acercarse a los resultados de preferencias reveladas.

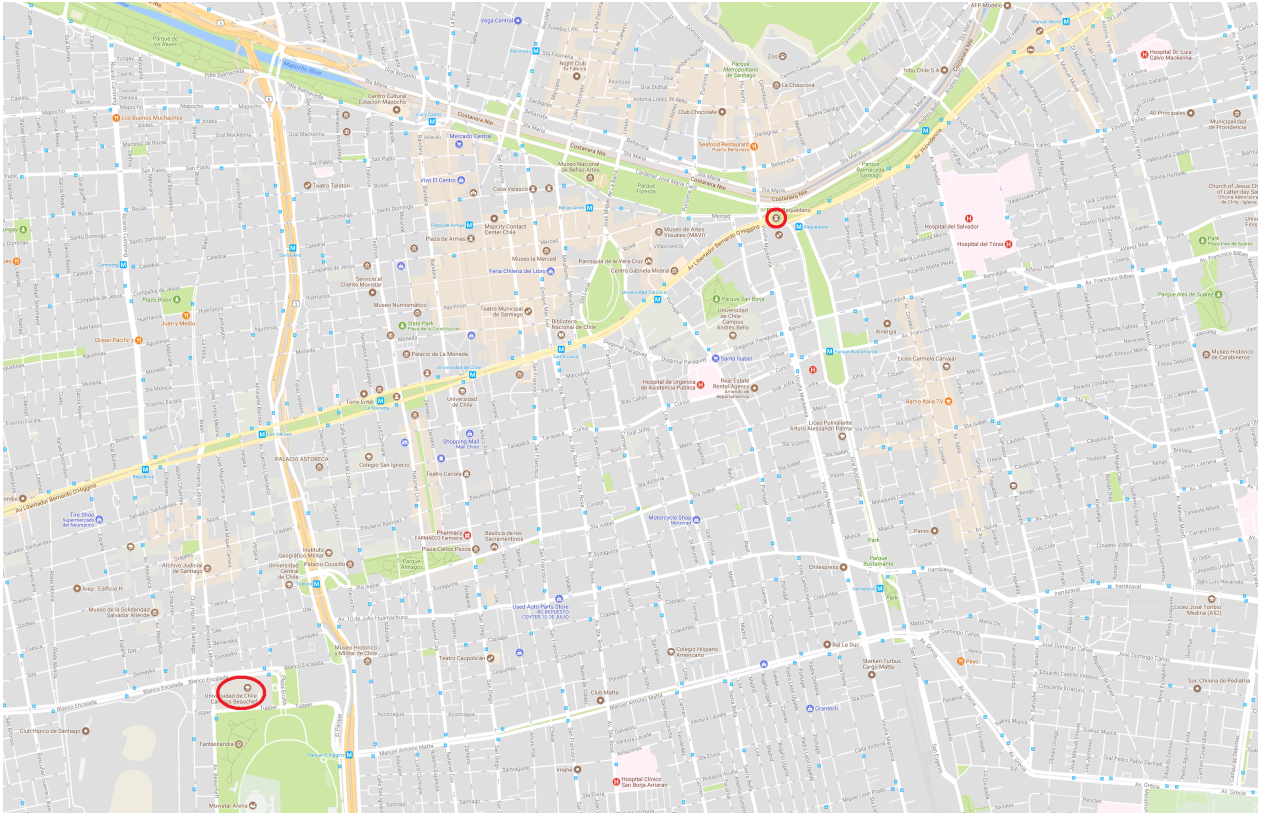


Figura 4.13: Mapa presentado para dibujar las rutas que el encuestado considera, 48x33 cm^2 .

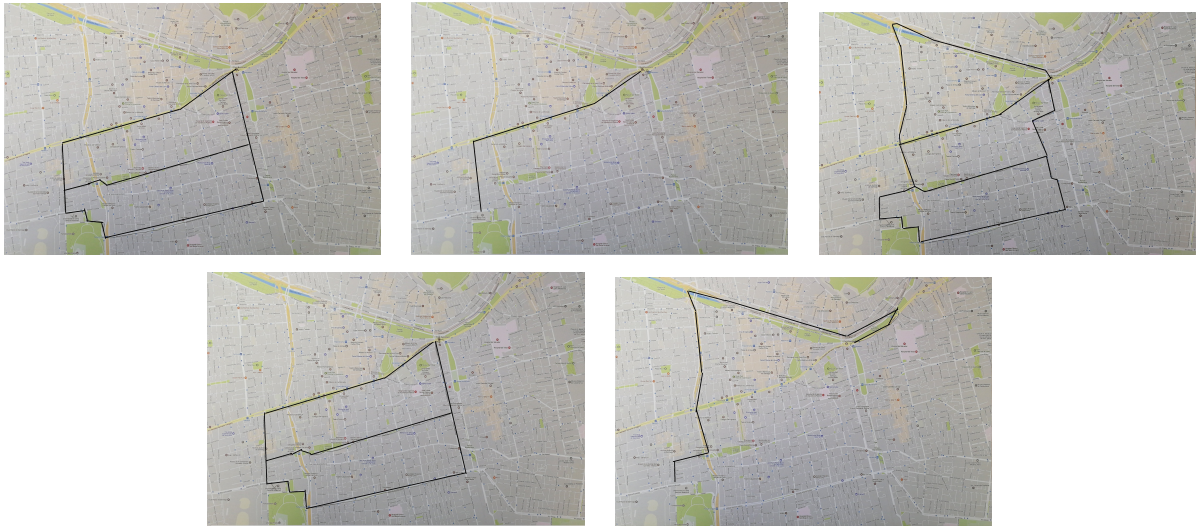


Figura 4.14: Respuestas de encuesta sobre mapa recolectadas.

En comparación a la encuesta online, en esta encuesta se presenta mucha menos información al usuario al momento de realizar el experimento, por ejemplo, no se presentan los tiempos de viaje mínimo y máximo, ni los niveles de congestión o la distancia, ya que las personas deben construir las rutas y no es posible brindar esta información. Además, con-

seguir esta pequeña muestra requiere de mucho más esfuerzo en cuanto a conseguir a los encuestados para participar de manera presencial. La encuesta en si tiene unos 10 minutos de duración aproximadamente, pero para procesar los datos de todas las rutas que la gente considera y dejarlos en una base de datos computacional se requieren horas de trabajo.

Capítulo 5

Caracterización del proceso de consideración mediante datos de preferencias declaradas

5.1. Introducción

A partir de las 3 herramientas para la recolección de datos sobre el conjunto de consideración, se tiene que sólo los datos de preferencias declaradas online permitieron obtener (en los plazos de esta memoria) una muestra suficientemente grande para realizar algún tipo de análisis estadístico. Los datos recopilados por las otras fuentes se utilizarán de manera cualitativa para comparar alguno de los análisis realizados.

En este capítulo caracteriza el conjunto de consideración en elección de rutas mediante distintos análisis y modelos. En primer lugar se estudia el tamaño del conjunto a partir del experimento de consideración realizado en la encuesta online mostrado en la sección 4.3 y del experimento de consideración realizado en la encuesta sobre mapa presentado en la sección 4.4, donde se busca saber si existe un único tamaño o si depende, por ejemplo, del grupo socio-económico. Posteriormente, se analizan los niveles de representación de las heurísticas presentadas en la encuesta online, con el fin de observar lo que las personas dicen que hacen para considerar rutas e identificar distintos grupos socio-económicos que se identifiquen con distintas heurísticas. Luego, se realiza un modelo logit binario de consideración utilizando las respuestas del experimento de consideración de la encuesta online, a partir del cual se busca modelar la heurística que las personas utilizan e identificar las variables que son relevantes en este proceso.

Por otro lado, un análisis adicional que se presenta a partir del modelo logit binario de consideración, es el de probar si los parámetros serían iguales a los de un modelo logit de elección generado a partir de las respuestas del experimento de elección de la encuesta online. Esta pregunta surge debido a que los parámetros son una representación del peso que tiene cada variable en la utilidad que las personas reciben de una alternativa, y por lo tanto podrían ser intrínsecos de las personas y/o contexto presentado.

Finalmente, se presenta un análisis de distintos métodos utilizados para generar el conjunto de consideración supuesto, de manera análoga al análisis presentado en el capítulo 3 donde se realiza una simulación de Monte Carlo. Esto busca corroborar con los datos de la encuesta online las conclusiones obtenidas anteriormente.

Se destaca de antemano que los modelos logit en este capítulo son estimados con el software R (R Development Core Team, 2008), utilizando la librería de maximización de verosimilitud “maxLik” de Henningsen y Toomet (2011). Además, los errores estándar de los modelos que utilizan más de una observación por persona, son estimados utilizando la corrección a los errores estándar para series de tiempo en modelos de elección discreta que se puede encontrar en la ecuación 10 del artículo de Daly y Hess (2010).

5.2. Tamaño del conjunto de consideración

Con la inspiración del artículo de Brown y Wildt (1992), donde se estudia el tamaño del conjunto de consideración para productos como comida rápida, refrescos y gasolina. En esta sección se utilizan ambos experimentos de preferencias declaradas para estudiar el tamaño del conjunto de consideración bajo el contexto de elección de ruta. En dicho artículo se habla de dos maneras para preguntar las alternativas consideradas. La primera corresponde al diseño asistido, que sería equivalente al caso de la encuesta de preferencias declaradas online, ya que a las personas se les presentan las alternativas existentes. La segunda consiste en un diseño no asistido, como es el caso de la encuesta de preferencias declarada sobre mapa, donde los encuestados deben pensar en cuáles son las alternativas existentes para considerarlas.

5.2.1. Utilizando encuesta de preferencias declaradas online

El primer resultado que se obtiene, utilizando las respuestas de las personas que respondieron el experimento de consideración, se obtiene que el tamaño promedio del conjunto de consideración entre las 189 observaciones válidas es de 5.74. Pero dada la formulación del experimento, donde a algunas personas se les mostraron 10 alternativas y a otras 15, surge la pregunta sobre si el número de alternativas consideradas depende de la cantidad de alternativas que se les mostraba a las personas.

Para responder dicha pregunta, primero se realiza un t-test calculado según la ecuación 5.1, donde se prueba la hipótesis nula de que el promedio del tamaño del conjunto para el grupo que se les presentaron 10 alternativas es igual al promedio del grupo al que se les presentaron 15 alternativas. Para esto, se tiene que el promedio de alternativas consideradas para las personas que se les mostró 10 es de 4.72, mientras que para las que se les mostró 15 es de 6.72. El valor del estadístico para probar la hipótesis nula mencionada, junto con el valor crítico del test se muestra en la ecuación 5.2, lo que indica que para un 95 % de confianza se rechaza la hipótesis nula y por lo tanto los valores son estadísticamente distintos.

$$t - test = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}} \quad (5.1)$$

Donde \bar{Y}_i es el promedio de la muestra i , S_i^2 es la varianza de la muestra i y N_i es el tamaño de la muestra i .

$$t - test = 6,72 > t_{critico,95\%,187df} = 1,973 \quad (5.2)$$

El resultado anterior, indica que el tamaño del conjunto no es intrínseco de las personas, sino que depende de la situación planteada. Por otro lado, al tener solo dos muestras distintas, no es posible realizar un análisis que permita parametrizar el tamaño del conjunto en función del número de alternativas mostradas en este experimento. De esta manera estos resultados dan luces de lo que podría ser el tamaño del conjunto en este contexto, dado el diseño asistido con el que se creó la encuesta para recopilar los datos.

Como segundo estudio, se hace uso del análisis de varianza llamado ANOVA que permite saber si existe alguna diferencia estadísticamente significativa en el tamaño del conjunto para distintos grupos de personas. Lo que hace este análisis es realizar un test entre todas las combinaciones de grupos posibles. Por ejemplo, si se utiliza la variable edad, entonces ANOVA realiza un test para ver si el rango de edad 1 tiene un tamaño del conjunto de consideración distinto del rango de edad 2, así también para el rango de edad 3, y así sucesivamente todas las combinaciones.

En la Tabla 5.1 se presenta el valor p entregado por el análisis ANOVA para 11 variables socio-económicas disponibles en los datos que fueron incluidas de forma lineal en el modelo. Esto indica con qué probabilidad la diferencia del tamaño del conjunto entre los distintos grupos de una variable es cero. Además, a continuación se presenta una descripción de las variables utilizadas junto con la hipótesis sobre lo que debería suceder en el análisis.

- Automovilista

Los grupos contenidos en esta variable son dos, el primero contiene a una persona si se traslada a su lugar de trabajo o estudio en un vehículo motorizado privado por lo menos una vez a la semana y el segundo si nunca lo hace. Esta variable puede ser significativa en el análisis debido a que los automovilistas tienen experiencia en el proceso de consideración de rutas y pueden resultar más o menos exigentes al momento de considerar.

- FCFM

Contiene a los grupos de las personas que trabajan o estudian en la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile y de los que no. Esta variable podría resultar significativa debido a que el experimento de preferencias declaradas (de donde provienen los datos) está construido con un contexto de viaje que tiene origen en dicha facultad.

- Género

Contiene a ambos géneros masculinos y femeninos. No existe algún motivo que haga pensar que esta variable debiera ser significativa.

- Edad

Esta variable contiene a los distintos grupos de edad preguntados en la encuesta online. Esta variable podría resultar significativa si existe algún tipo de hábito que sea notable entre los distintos grupos.

- Nivel Educativo

Esta variable contiene a los distintos grupos de distintos niveles educativos preguntados en la encuesta online. Si se identificara alguna diferencia significativa se puede deber principalmente a la capacidad cognitiva que se podría representar por esta variable.

- Ocupación

Los grupos contenidos en esta variable son los de distintas ocupaciones presentados en la encuesta online. La hipótesis sobre esta variable es similar a la de “Nivel Educativo”.

- Tiempo en Santiago

Esta variable contiene a los distintos grupos según el tiempo que han vivido en la ciudad de Santiago. Puede ser considerada como un indicador del conocimiento de la red, por lo que puede causar una diferencia significativa debido a que una persona que conoce Santiago podría rechazar una ruta con tan solo ver por donde pasa, en cambio alguien que no podría llegar a procesarla.

- Licencia

Contiene a los grupos de las personas que tienen licencia de conducir y de las que no. A priori es una variante más débil de la variable “Automovilista”, por lo que no debería resultar significativa.

- Personas en el Hogar

Esta variable separa a los grupos por el número de personas que viven en el hogar. No existe ningún indicador de que esta variable deba ser significativa.

- Ingreso

Esta variable separa a los distintos grupos de ingreso preguntados en la encuesta online. No hay algún indicador que diga que debería ser significativa en el contexto de la encuesta realizada, ya que el experimento de consideración no contiene una variable de costo que se pueda relacionar.

- Rutas Mostradas

Contiene a los grupos presentados en el análisis anterior, es decir a los que se les presentaron 10 o 15 variables en el experimento de consideración. Ya que el test-t realizado indicó que la diferencia en el tamaño era significativa para estos grupos, entonces es de esperar que se corroboren dichos resultados.

Tabla 5.1: Resumen de Anova para explicar las diferencias en el tamaño del conjunto de consideración declarado, valor p de cada variable.

Variable Explicada	Prob(>t)										
	Automovilista	FCFM	Género	Edad	Nivel Educativo	Ocupación	Tiempo en Santiago	Licencia	Personas en el Hogar	Ingreso	Rutas Mostradas
Tamaño del conjunto de consideración	<2e-5 (***)	0.282	0.619	0.277	0.168	0.142	0.144	0.383	0.0707 (.)	0.294	<2e-7 (***)

Códigos de significancia: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Como se puede ver, las variables “Automovilista” y “Rutas mostradas” son las que presentan diferencias significativas con un 95 % de confianza, tal como se mencionaba anteriormente. El resto de las variables no logran mostrar un efecto significativo entre los grupos que contienen. El hecho de que la variable “Automovilista” cause una diferencia significativa entre sus grupos puede deberse a que un automovilista comprende mejor el experimento, entiende las rutas y se pone en una posición más realista, por lo que la cantidad de alternativas consideradas se ve afectada por esto. En el caso de las rutas mostradas, simplemente se comprueba lo que se presentó anteriormente con el t-test.

5.2.2. Utilizando encuesta de preferencias declaradas sobre mapa

En esta etapa se estudia el tamaño del conjunto con los datos obtenidos con la encuesta sobre mapa presentada en la sección 4.4. La hipótesis para este análisis es que el número de rutas consideradas por las personas debería ser menor que en el caso de los resultados obtenidos con los datos de la encuesta online. Esta diferencia debería existir debido a que la encuesta fue construida con un diseño no asistido, es decir, no se presenta ninguna ruta a priori y las personas deben construirlas en sus mentes sin ayuda. En este caso, como se mostró en la sección 4.4, la cantidad de encuestados no es suficiente para hacer una afirmación con sustento estadístico. Sin embargo, es posible explorar la diferencia que podría causar el diseño no asistido sobre el tamaño del conjunto, particularmente en comparación con análisis recién mostrado de preferencias declaradas online.

De los datos recopilados, se obtiene que el tamaño promedio del conjunto con 5 observaciones es de 2.4, lo que demuestra ser sustancialmente menor a los dos casos anteriormente mostrados, cuando se presentaban 10 o 15 alternativas en la encuesta online. Esto da indicios sobre la dificultad de generar una alternativa por parte de las personas para poder considerarlas, lo cual en el diseño asistido no es necesario, ya que las alternativas son generadas con antelación.

Como se viene mencionando en esta sección, la principal diferencia es causada por el diseño de los experimentos, donde en el caso de la encuesta online se presentan las alternativas ya creadas, facilitando el procesamiento, mientras que en la encuesta sobre mapa se deben crear a partir de la lectura de un mapa, con menos variables (como se menciona en la sección 4.4) y con muchas más posibilidades de rutas.

5.3. Heurísticas declaradas

En esta sección se analizan los resultados de las preguntas de la encuesta de preferencias declaradas online, donde se les pidió a las personas que respondieran que tan representados se sienten por 9 heurísticas que pueden ser utilizadas para generar el conjunto de consideración. Dichas heurísticas son representaciones generadas a partir de funciones de costo generalizado que se utilizan para calcular rutas mínimas en el artículo de Bekhor et al. (2006). La Tabla 5.2 muestra las heurísticas preguntadas junto con una notación que será utilizada en los resultados presentados a continuación.

De manera de resumir los resultados, se creó una variable binaria de representatividad, donde las personas que marcaron las opciones “me representa muy bien” y “me representa perfectamente” según el ejercicio presentado en la encuesta online, cuentan como representados. Luego, con esta variable se generó un gráfico ordenado por representatividad que se presenta como la Figura 5.1. En esta tabla destacan las heurísticas H1 y H8, superando el 50 % de representatividad, ambas heurísticas están relacionadas con el tiempo y distancia de las rutas, lo que da luces sobre lo que la gente podría observar principalmente para considerar o no una alternativa. Por otro lado, se destaca la heurística H4 con menos de un 20 % de representatividad, esta heurística está relacionada con los virajes a la izquierda (los

cuales en Chile pueden ser complejos en una situación de viraje con oposición), esto puede ser un indicador de que la gente no se logra poner en un escenario realista al momento de realizar el experimento o de que simplemente es algo que pasa a segundo plano en el proceso de consideración. Una de las cosas curiosas que se pueden notar en estos resultados, es que la heurística H4, que tiene relación con los virajes a la izquierda, y la heurística H9, que tiene que ver con utilizar rutas con un menor número de calles distintas presentan niveles de representatividad alejados, pero si se piensa detenidamente, los virajes a la izquierda son una causa de utilizar calles distintas. Lo anterior se puede explicar por el diseño del experimento, al mostrar la ruta dibujada sobre un mapa, es más notoria la trayectoria de la ruta que los virajes en específico. Se debe recordar que las preguntas sobre las heurísticas se realizaron después de realizar los experimentos de consideración y de elección, lo cual, por un lado, debiera favorecer la calidad del reporte al preguntarse por algo que se realizó recientemente, pero por otro lado podía sesgarlo.

En general los resultados que se obtengan aquí son para dar una idea de lo que la gente cree que hace, pero no necesariamente representan lo que realmente están haciendo. En la sección 5.4 se muestra un contraste a lo presentado aquí, presentando un modelo con las variables significativas en un modelo logit de consideración de rutas.

Al igual que para explicar el tamaño del conjunto (ver sección 5.2), se realizó un análisis de varianza ANOVA, utilizando las mismas variables para saber si existe algún grupo que causa una diferencia en el nivel de representatividad de alguna heurística. La diferencia es que en este caso es difícil plantear hipótesis sobre que sucede en los distintos grupos socio-económicos, a priori lo único que se puede esperar es que la variable de “Rutas Mostradas” no debería causar ninguna diferencia.

En la Tabla 5.3 se presentan los resultados de este análisis, donde para algunas heurísticas destacan las variables “Automovilista”, “FCFM” y “Tiempo en Santiago”. La variable “Automovilista” muestra diferencia en las heurísticas H4, H5 y H7, las cuales indican los virajes a la izquierda, las calles seguras y un mayor uso de autopistas, lo que es atribuible a un escenario mas realista de parte del grupo de los automovilistas y un mayor gusto por las autopistas dadas por la experiencia. La variable “FCFM” muestra diferencias significativas en las heurísticas H1, H2 y H8, las cuales tienen que ver con los tiempos y la distancia presentados, lo cual se puede atribuir a un pensamiento más matemático de parte del grupo que trabaja o estudia en la facultad. Finalmente, la variable “Tiempo en Santiago” se presenta como significativa para las heurísticas H4 y H7, las cuales se relacionan con los virajes a la izquierda y el uso de autopistas, lo cual, similar al caso de los automovilistas, se puede atribuir a mayor experiencia, en este caso, con la red de santiago.

Tabla 5.2: Heurísticas mostradas.

Notación	Heurísticas
H1	Considero las alternativas de menor tiempo promedio
H2	Considero las alternativas de menor tiempo en un escenario optimista
H3	Considero las alternativas de menor tiempo y distancia en promedio
H4	Considero las alternativas con menos virajes a la izquierda
H5	Considero las alternativas con más tramos en calles seguras
H6	Considero las alternativas que utilizan vías de alta capacidad (Ej. Alameda)
H7	Considero las alternativas que utilizan más autopistas
H8	Considero las alternativas de menor distancia
H9	Considero las alternativas con menor número de calles distintas

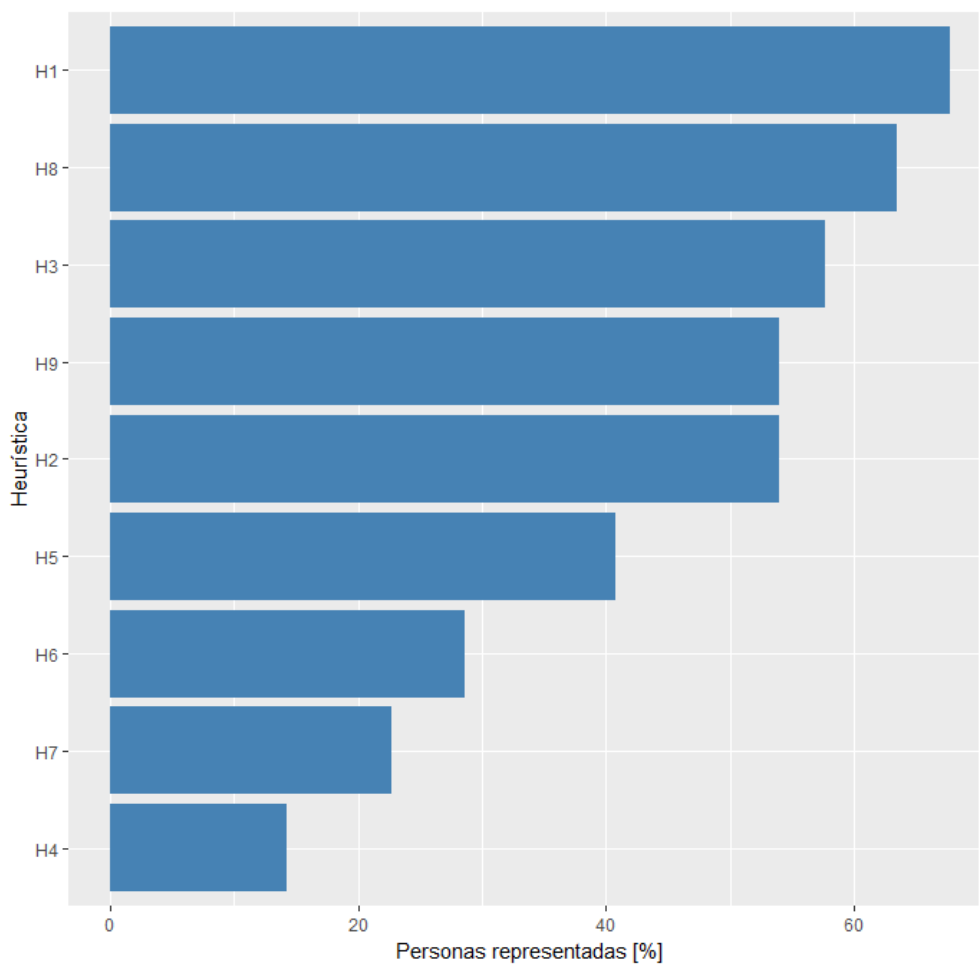


Figura 5.1: Porcentaje de personas representadas por cada heurística presentada.

Tabla 5.3: Resumen de Anova para explicar la representatividad de las heurísticas por grupo, valor p de cada variable.

Variable Explicada	Prob(>t)										
	Automovilista	FCFM	Género	Edad	Nivel Educativo	Ocupación	Tiempo en Santiago	Licencia	Personas en el Hogar	Ingreso	Rutas Mostradas
H1	0.904	0.0400 (*)	0.296	0.454	0.928	0.927	0.171	0.358	0.235	0.978	0.418
H2	0.397	0.0217 (*)	0.325	0.114	0.768	0.460	0.684	0.698	0.903	0.971	0.934
H3	0.194	0.198	0.803	0.937	0.899	0.683	0.445	0.177	0.216	0.993	0.313
H4	0.0444 (*)	0.858	0.121	0.304	0.117	0.450	0.0230 (*)	0.274	0.310	0.135	0.112
H5	0.0448 (*)	0.594	0.250	0.796	0.0993 (.)	0.902	0.967	0.959	0.548	0.715	0.742
H6	0.358	0.895	0.215	0.752	0.565	0.915	0.107	0.794	0.94	0.991	0.786
H7	0.0785 (.)	0.614	0.784	0.475	0.825	0.822	0.00132 (**)	0.363	0.382	0.304	0.369
H8	0.632	0.0874 (.)	0.111	0.364	0.404	0.325	0.568	0.364	0.921	0.768	0.535
H9	0.119	0.742	0.703	0.458	0.177	0.208	0.574	0.532	0.813	0.285	0.910

Códigos de significancia: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

5.4. Modelo binario de consideración

Con la finalidad de identificar las variables que afectan al proceso de consideración, se desarrolla un modelo logit binario de consideración (ver ecuación 3.3), donde se utiliza cada pregunta del ejercicio de consideración de la encuesta online presentada en la sección 4.3, sobre considerar o no una alternativa, como una observación válida para el modelo. La modelación consistió, en una primera etapa, en utilizar todas las variables disponibles que pudieran tener sentido en el modelo, tanto de las rutas como variables socio-económicas recolectadas. Este modelo preliminar se muestra en la Tabla 5.4. La intuición para realizar un modelo de este estilo es que la heurística que está detrás del proceso de consideración puede ser representado por un modelo logit binario, donde un parámetro negativo representará que una variable de una alternativa al aumentar su valor causa que la probabilidad de considerar dicha alternativa disminuya. Finalmente, una explicación de las variables y sus parámetros es presentada a continuación.

- Constante

La “Constante” es necesaria cuando se realiza un logit binario, en el que la utilidad de una alternativa es comparada con una utilidad de base que es fijada en un valor arbitrario, que en este caso es cero. Su signo depende de los demás atributos, pero debería ser estadísticamente significativa.

- Rojo

La variable “Rojo”, representa el porcentaje de color rojo que tiene una ruta (presentado con la escala de congestión en la encuesta online), es decir, es un representante de la congestión de la ruta, en este caso significa que tanta más utilidad reporta una unidad más de porcentaje de rojo en comparación a los otros colores presentes (que representan menor congestión). Es de esperar que el parámetro asociado sea significativo y tenga signo negativo ya que si una ruta tiene un mayor porcentaje de rojo entonces será considerada con menor probabilidad.

- Virajes Derecha

La variable “Virajes Derecha” representa el número de virajes a la derecha que tiene una ruta. A priori, el parámetro asociado a esta variable debería ser significativo y tener signo negativo, ya que el ideal de una ruta sería ir en línea recta, por lo tanto cualquier viraje extra es una disminución a la utilidad.

- Virajes Izquierda

La variable “Virajes izquierda” indica la cantidad de virajes a la izquierda que se deben realizar en una ruta. Al igual que los virajes a la derecha, el parámetro asociado debería ser significativo y tener signo negativo. Incluso se puede llegar a pensar que la magnitud de este parámetro debería ser mayor a la de los virajes a la derecha por la complejidad que representan

- Tiempo Mínimo

De los tiempos presentados en la encuesta online, aquel que se comporta mejor en los modelos es el “Tiempo Mínimo”. El parámetro asociado a esta variable debería ser significativo y tener signo negativo, ya que una ruta tendrá menor probabilidad de ser considerada si su tiempo de viaje mínimo aumenta.

- Orden >10

La variable mencionada como “Orden >10 ”, es una variable de corrección al modelo por el sesgo lexicográfico causado por mostrar las alternativas una tras otra. Se espera que el parámetro asociado a esta variable sea significativo y su signo debe ser negativo para indicar que aquellas rutas que son presentadas después de la décima ruta tengan menor probabilidad de ser consideradas. Lo anterior se puede adjudicar a las limitaciones cognitivas o cansancio de las personas encuestadas.

- FCFM

La variable “FCFM” indica si la persona trabaja o estudia en la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Esta variable podría ser de interés dada la forma en la que está construido el experimento, donde el origen del viaje es justamente la facultad, resultando tener un parámetro estadísticamente significativo. Sin embargo, no se pueden realizar hipótesis sobre el signo de dicho parámetro.

- Mujer

La variable “Mujer”, indica si la persona es mujer. En un contexto de rutas, el género no debería ser significativo.

- Edad

En las variables de edad, se omite el rango más alto de edad. A priori estas variables no deberían ser significativas, pero podrían serlo si se presentan como una corrección al experimento por ser online y resultar más fácil para cierto rango etario.

- Educación

En las variables de educación, “E.M. Científico-Humanista” representa a las personas que su último nivel educacional fue en un colegio científico-humanista, “E.M. Técnico profesional” es análogo e “Instituto Profesional” es auto-explicativa, la variable omitida es la del grupo “Universitario”. A priori no hay ningún sustento para indicar que el parámetro asociado a las variables deba ser significativo.

- Ocupación

Las variables de ocupación, presentan la categoría de “Ocupación Estudiante” y “Ocupación Trabajador”, omitiendo la “Ocupación Otros”. A priori no se pueden realizar hipótesis sobre la significancia de los parámetros asociados, a no ser que alguna de las categorías sea más exigente al considerar las alternativas.

- Más de 1 Año en Santiago

La variable “Más de 1 Año en Santiago”, que representa justamente lo que dice, si una persona ha vivido al menos un año en la ciudad de Santiago. Se esperaría que esta variable pudiera hacerse cargo del conocimiento que tiene una persona sobre la red de transporte u por lo tanto tener un parámetro asociado significativo, sin realizar hipótesis sobre su signo.

- Chofer y Acompañante

Las variables “Chofer” y “Acompañante”, representan que una persona maneja un vehículo motorizado para ir a su lugar de trabajo o estudio, o bien si es acompañante. Otros modos de transporte, como los “No Motorizados” son omitidos. A priori se esperaría que hubiese un efecto significativo, ya que las personas que manejan podrían tomar el experimento de una manera diferente. Sin embargo no se asume el signo que podría tener el parámetro asociado a cada una.

- Automovilista

Finalmente, la variable “Automovilista” representa si una persona utiliza algún vehículo motorizado para ir al trabajo o estudio por lo menos 1 vez a la semana. Esta variable, similar a la anterior se espera que muestre alguna diferencia en el realismo con el que las personas se tomar el experimento, por lo que el efecto debería ser significativo. Sin embargo no se pueden realizar hipótesis con sustento para el signo de su parámetro asociado.

Tabla 5.4: Modelo de consideración con todas las variables.

	Logit Binario de Consideración										
	Parámetro	Error Estándar	t-valor	Prob(>t)							
Constante	4.80	6.943	0.692	0.489							
Rojo	-2.81	1.167	-2.41	0.0159	*						
Virajes Derecha	-0.231	0.0638	-3.62	0.000294	***						
Virajes Izquierda	-0.217	0.0706	-3.07	0.00214	**						
Tiempo Mínimo	-0.245	0.0188	-13.0	<2e-16	***						
Orden >10	-0.208	0.119	-1.74	0.0816	.						
FCFM	0.0952	0.0964	0.988	0.323							
Mujer	-0.121	0.0872	-1.39	0.164							
Edad entre 18 y 24	0.914	6.96	0.131	0.896							
Edad entre 25 y 34	0.949	6.96	0.136	0.892							
Edad entre 35 y 44	0.797	6.97	0.114	0.909							
Edad entre 45 y 54	-0.0000589	6.96	0.00	1.00							
E.M. Científico-Humanista	-0.0728	0.118	-0.616	0.538							
E.M. Técnico-Profesional	1.03	0.374	2.76	0.00584	**						
Instituto Profesional	-0.161	6.93	-0.0230	0.982							
Ocupación Estudiante	-0.187	0.160	-1.17	0.241							
Ocupación Trabajador	-0.565	0.161	-3.50	0.000463	***						
Más de 1 Año en Santiago	-0.136	0.198	-0.684	0.494							
Chofer	0.0428	0.266	0.160	0.873							
Acompañante	0.0451	0.142	0.317	0.751							
Automovilista	-0.356	0.0926	-3.85	0.000120	***						
Observaciones	2370										
ρ^2 ajustado	0.236										
Log-verosimilitud	-1234.575										
Códigos de significancia:	0	***	0.001	**	0.01	*	0.05	.	0.1	'	1

Como se menciona en la hipótesis, la constante es necesaria para un modelo logit binario donde la utilidad de una alternativa es comparada con una utilidad base desconocida (que por temas prácticos se fija en 0), por lo que a pesar de no ser significativa en este modelo se seguirá utilizando en futuras especificaciones. Por otro lado, de los parámetros significativos presentados en la tabla anterior, destaca aquel asociado a la variable de “Instituto Profesional”, sin embargo, no existe ningún trasfondo que pudiera explicar esto para este grupo en específico. Al revisar en detalle, se puede notar que los encuestados de dicha categoría son muy pocos y el efecto se puede deber a una simple coincidencia, de manera que ningún nivel educacional es incluido en el modelo definitivo. El resto de las variables significativas cumplen con las hipótesis planteadas. Por un lado, la ocupación se puede explicar por la percepción del experimento de parte de este grupo o por la falta del tiempo para responder la encuesta que causa esta ocupación. Por otro lado la variable “Automovilista” da cuenta del efecto de realismo sobre el experimento o de experiencia en elección de rutas, donde al tener un parámetro con signo negativo se presenta como una mayor exigencia para considerar alternativas. Además, es esta última variable la que opaca la significancia esperada de parte de los parámetros de las variables “Chofer” y “Acompañante”. De esta manera, luego de probar

distintas especificaciones para un modelo definitivo, se presenta en la Tabla 5.5 el modelo de consideración definitivo, donde todas las variables son significativas.

Tabla 5.5: Modelo de consideración definitivo.

Logit Binario de Consideración					
	Parámetro	Error Estándar	t-valor	Prob(>t)	
Constante	5.36	0.249	21.5	<2e-16	***
Rojo	-2.70	1.12	-2.40	0.0164	*
Virajes Derecha	-0.226	0.0616	-3.67	0.000243	***
Virajes Izquierda	-0.216	0.0668	-3.23	0.00123	**
Tiempo Mínimo	-0.243	0.0181	-13.5	<2e-16	***
Orden >10	-0.210	0.110	-1.91	0.0565	.
Ocupación Trabajador	-0.399	0.108	-3.68	0.000235	***
Automovilista	-0.459	0.0816	-5.63	<2e-8	***
Observaciones	2370				
ρ^2 ajustado	0.237				
Log-verosimilitud	-1244.843				
Códigos de significancia:	0	0.001	0.01	0.05	0.1
	‘***’	‘**’	‘*’	‘.’	‘.’

Este modelo indica que las variables asociadas a las rutas son de gran importancia y se encuentran presentes en el modelo, cumpliendo la hipótesis presentada anteriormente. Además de estas variables, una corrección a causa del diseño experimento es necesaria, representada por la variable “Orden >10”. Por otro lado, se muestra que existen dos grupos de personas (trabajadores y automovilistas) que por motivos como su percepción del experimento o el tiempo disponible para responder, son más exigentes al momento de considerar una ruta y esto se ve representado en las últimas dos variables del modelo con signo negativo. En la Tabla 5.6 se presenta el porcentaje de la utilidad sistemática que presenta cada una de las variables cuando se evalúa en el promedio de cada atributo (sin considerar la constante), donde se puede apreciar que las variables socio-económicas son las de mayor aporte, mientras que la variable asociada a la congestión de las rutas es aquella que presenta menor impacto. Además, se presenta la elasticidad agregada de las variables, lo que representa cuál es el cambio en la probabilidad de considerar una ruta cuando se aumenta en una unidad la variable que corresponde. En este caso se puede apreciar que la variable “Tiempo Mínimo” es la más sensible de todas, seguido por las dos variables de virajes.

Este modelo se presenta como una buena manera de entender las variables que afectan en el proceso de generación del conjunto de consideración, además logra dar luces de la heurística que las personas utilizan para genera el conjunto de consideración. Se puede notar que el modelo final, presentado en la Tabla 5.5, en comparación con la representatividad de las heurísticas mostradas en la Tabla 5.1, muestra que las variables de virajes son estadísticamente significativas, mientras que en el análisis anterior se presentaba lo contrario, esto se puede deber a que los virajes son representados por otra heurística como la de considerar un menor número de calles distintas, por otro lado, variables como el tiempo mínimo se ratifican con este análisis.

Tabla 5.6: Porcentaje de la utilidad sistemática que aporta cada variable sin contar la constante específica utilizada para dar cuenta de la utilidad base con la que se compara en un logit binario.

Variable	Porcentaje	Elasticidad Agregada
Automovilista	25.33 %	-0.0308
Ocupación Trabajador	22.00 %	-0.0268
Tiempo Mínimo	13.43 %	-1.41
Virajes Derecha	12.47 %	-0.281
Virajes Izquierda	11.91 %	-0.211
Orden >10	11.60 %	-0.0229
Rojo	3.26 %	-0.0731

5.5. Modelo de elección

Una de las preguntas que surge al realizar el modelo de consideración recién presentado, es si los parámetros serán iguales a los de un modelo de elección. La hipótesis presentada en esta sección es que los parámetros no deberían ser iguales, ya que el modelo de consideración es una representación de una heurística y por lo tanto los parámetros no son los que las personas utilizan en su mente, sino que solo ayudan a modelar el fenómeno, en cambio los parámetros de un modelo de elección representan el peso que una persona le da a una variable para calcular la utilidad de dicha alternativa.

Para probar lo anterior, se realiza un modelo logit multinomial que utiliza el ranking de alternativas como elecciones. Es decir, que si una persona hizo un ranking de 3 alternativas, entonces se obtienen dos observaciones a partir de ello, donde la alternativa del ranking mayor se va eliminando y da paso a una nueva observación con la alternativa de ranking siguiente como la elección. Por otro lado, se utiliza como conjunto de consideración aquellas alternativas que declararon en la sección de consideración de la encuesta online. Además, ya que algunos parámetros de algunas variables no se podrían estimar debido a que no cambian entre los bloques generados (por ejemplo, virajes o colores de congestión), se utiliza un orden aleatorio en las alternativas, es decir, para la persona 1, la alternativa 10 podría ser la primera en el modelo, sin embargo, para la persona 2, la misma alternativa podría ser la quinta en el modelo, todo esto es para poder generar varianza en los atributos, esto causa que las constante específicas de alternativa no se incluyan debido a que no tendrían sentido al representar a alternativas diferentes.

Como primera instancia, se intentaron utilizar las mismas variables que el modelo de consideración, pero la variable sobre el orden no tienen sentido en este modelo (dado que la variable hacia referencia a la posición en el experimento de consideración para una alternativa) y las variables socio-económicas presentan un problema de especificación, donde no se pueden presentar las mismas variables con el mismo valor en todas las alternativas, ya que esto no permite estimar los parámetros. Para solucionar esto se probaron diversas especificaciones, donde se utilizaron dichas variables para separar grupos de personas que tuvieran parámetros diferentes, sin embargo, ninguna especificación de este tipo fue significativa. Las hipótesis y una breve explicación sobre las variables utilizadas en el modelo se presentan a continuación:

- Rojo

Es el porcentaje de color rojo que contiene una ruta (según la escala de congestión). El parámetro asociado a esta variable debería tener signo negativo y ser significativo ya que una mayor cantidad de rojo implica mayor congestión en la ruta, por lo que la probabilidad de elegir dicha ruta debería disminuir.

- Virajes Derecha

Representa la cantidad de virajes a la derecha que contiene una ruta. A priori esta variable debería ser significativa y tener signo negativo, ya que el ideal de una ruta sería ir en línea recta, por lo tanto cualquier viraje extra es una disminución a la utilidad.

- Virajes Izquierda

Representa el número de virajes a la izquierda que una ruta posee. Al igual que en los virajes a la derecha, el parámetro asociado debería ser significativo y con signo negativo y posiblemente de mayor magnitud en comparación con los virajes a la derecha debido a la complejidad que estos virajes representan en Chile, especialmente en situaciones con oposición de vehículos.

- Tiempo Mínimo

Esta variable representa al tiempo mínimo que alguien se puede demorar en recorrer una ruta. El parámetro asociado a esta variable debería ser significativo y tener signo negativo, al igual que para el modelo presentado en la Tabla 5.5, ya que un mayor tiempo en ruta manteniendo todo lo demás constante causa una disminución en la utilidad que esta brinda.

Los resultados de la estimación del mejor modelo de elección se presentan en la Tabla 5.7, donde se mantienen las variables “Rojo” y “Virajes Derecha” asociadas a las rutas, a pesar de no ser significativas, para facilitar la comparación con el modelo de consideración. Además, se debe observar que el ajuste de este modelo es bastante bajo, lo cual es atribuible a la cantidad de observaciones y la falta de una mejor especificación, o inclusive de un modelo más complejo como logit anidado o probit. Por otro lado, las cuatro variables presentan signo negativo, lo que apoya las hipótesis presentadas anteriormente. Más adelante, en la Tabla 5.10 se presenta un modelo de elección, el cual es una versión pulida del aquí presentado, utilizando solo las variables cuyos parámetros son significativos en este modelo. Además, se presenta como un modelo que no utiliza el ranking de alternativas para generar observaciones, sino que utiliza la alternativa de ranking más alto como la elección, obteniendo de esta manera sólo una observación por persona. En dicho modelo, las variables son significativas y con signo positivo como se indica en las hipótesis. Por otro lado, se puede observar que se presenta un mayor ρ^2 ajustado, pero no se debe comparar debido a que el número de observaciones es diferente.

Tabla 5.7: Modelo de elección utilizando el ranking de alternativas.

Logit Multinomial de Elección Utilizando Ranking											
	Parámetro	Error Estándar	t-valor	Prob(>t)							
Rojo	-0.166	0.730	-0.228	0.820							
Virajes Derecha	-0.0290	0.0659	-0.440	0.660							
Virajes Izquierda	-0.201	0.0745	-2.70	0.00692	**						
Tiempo Mínimo	-0.212	0.0156	-13.6	<2e-16	***						
Observaciones	895										
ρ^2 ajustado	0.097										
Log-verosimilitud	-1106.265										
Códigos de significancia:	0	***	0.001	**	0.01	*	0.05	.	0.1	'	1

El siguiente paso para cumplir el objetivo planteado es modelar la consideración y elección dentro de una misma función de maximización de verosimilitud de manera de ser capaz de realizar test formales para comparar los parámetros. Para poder hacer este tipo de modelo, en el caso de querer restringir los parámetros para probar si son iguales o distintos, se debe explicitar una de las escalas de uno de los modelos, como se sabe, al estimar un logit se fija la escala en 1 y se estiman los parámetros, pero al estimar dos logit en simultaneo una de estas escalas es estimable. En los modelos estimados esta escala se ve presentada como $\mu_{elección}$.

Para poder realizar un test formal, en particular el de razón de verosimilitud, se realizan dos modelos. El primer modelo donde todos los parámetros son distintos (modelo sin restringir), dicho modelo se muestra en la Tabla 5.8. Y el segundo modelo donde los 4 parámetros comunes son iguales (modelo restringido), cuyo resultado se muestra en la Tabla 5.9.

El test de razón de verosimilitud se utiliza para probar la hipótesis nula de que los parámetros son iguales, de manera que se calcula el estadístico del test y se compara con el valor crítico como se muestra en la ecuación 5.3. Dicho test indica que se rechaza la hipótesis nula y por lo tanto los parámetros son distintos.

$$\begin{aligned}
 & -2(-L(\text{restringido}) - L(\text{sinrestringir})) = \\
 & -2(-2365,812 - -2351,108) = 11,208 > \chi^2(4, 95\%) = 9,488 \quad (5.3)
 \end{aligned}$$

Este resultado tiene sentido si la hipótesis de que la gente utiliza una heurística para generar su conjunto de consideración es cierta. Además, no tendría sentido que fuesen los mismos si las personas no tienen la capacidad cognitiva para calcular la utilidad de todas las alternativas existentes, de hecho, es este supuesto el que da vida al concepto del conjunto de consideración. Como se menciona en Swait y Ben-Akiva (1987) el proceso de consideración se basa en restricciones, mientras que en el proceso de elección se calcula la utilidad de cada alternativa, lo que respalda el resultado de que los parámetros son diferentes. Este resultado de ninguna manera invalida el modelo de consideración, pero indica que dicho modelo es una representación de algo más y no lo que las personas realmente hacen.

Tabla 5.8: Modelo de combinado de consideración y elección sin restringir los parámetros comunes.

	Logit Combinado de Consideración y Elección										
	Parámetro	Error Estándar	t-valor	Prob(>t)							
Consideración											
Constante	5.36	0.257	20.9	<2e-16	***						
Rojo	-2.70	1.13	-2.39	0.0167	*						
Virajes Derecha	-0.226	0.0626	-3.61	0.000305	***						
Virajes Izquierda	-0.216	0.0679	-3.18	0.00147	**						
Tiempo Mínimo	-0.243	0.0184	-13.2	<2e-16	***						
Orden >10	-0.210	0.112	-1.87	0.0614	.						
Ocupación Trabajador	-0.399	0.110	-3.62	0.000297	***						
Automovilista	-0.459	0.0854	-5.38	<2e-7	***						
Elección											
Rojo	-0.167	0.753	-0.221	0.825							
Virajes Derecha	-0.0290	0.0681	-0.426	0.670							
Virajes Izquierda	-0.201	0.0780	-2.58	0.00987	**						
Tiempo Mínimo	-0.212	0.0161	-13.2	<2e-16	***						
$\mu_{elección}$	1.00										
Observaciones	3265										
ρ^2 ajustado	0.439										
Log-verosimilitud	-2351.108										
Códigos de significancia:	0	***)	0.001	**)	0.01	*)	0.05	.)	0.1	.)	1

Tabla 5.9: Modelo de combinado de consideración y elección restringiendo los parámetros comunes.

	Logit Combinado de Consideración y Elección										
	Parámetro	Error Estándar	t-valor	Prob(>t)							
Consideración											
Constante	5.52	0.249	22.2	<2e-16	***						
Rojo	-1.45	0.774	-1.87	0.0611	.						
Virajes Derecha	-0.148	0.0559	-2.65	0.00802	**						
Virajes Izquierda	-0.261	0.0617	-4.23	<2e-4	***						
Tiempo Mínimo	-0.269	0.0160	-16.8	<2e-16	***						
Orden >10	-0.207	0.112	-1.85	0.0640	.						
Ocupación Trabajador	-0.396	0.110	-3.61	0.000310	***						
Automovilista	-0.453	0.0828	-5.47	<2e-7	***						
Elección											
Rojo	-1.45	0.774	-1.87	0.0611	.						
Virajes Derecha	-0.148	0.0559	-2.65	0.00802	**						
Virajes Izquierda	-0.261	0.0617	-4.23	<2e-4	***						
Tiempo Mínimo	-0.269	0.0160	-16.8	<2e-16	***						
$\mu_{elección}$	0.684	0.0424	16.1	<2e-16	***						
Observaciones	3265										
ρ^2 ajustado	0.440										
Log-verosimilitud	-2356.812										
Códigos de significancia:	0	***	0.001	**	0.01	*	0.05	.	0.1	'	1

5.6. Análisis comparativo de métodos generadores del conjunto de consideración

El último aporte que se realiza en este trabajo, consiste en la evaluación de distintos métodos para generar el conjunto de consideración supuesto. Este estudio es análogo al realizado en la sección de la simulación de Monte Carlo presentada en el capítulo 3, con la diferencia de que se utilizan los datos de preferencias declaradas de la encuesta online y no es posible evaluar todos los métodos por el tipo de datos disponibles.

5.6.1. Modelo utilizando el conjunto de consideración verdadero

El modelo presentado aquí se utilizará de base para comparar con los modelos que estiman los parámetros mediante conjuntos de consideración supuestos generados de diversas maneras. Se puede notar de la Tabla 5.10 que la especificación utilizada para estimar este modelo utiliza sólo las variables significativas del modelo presentado en la Tabla 5.7, el cual utiliza más observaciones y se considera más confiable.

En la Tabla 5.10 se presenta un modelo de elección que estima los parámetros utilizando la primera alternativa del ranking generado por cada persona, esto se debe a que el ranking no permite simular conjuntos de consideración para cada observación generada ya que al ir eliminando la elección de una observación para generar la siguiente no se puede asumir que la elección siguiente pertenece al conjunto supuesto, puesto que al hacerlo se estaría incluyendo siempre el conjunto original. Además, se destaca que se utiliza el conjunto de consideración verdadero (CV), que corresponde al declarado por cada individuo para realizar la estimación del modelo. Finalmente, las hipótesis presentadas en la sección 5.5 son las mismas que se utilizan en este caso para validar los parámetros.

Tabla 5.10: Modelo de elección sin utilizar el ranking de alternativas, conjunto de consideración utilizado corresponde a CV.

Logit Multinomial de Elección sin Utilizar Ranking												
	Parámetro	Error Estándar	t-valor	Prob(>t)								
Virajes Izquierda	-0.431	0.108	-4.00	<2e-4	***							
Tiempo Mínimo	-0.298	0.0484	-6.20	<2e-9	***							
Observaciones	189											
ρ^2 ajustado	0.231											
Log-verosimilitud	-239.388											
Códigos de significancia:	0	***	0.001	**	0.01	*	0.05	.	0.1	'	'	1

Del modelo presentado se puede apreciar que ambas variables son significativas y podrán ser utilizadas para comparar la proporción de parámetros con el resto de modelos presentados en esta sección. El parámetro de “Virajes Izquierda” y “Tiempo Mínimo” presentan el signo y nivel de significancia planteado en las hipótesis.

5.6.2. Modelos utilizando otros conjuntos de consideración

Los métodos utilizados para generar el conjunto de consideración supuesto son los mismos que en el capítulo 3 sobre la simulación de Monte Carlo, excepto por el método de las alternativas experimentadas (CE), debido a que no es posible utilizarlo con el tipo de datos recolectados en la encuesta online. A continuación se explican los métodos generadores del conjunto de consideración supuesto.

1. Conjunto de consideración de las k-ésimas rutas mínimas, utilizando los parámetros verdaderos (KRM_V)

En este método se simula una de las reglas deterministas más conocidas, correspondiente a rutas mínimas, tratando de emular distintos métodos señalados en artículos como Prato (2009) y Ramming (2001). Este método consiste en calcular la componente sistemática de la utilidad, con la misma forma funcional que el modelo logit multinomial de elección de la Tabla 5.10, utilizando parámetros a priori estimados por este modelo. Luego, con la utilidad sistemática calculada se seleccionan las de 3 mayor utilidad

para formar el conjunto de consideración supuesto, además si la elección original no se encuentra en el conjunto de consideración supuesto generado, entonces se agrega.

2. Conjunto de consideración de las k-ésimas rutas mínimas, utilizando parámetros erróneos (KRM_E)

Este método es equivalente al anterior, pero se utilizan parámetros a priori arbitrariamente distintos a los verdaderos del modelo de elección presentados en la Tabla 5.11. De manera que se calcula la utilidad sistemática utilizando la especificación del modelo de elección de la Tabla 5.10, luego se hace un ranking de las alternativas y se eligen las 3 mejores para formar el conjunto de consideración, además si la alternativa escogida no es una de las 3 mejores, se agrega al conjunto de consideración supuesto.

Tabla 5.11: Parámetros a priori para método KRM_E.

Parámetro	Valor
<i>th_tiempo</i>	-0.1492
<i>th_costo</i>	-0.2984

3. Conjunto de consideración completo (CC)

Este método es el más simple de modelar, ya que incluye a todas las alternativas en el conjunto de consideración supuesto. En este caso se distinguen dos situaciones. Primero, cuando el conjunto completo se genera a partir de las alternativas mostradas, es decir, de las 10 o 15 alternativas que se presentan en el experimento de consideración de la encuesta online. Segundo, cuando se genera el conjunto a partir de las 20 alternativas conocidas por el modelador.

4. Conjunto de consideración con logit binario (CLB)

Por último, este método es el único que pertenece a la familia de los métodos estocásticos. En este caso se estima un modelo logit binario, donde conociendo el conjunto de consideración de las personas (declarado en la encuesta online), se utiliza cada alternativa como una observación, donde la elección es haberla considerado o no. En este caso, se utiliza el modelo generado en la sección 5.4 de la Tabla 5.5 para simular el conjunto a partir de las alternativas mostradas (10 o 15) o de las 20 alternativas conocidas por el modelador. La simulación es realizada a partir del cálculo de la probabilidad de cada alternativa de ser considerada y de la generación de números aleatorios entre 0 y 1, de manera que una alternativa es considerada si su probabilidad es mayor que el número aleatorio generado.

La diferencia en este análisis es que existen dos maneras en cómo se pueden utilizar estos métodos. La primera, consiste en utilizar las 10 o 15 alternativas presentadas como conjunto universal de alternativas. La segunda, utiliza las 20 alternativas que el modelador conoce que existen como universo de alternativas. Si bien la primera tiene más similitudes con el análisis hecho en el capítulo sobre la simulación de Monte Carlo, la segunda se acerca más al caso realista que tienen que enfrentar los analistas, debido a que conoce puede conocer más alternativas que las que los individuos pueden llegar a conocer.

Conjunto de consideración supuesto generado a partir de las alternativas mostradas

Para esta etapa, se utilizan los métodos mencionados anteriormente para generar el conjunto de consideración supuesto a partir de las alternativas mostradas a las personas y luego estimar modelos de elección que permitan comparar la proporción de los parámetros con el modelo mostrado en la Tabla 5.10 que utiliza el conjunto de consideración verdadero.

En la Tabla 5.12 se presentan los resultados de estimar un modelo de elección utilizando el conjunto de consideración generado por el método CC. Se puede ver que los parámetros son significativos y tienen signo correcto según lo discutido anteriormente.

La Tabla 5.13 muestra los resultados de la estimación del modelo de elección utilizando el método KSP_V para generar el conjunto de consideración supuesto. Se puede ver que a pesar de que los parámetros tienen signo correcto según lo discutido, el de Virajes Izquierda no es estadísticamente significativo.

En la Tabla 5.14, se muestran los resultados de estimar un modelo de elección utilizando el conjunto de consideración generado mediante KSP_E. Como se puede apreciar ambos parámetros son significativos y con signo correcto según lo discutido.

Por último, la Tabla 5.15 muestra los resultados de la estimación del modelo de elección utilizando el conjunto de consideración generado por el método CLB. Nuevamente los parámetros son significativos y tienen signo correcto según lo discutido.

Tabla 5.12: Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante CC a partir de las alternativas mostradas.

Logit Multinomial de Elección sin Utilizar Ranking											
	Parámetro	Error Estándar	t-valor	Prob(>t)							
Virajes Izquierda	-0.534	0.0952	-5.611	<2e-7	***						
Tiempo Mínimo	-0.322	0.0431	-7.472	<2e-13	***						
Observaciones	189										
ρ^2 ajustado	0.338										
Log-verosimilitud	-311.827										
Códigos de significancia:	0	***)	0.001	**)	0.01	*)	0.05	.)	0.1	.)	1

Tabla 5.13: Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante KRM_V a partir de las alternativas mostradas.

Logit Multinomial de Elección sin Utilizar Ranking					
	Parámetro	Error Estándar	t-valor	Prob(>t)	
Virajes Izquierda	-0.272	0.104	-2.62	0.00869	**
Tiempo Mínimo	-0.242	0.0459	-5.28	<2e-7	***
Observaciones	189				
ρ^2 ajustado	0.117				
Log-verosimilitud	-246.624				
Códigos de significancia:	0	0.001	0.01	0.05	0.1

Tabla 5.14: Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante KRM_E a partir de las alternativas mostradas.

Logit Multinomial de Elección sin Utilizar Ranking					
	Parámetro	Error Estándar	t-valor	Prob(>t)	
Virajes Izquierda	-0.362	0.105	-3.44	<2e-3	***
Tiempo Mínimo	-0.207	0.0467	-4.44	<2e-5	***
Observaciones	189				
ρ^2 ajustado	0.122				
Log-verosimilitud	-246.398				
Códigos de significancia:	0	0.001	0.01	0.05	0.1

Tabla 5.15: Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante CLB a partir de las alternativas mostradas.

Logit Multinomial de Elección sin Utilizar Ranking					
	Parámetro	Error Estándar	t-valor	Prob(>t)	
Virajes Izquierda	-0.379	0.0975	-3.89	<2e-4	***
Tiempo Mínimo	-0.245	0.0458	-5.35	<2e-7	***
Observaciones	189				
ρ^2 ajustado	0.190				
Log-verosimilitud	-263.998				
Códigos de significancia:	0	0.001	0.01	0.05	0.1

Conjunto de consideración supuesto generado a partir del universo de alternativas

En esta siguiente etapa, utilizando los métodos mencionados anteriormente se genera el conjunto de consideración supuesto a partir de las 20 alternativas conocidas por el modelador,

para luego estimar modelos de elección que permitan comparar los parámetros con el modelo mostrado que utiliza el conjunto de consideración verdadero.

En la Tabla 5.16 se presentan los resultados de estimar un modelo de elección utilizando el conjunto de consideración generado por el método CC. Se puede ver que los parámetros son significativos y tienen signo correcto según lo discutido anteriormente.

La Tabla 5.17 muestra los resultados de la estimación del modelo de elección utilizando el método KSP_V para generar el conjunto de consideración supuesto. Se puede ver que este método falla en todo sentido, ya que ambos parámetros tiene signo incorrecto, lo que llevaría a una predicción totalmente al revés.

En la Tabla 5.18, se muestran los resultados de estimar un modelo de elección utilizando el conjunto de consideración generado mediante KSP_E. Este es exactamente igual al anterior, debido a que al utilizar todas las alternativas existen algunas que son dominantes y sin importar cuanto se muevan los parámetros a priori utilizados resultan ser las de mejor utilidad sistemática. Se menciona al igual que en el caso anterior que los signos son incorrectos y por lo tanto el modelo es malo ya que llevaría a predicciones al revés.

Finalmente, la Tabla 5.19 muestra los resultados de la estimación del modelo de elección utilizando el conjunto de consideración generado por el método CLB. Nuevamente los parámetros son significativos y tienen sentido correcto.

Tabla 5.16: Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante CC a partir del universo de alternativas.

Logit Multinomial de Elección sin Utilizar Ranking											
	Parámetro	Error Estándar	t-valor	Prob(>t)							
Virajes Izquierda	-0.499	0.0944	-5.29	<2e-7	***						
Tiempo Mínimo	-0.290	0.0431	-6.73	<2e-11	***						
Observaciones	189										
ρ^2 ajustado	0.273										
Log-verosimilitud	-409.779										
Códigos de significancia:	0	***	0.001	**	0.01	*	0.05	.	0.1	'	1

Tabla 5.17: Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante KRM_V a partir del universo de alternativas.

Logit Multinomial de Elección sin Utilizar Ranking					
	Parámetro	Error Estándar	t-valor	Prob(>t)	
Virajes Izquierda	0.154	0.138	1.12	0.264	
Tiempo Mínimo	0.113	0.0562	2.02	0.0439	*
Observaciones	189				
ρ^2 ajustado	0.0146				
Log-verosimilitud	-268.935				
Códigos de significancia:	0	***	0.001	**	0.01 *
				0.05 .	0.1 ' 1

Tabla 5.18: Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante KRM_E a partir del universo de alternativas.

Logit Multinomial de Elección sin Utilizar Ranking					
	Parámetro	Error Estándar	t-valor	Prob(>t)	
Virajes Izquierda	0.154	0.138	1.12	0.264	
Tiempo Mínimo	0.113	0.0562	2.02	0.0439	*
Observaciones	189				
ρ^2 ajustado	0.0146				
Log-verosimilitud	-268.935				
Códigos de significancia:	0	***	0.001	**	0.01 *
				0.05 .	0.1 ' 1

Tabla 5.19: Modelo de elección sin utilizar el ranking de alternativas, estimado utilizando el conjunto de consideración supuesto mediante CLB a partir del universo de alternativas.

Logit Multinomial de Elección sin Utilizar Ranking					
	Parámetro	Error Estándar	t-valor	Prob(>t)	
Virajes Izquierda	-0.338	0.0948	-3.60	<2e-3	***
Tiempo Mínimo	-0.228	0.0432	-5.29	<2e-7	***
Observaciones	189				
ρ^2 ajustado	0.144				
Log-verosimilitud	-359.861				
Códigos de significancia:	0	***	0.001	**	0.01 *
				0.05 .	0.1 ' 1

5.6.3. Comparación de modelos

En esta última etapa, se comparan los modelos recién mostrados de manera gráfica y mediante un test formal para ambos escenarios presentados, es decir, cuando los métodos

generan el conjunto a partir de las alternativas mostradas o a partir del universo de alternativas. Se debe destacar que los métodos KRM_V y KRM_E para el caso generado a partir del universo de alternativa tienen signo incorrecto en ambos parámetros y por la tanto son incorrectos sin necesidad de comparar la proporción de los parámetros.

En la Figura 5.2 y Figura 5.3 se presentan los intervalos de confianza de la proporción de los parámetros de Virajes Izquierda y Tiempo Mínimo. Para calcular los errores estándar que generan rangos presentados, debido que se trata de proporciones de parámetros, se utiliza el método Delta, que aplica una aproximación de Taylor de primer orden a las desviaciones al rededor del punto estimado, este método puede revisarse con más detalle en el artículo de Daly et al. (2012), donde la expresión para la desviación estándar del ratio de parámetros se puede encontrar en la tabla 1 del mismo.

Utilizando las alternativas mostradas

En la Figura 5.2 se presentan los intervalos de todos los métodos utilizados, partiendo por el conjunto verdadero hasta los que generan el conjunto de consideración supuesto utilizando las alternativas mostradas a las personas.

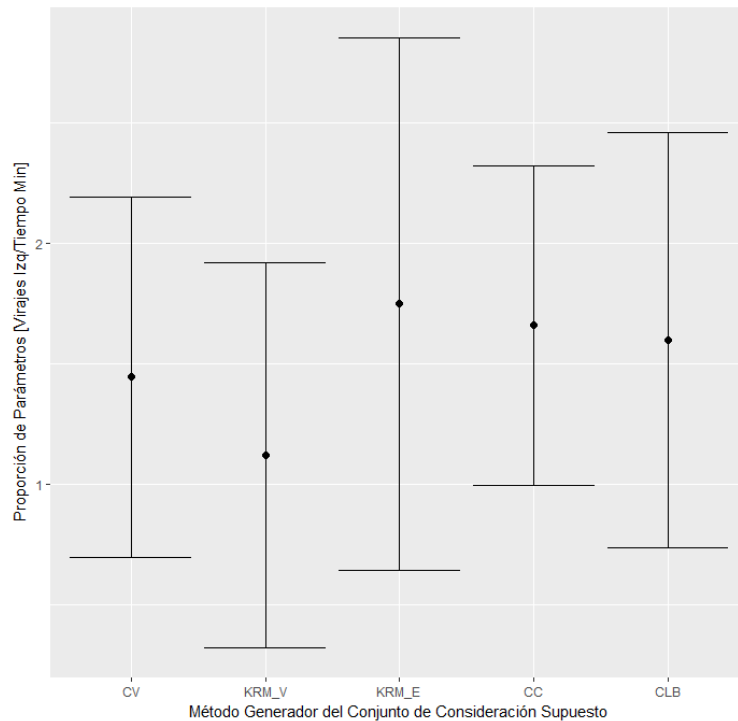


Figura 5.2: Comparación de modelos de elección utilizando distintos métodos de generación del conjunto de consideración supuesto, utilizando las alternativas mostradas.

De este gráfico no se pueden obtener mayores conclusiones ya que todos los modelos estimados contienen al valor real dentro de su rango. Lo que si se puede destacar es una mayor varianza y un pequeño sesgo en los métodos KRM_V y KRM_E, lo que de cierta

forma continúa con las sospechas planteadas en el capítulo 3 utilizando la simulación de Monte Carlo. Además, el método CLB muestra un gran parecido al modelo que utiliza el conjunto de consideración verdadero (CV), tanto en el valor de la proporción de los parámetros como en la desviación estándar de los mismos, lo que hace pensar que este método se comporta bien para representar la heurística que está detrás del proceso de generación del conjunto verdadero. Por otro lado, el método CC presenta un pequeño sesgo a cambio de una menor desviación estándar, lo que presenta menos fallos que en el caso de los datos simulados en el capítulo 3 sobre la simulación de Monte Carlo.

En la Tabla 5.20, se presentan los valores del estadístico para realizar un test t asintótico de dos colas, donde la hipótesis nula es que los parámetros de ambos modelos son iguales. El valor crítico del test para un 95% de confianza es 1.96, por lo que para ninguno de los modelos se puede rechazar la hipótesis de que los parámetros sean iguales a los del modelo que utiliza el conjunto de consideración verdadero presentado en la Tabla 5.10.

Tabla 5.20: Valores de test de especificación de razón de verosimilitud, método CV contra el resto, cuando se utilizan las alternativas mostradas para generar el conjunto de consideración supuesto.

Método	Valor del Test
KRM_V	0.485
KRM_E	0.375
CC	0.353
CLB	0.221

Utilizando el universo de alternativas

En la Figura 5.3 se presentan los intervalos de todos los métodos utilizados, desde el conjunto verdadero hasta los métodos generan el conjunto de consideración a partir del universo de alternativas. Los métodos KRM_V y KRM_E son presentados pero ya han sido rechazados a partir del análisis de los signos presentados en la Tabla 5.17 y Tabla 5.18.

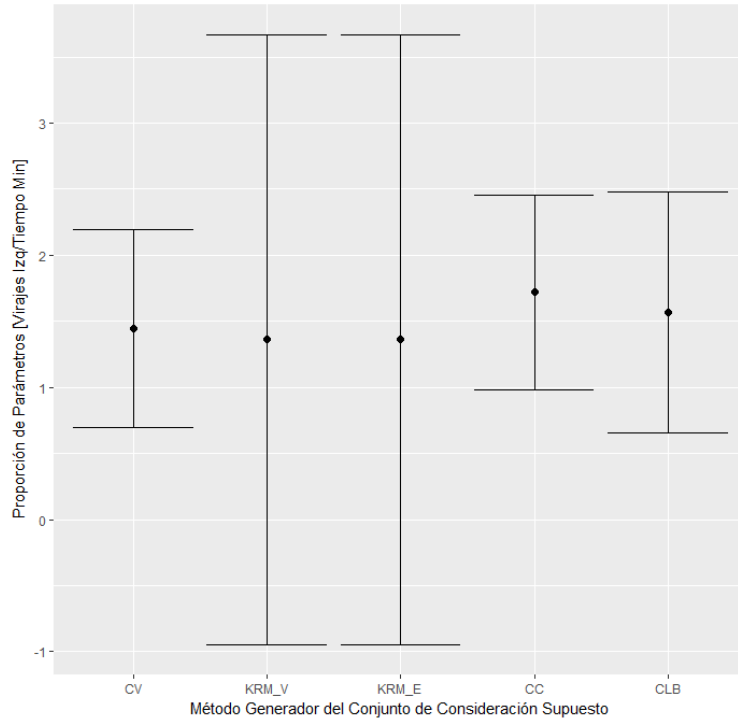


Figura 5.3: Comparación de modelos de elección utilizando distintos métodos de generación del conjunto de consideración supuesto, utilizando el universo de alternativas.

A partir de la figura, lo que más destaca es un gran parecido entre el método CLB y el modelo verdadero, tanto en el valor de la proporción los parámetros como en la desviación estándar de esta, lo que indica que este método muestra un buen potencial de representar a la heurística que está detrás del proceso de generación del conjunto verdadero. Por otra parte el método CC muestra un pequeño sesgo pero una menor desviación estándar de la proporción de los parámetros, al igual que en el caso anterior, lo cual nuevamente resulta ser un método más robusto que lo presentado en el capítulo 3 sobre la simulación de Monte Carlo.

En la Tabla 5.21, se presentan los valores del estadístico para realizar un test t asintótico de dos colas, donde la hipótesis nula es que los parámetros de ambos modelos son iguales. El valor crítico para un 95 % de confianza es 1.96.

Tabla 5.21: Valores de test de especificación de razón de verosimilitud, método CV contra el resto, cuando se utiliza el universo de alternativas para generar el conjunto de consideración supuesto.

Método	Valor del Test
KRM_V	-
KRM_E	-
CC	0.432
CLB	0.172

La tabla anterior ratifica las conclusiones presentadas por la Figura 5.3, donde los métodos CLB y CC se comportan bien en relación al modelo que utiliza el conjunto de consideración verdadero.

Discusión

Lo primero que se debe analizar es la diferencia que se causa al utilizar los métodos KRM_V y KRM_E en los dos escenarios planteados. Se debe recordar que estas reglas realizan un ranking de la utilidad sistemática de las alternativas y eligen las 3 mejores para formar parte del conjunto de consideración supuesto, sumando la elección al final. En el caso de utilizar solo las alternativas mostradas, las rutas que ve el modelador y el individuo que elige son las mismas, entonces al utilizar estos métodos no se incluyen alternativas que existen pero que pueden ser mucho mejores que la elección de la persona. En cambio, cuando se utiliza el universo de alternativas (20 rutas), el modelador ve todas las rutas y es posible que en muchos casos los métodos lleguen a incluir rutas que son totalmente dominantes sobre la elección de la persona. De esta manera, cuando los modelos de la Tabla 5.17 y Tabla 5.18 estiman los parámetros, se estima una lógica que dice que las personas eligen alternativas con más virajes a la izquierda y con más tiempo mínimo, porque en el conjunto de consideración supuesto ahora existen alternativas que la persona no vio. Si se considera la hipótesis de que el escenario que utiliza el universo de alternativas es más realista, entonces los resultados son preocupantes ya que los métodos deterministas entregan resultados al revés.

Por otro lado, los métodos CC y CLB se presentan como métodos robustos en este análisis, lo cual no se compara con los resultados obtenidos en el capítulo 3 sobre la simulación de Monte Carlo en los escenarios realistas. Esto puede indicar que las observaciones no son suficientes para mostrar un sesgo estadísticamente significativo, o que las limitaciones en la especificación de los modelos causan otro tipo de sesgos que se deben corregir antes de comparar estos métodos.

Finalmente, se debe mencionar que una extensión posible para métodos como CLB es realizar varias simulaciones del conjunto de consideración supuesto y obtener resultados del promedio de todas las estimaciones.

Capítulo 6

Conclusiones

6.1. Introducción

En este trabajo se presentan una serie de aportes para comprender el proceso de generación del conjunto de consideración. En primer lugar, se realiza una simulación de Monte Carlo para estudiar la robustez de cinco métodos generadores del conjunto de consideración por parte del analista bajo cuatro escenarios sobre el verdadero proceso de construcción. En segundo lugar, se realiza un análisis de tres métodos para recolectar datos sobre el conjunto de consideración, donde se estudian fuentes de preferencias declaradas y reveladas. Finalmente, mediante el uso de datos de preferencias declaradas, obtenidos a través de una encuestas online y sobre mapa, donde se les solicita a los participantes realizar un proceso de generación del conjunto de consideración. Con esto se logra estudiar características del conjunto de consideración, además de dar luces sobre el proceso de generación del conjunto y probar la robustez de distintos métodos utilizados para construirlo.

La simulación de Monte Carlo logra mostrar la robustez de los métodos, destacando el método que genera el conjunto de consideración con las alternativas experimentadas, además muestra que el método que emula a los métodos deterministas tradicionales falla sistemáticamente. Por otro lado, los resultados muestran que el método que considera el conjunto completo, falla cuando el conjunto se genera utilizando el principio de eliminación por aspectos. Por otra parte, el análisis de tres herramientas para la recolección de datos sobre el conjunto de consideración describe los pros y contras de cada una, identificando que los datos pasivos son en particular una buena herramienta para trabajar con rutas, pero las preferencias declaradas permiten adecuarse a la situación que se requiere estudiar. Luego, se logra realizar un análisis sobre el tamaño del conjunto, el cual requiere mayor estudio utilizando distintos métodos. El proceso de consideración se logra describir utilizando un modelo de consideración binario, además, las heurísticas declaradas por la gente requieren un mayor análisis debido a que se presentan algunas inconsistencias con lo declarado en el ejercicio de consideración. También se logra mostrar con éxito que los parámetros en un modelo de consideración son distintos a los de elección. Finalmente, el análisis comparativo de métodos de generación del conjunto de consideración supuesto, muestra conclusiones similares (un poco más débiles) que las del análisis de Monte Carlo.

6.2. Conclusiones generales

Como primera conclusión, se puede ver que el análisis realizado en la simulación de Monte Carlo y utilizando los datos de preferencias declaradas, indican que los métodos deterministas son los que suelen fallar más. Sin embargo, se debe recordar que estos métodos son una simplificación de lo que existe en la literatura. Además, de la simulación de Monte Carlo, se desprende que el método que utiliza las rutas experimentadas muestra un gran potencial para la construcción del conjunto de consideración, sobre todo gracias al aumento en disponibilidad de los datos de serie de tiempo. Una de las características que hacen más robusto a dicho método, es que todas las rutas que se incluyen en el conjunto de consideración pertenecen al conjunto verdadero, solo pueden faltar alternativas no observadas. Incluso, es posible conjeturar que las alternativas no observadas, si pertenecen al conjunto de consideración verdadero, es posible que su probabilidad de elección sea baja.

Por otro lado, luego del análisis de 3 herramientas para la recolección de datos sobre el conjunto de consideración, se puede concluir que, si bien los datos de preferencias reveladas son ideales, ya que se presentan sobre una estructura de red, éstos no logran revelar suficiente información sobre el conjunto de consideración como para generar una caracterización completa. Por otro lado, se debe tener particular precaución al utilizar datos de preferencias declaradas que obligan a la gente a interactuar con un mapa, ya que puede generarse un sesgo debido a que no toda la gente sabe como leer un mapa cuando se enfrenta a él, si embargo este tipo de datos se presenta como una buena fuente de información que permite adecuarse al contexto y se enfoca en responder las preguntas de investigación en esta memoria.

Finalmente, se concluye que se puede lograr una buena caracterización del proceso de generación del conjunto de consideración, mostrando un análisis sobre el tamaño del conjunto y especialmente estimando el modelo de consideración, que destaca las variables relevantes, siendo además un buen método para simular el conjunto de consideración.

6.3. Recomendaciones metodológicas

La principal recomendación metodológica que se pueden obtener a partir de este trabajo es con respecto a los tipos de datos que se pueden utilizar para analizar el fenómeno estudiado. El ejercicio de consideración y elección presentado es bueno para esta situación específica de caracterización, sin embargo, los datos de preferencias reveladas siempre serán preferibles ya que muestran el verdadero comportamiento de las personas en una situación real. Es por esto que se recomienda estudiar los datos pasivos para poder construir conjuntos de consideración que superen a los métodos actuales, principalmente con el método CE mencionado en el capítulo 3 sobre la simulación de Monte Carlo.

Por otra parte, el modelo de consideración logró demostrar ser una buena herramienta, a pesar de la simpleza de la especificación presentada. Desde el punto de vista de la caracterización del conjunto, es una forma muy completa que se recomienda seguir estudiando.

6.4. Extensiones

En la etapa de esta investigación que se basó en una simulación de Monte Carlo, existe una infinidad de extensiones que se pueden hacer, partiendo desde hacer el caso más realista según elección de rutas, hasta probar una serie de métodos diferentes para generar el conjunto de consideración supuesto. Además, es interesante poder desarrollar con mayor profundidad el método propuesto para generar el conjunto de consideración supuesto que utiliza las alternativas experimentadas (CE). Con certeza deben existir datos de serie de tiempo con los que se pueda experimentar y llegar a una mejor conclusión sobre el potencial de este método práctico. Por otro lado, el método de las k -ésimas rutas mínimas utilizando parámetros a priori errados (KRM_E) se presenta con un solo set de parámetros a priori (ver Tabla 3.3), sin embargo, un análisis de sensibilidad puede ser realizado al modificar estos parámetros.

Debido a que los datos de preferencias reveladas muestran un buen potencial, se pueden realizar estudios en profundidad para poder utilizar con mayor facilidad este tipo de datos. Si bien puede que no demuestren exactamente el conjunto de consideración de una persona, muestran ser prácticos y pueden ayudar a entender otros fenómenos de elección de rutas.

Finalmente, la encuesta de preferencias declaradas presencial sobre mapa muestra ser buena para estudiar el tamaño del conjunto, ya que se muestra como un experimento más realista que su contraparte. Para que sean factibles de utilizar, se requiere un mayor número de observaciones y un mejor diseño de la encuesta, posiblemente con diversos pares origen-destino y distintos periodos del día.

Bibliografía

- Bekhor, S., Ben-Akiva, M. E., y Ramming, M. S. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144(1), 235–247.
- Ben-Akiva, M., Bergman, M., Daly, A. J., y Ramaswamy, R. (1984). Modeling inter-urban route choice behaviour. En *Proceedings of the 9th international symposium on transportation and traffic theory* (pp. 299–330).
- Ben-Akiva, M., y Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, 12(1), 9–24.
- Ben-Akiva, M., y Lerman, S. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9). MIT press.
- Brown, J. J., y Wildt, A. R. (1992). Consideration set measurement. *Journal of the Academy of Marketing Science*, 20(3), 235–243.
- Cain, P. (2016). *Google maps timeline: Why a little-known google feature tracked me for months*. Descargado 2016-06-11, de <https://globalnews.ca/news/2746703/google-maps-timeline-why-a-little-known-google-feature-tracked-me-for-months/>
- Daly, A., y Hess, S. (2010). Simple approaches for random utility modelling with panel data. En *European transport conference*.
- Daly, A., Hess, S., y de Jong, G. (2012). Calculating errors for measures derived from choice modelling estimates. *Transportation Research Part B: Methodological*, 46(2), 333–341.
- Gandhi, U. (2017). *Buscando y descargando datos de openstreetmap*. Descargado 2017-12-05, de http://www.qgistutorials.com/es/docs/downloading_osm_data.html
- Hauser, J. R. (2014). Consideration-set heuristics. *Journal of Business Research*, 67(8), 1688–1699.
- Henningsen, A., y Toomet, O. (2011). maxlik: A package for maximum likelihood estimation in R. *Computational Statistics*, 26(3), 443–458. Descargado de <http://dx.doi.org/10.1007/s00180-010-0217-1> doi: 10.1007/s00180-010-0217-1
- Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, 8(3), 229–254.
- McFadden, D. L. (1984). Econometric analysis of qualitative response models. *Handbook of Econometrics*, 2, 1395–1457.
- Papinski, D., y Scott, D. M. (2011). A gis-based toolkit for route choice analysis. *Journal of Transport Geography*, 19(3), 434–442.
- Prato, C. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, 2(1), 65–100.
- Prato, C., y Bekhor, S. (2007). Modeling route choice behavior: how relevant is the composition of choice set? *Transportation Research Record: Journal of the Transportation*

- Research Board*(2003), 64–73.
- QGIS Development Team. (2009). Qgis geographic information system [Manual de software informático]. Descargado de <https://www.qgis.org/es/site/>
- R Development Core Team. (2008). R: A language and environment for statistical computing [Manual de software informático]. Vienna, Austria. Descargado de <http://www.R-project.org> (ISBN 3-900051-07-0)
- Ramming, M. S. (2001). Network knowledge and route choice. *Unpublished Ph. D. Thesis, Massachusetts Institute of Technology*.
- Raveau, S., Guo, Z., Muñoz, J. C., y Wilson, N. H. (2014). A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. *Transportation Research Part A: Policy and Practice*, 66, 185–195.
- Rieser-Schüssler, N., Balmer, M., y Axhausen, K. W. (2013). Route choice sets for very high-resolution data. *Transportmetrica A: Transport Science*, 9(9), 825–845.
- SECTRA. (2014). *Encuesta origen destino de viajes 2012*. Santiago, Chile: Sectra - Universidad Alberto Hurtado. Descargado de <http://www.sectra.gob.cl/biblioteca/detalle1.asp?mfn=3253>
- Swait, J., y Ben-Akiva, M. (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*, 21(2), 91–102.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Descargado de <http://ggplot2.org>

Apéndice A

Tablas adicionales

Tabla A.1: Resumen de Estadísticas de Datos Generales de los Participantes.

Número de participantes	189	
Trabaja o estudia en Beauchef		
Sí	97	51.3 %
No	92	48.7 %
Género		
Femenino	116	61.4 %
Masculino	73	38.6 %
Edad		
18 - 24	129	68.3 %
25 - 34	46	24.3 %
35 - 44	4	2.1 %
45 - 54	8	4.2 %
55 - 64	2	1.1 %
Nivel educacional		
Media Científico - Humanista	107	56.6 %
Media Técnico - Profesional	5	2.7 %
Instituto Profesional	2	1.1 %
Universitaria	75	39.7 %
Ocupación		
Busca trabajo por primera vez	6	3.2 %
Desempleado	2	1.1 %
Dueño(a) de casa	5	2.7 %
Estudia	138	73.0 %
Trabaja	34	18.0 %
Otro	4	2.1 %
Tiempo vivido en Santiago		
Nunca he vivido en Santiago	5	2.7 %
Menos de 6 meses	4	2.1 %
Entre 1 año y 3 años	13	6.9 %
Entre 6 meses y 1 año	3	1.6 %
Más de 3 años	164	86.8 %
Principal modo de transporte al trabajo o estudio		
A pie	15	7.9 %
Auto Acompañante	4	2.1 %
Auto Chofer	14	7.4 %
Bicicleta	8	4.2 %
Bus	38	20.1 %
Bus-Metro	83	43.9 %
Bus interurbano o rural	4	2.1 %
Bus urbano con pago al conductor (Metrobus y otros)	2	1.1 %
Metro	19	10.1 %
Motocicleta Chofer	1	0.5 %
Taxi o radiotaxi	1	0.5 %

Tabla A.2: Resumen de Estadísticas de Datos del Hogar de los Participantes.

Número de participantes	189	
Posesión de licencia		
Sí	89	47.1 %
No	100	52.9 %
Número de personas en el hogar		
1	8	4.2 %
2	35	18.5 %
3	34	18.0 %
4	55	29.1 %
5	42	22.2 %
6 o más	15	7.9 %
Número de licencias en el hogar		
0	25	13.2 %
1	59	31.2 %
2	52	27.5 %
3	35	18.5 %
4	17	9.0 %
5	1	0.5 %
Vehículos motorizados disponibles		
0	48	25.4 %
1	72	38.1 %
2	52	27.5 %
3 o más	17	9.0 %
Días de uso de vehículo motorizado		
0	122	64.6 %
1	12	6.4 %
2	12	6.4 %
3	9	4.8 %
4	5	2.7 %
5	15	7.9 %
6	5	2.7 %
7	9	4.8 %
Ingreso mensual líquido del hogar		
Menos de 400.000 pesos	20	10.6 %
Entre 1.600.001 y 2.400.000 pesos	27	14.3 %
Entre 2.400.001 y 4.800.000 pesos	26	13.8 %
Entre 400.001 y 800.000 pesos	44	23.3 %
Entre 800.001 y 1.600.000 pesos	51	27.0 %
Más de 4.800.000 pesos	8	4.2 %
No contesta	13	6.9 %

Tabla A.3: Resultados porcentuales de las respuestas sobre las heurísticas utilizadas para generar el conjunto de consideración

	No me representa	Me representa ligeramente bien	Me representa moderadamente bien	Me representa muy bien	Me representa perfectamente
Considero las alternativas de menor tiempo promedio.	1 %	8 %	23 %	38 %	30 %
Considero las alternativas de menor tiempo en un escenario optimista.	8 %	13 %	25 %	35 %	19 %
Considero las alternativas de menor tiempo y distancia en promedio.	6 %	10 %	27 %	33 %	24 %
Considero las alternativas con menos virajes a la izquierda.	67 %	11 %	8 %	7 %	7 %
Considero las alternativas con más tramos en calles seguras.	25 %	15 %	19 %	28 %	13 %
Considero las alternativas que utilizan vías de alta capacidad (Ej. Alameda).	20 %	17 %	34 %	18 %	11 %
Considero las alternativas que utilizan más autopistas.	35 %	18 %	24 %	16 %	6 %
Considero las alternativas de menor distancia.	3 %	8 %	25 %	37 %	27 %
Considero las alternativas con menor número de calles distintas.	12 %	13 %	21 %	28 %	26 %