

# CARACTERIZACIÓN DEL CONJUNTO DE CONSIDERACIÓN EN ELECCIÓN DE RUTA

G. Nicolás Villalobos <sup>1\*</sup>, C. Angelo Guevara <sup>2</sup>

<sup>1</sup> Secretaría de Planificación de Transporte, Ministerio de Transportes y Telecomunicaciones, Chile

<sup>2</sup> Universidad de Chile, Chile

---

\*Autor para correspondencia: **RESUMEN**

[guillermo.nicolas@gmail.com](mailto:guillermo.nicolas@gmail.com)

En la modelación de elecciones discretas, el conjunto de consideración contiene las alternativas para realizar una elección. Este conjunto es difícil de conocer y, representarlo erróneamente, redundará en modelos incapaces de representar la realidad, entregando estimadores inconsistentes y errores de predicción. Este trabajo realiza tres contribuciones abordando este desafío. La primera utiliza simulaciones de Monte Carlo para explorar la robustez de métodos prácticos utilizados para construir este conjunto, mostrando que métodos tradicionales generan sesgos en la estimación y que utilizar elecciones históricas presenta resultados prometedores. La segunda analiza tres métodos para recolectar datos sobre este conjunto: datos pasivos, encuesta online y encuesta sobre mapa. Los datos pasivos tienen la ventaja de mostrar preferencias reveladas, pero requieren mucho procesamiento. Las encuestas mostraron ser factibles a nivel exploratorio, pero son susceptibles al sesgo hipotético y recolectarlos es costoso. La tercera utiliza un experimento de preferencias declaradas intentando replicar el proceso de generación del conjunto para estudiar sus características y probar la robustez de ciertos métodos para construirlo. Este ejercicio permitió describir el tamaño del conjunto, identificar heurísticas que la gente dice utilizar para generarlo, además de modelar el proceso de consideración. El artículo concluye resumiendo las contribuciones, limitaciones y alcances prácticos del trabajo.

**Palabras clave** conjunto de consideración, elecciones discretas, Monte Carlo, RP, SP.

## **ABSTRACT**

*In discrete choices, the consideration set offers the alternatives to make a choice. This set is difficult to know, and any mistake will imply that models will not represent reality, delivering inconsistent estimator and prediction error. This paper makes three contribution to face this challenge. The first one uses Monte Carlo simulations to explore the robustness of practical methods to build this set, showing that traditional methods cause bias in the estimations, whereas using historical events provides promising results. The second one analyzes three methods to gather data on this set: passive data, online and map survey. Passive data have the advantage of showing revealed preferences but require a lot of processing. Data gathered by surveys showed to be feasible, but susceptible to hypothetical bias and gathering them was expensive. The third one uses a declared preferences experiment trying to replicate the process of generating this set, to study its characteristics and demonstrate robustness of certain methods used to build it. This exercise allowed us to describe the size of the set, identify heuristics that people say that they use, and to model the process of consideration. The article finishes with the contributions, limitations, and practical implications of this paper.*

**Keywords** consideration set, discrete choice, Monte Carlo, RP, SP.

---

---

## 1. INTRODUCCIÓN

En muchas situaciones prácticas, un individuo se ve enfrentado a elegir una alternativa entre un gran número de opciones disponibles. A pesar de que existe un beneficio asociado a considerar un mayor número de alternativas (debido a que esto aumenta la máxima utilidad alcanzable), también existe un costo de explorar una alternativa extra, causado por varios factores como el agotamiento o el límite cognitivo de las personas. Por esta razón, se supone que, al enfrentar un gran número de alternativas, las personas utilizan ciertas heurísticas para reducir el número de opciones a escrutar en detalle, en lo que se define como el conjunto de consideración (Hauser, 2014).

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, como la elección de rutas. El problema es que violar este supuesto puede causar sesgos en la estimación de los parámetros en modelos de elección y pobres predicciones de elecciones.

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 tienen ventajas y desventajas, que serán revisadas junto a más literatura en la sección 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.

El objetivo principal de este trabajo 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:

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

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 declaradas se estudian mediante dos encuestas. La primera es una encuesta online, que recopila datos socioeconómicos y realiza un ejercicio de simulación del proceso de generación del conjunto de consideración, junto con un ejercicio de elección y la recopilación de otros datos de interés. La segunda es una encuesta presencial sobre un mapa, donde se preguntan los mismos datos socioeconómicos 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. Todos estos métodos se realizan en base a la ciudad de Santiago, Chile.

Finalmente, con el tercer 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.

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.

### **1.1 Modelación de elecciones discretas incorporando el conjunto de consideración**

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 1 se muestra la expresión desarrollada, la cual puede parecer simple, pero poder conocer la probabilidad de cada posible conjunto de consideración para evaluar dicha expresión es impracticable.

$$P_n(i) = \sum_{C \in G_n} P_n(i | C) \cdot P_n(C) \quad (1)$$

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 necesite más estudio en esta línea de trabajo.

## 1.2 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 este trabajo, 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 este trabajo, brindando un enfoque de comprensión del conjunto de consideración en lugar de soluciones prácticas.

### 1.3 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 superponen.

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.

## 2. ESTUDIO DEL IMPACTO DE UTILIZAR DISTINTOS CONJUNTOS DE CONSIDERACIÓN MEDIANTE MÉTODO DE MONTE CARLO

En esta sección 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 cada iteración de la simulación de Monte Carlo, se generan 2000 observaciones, donde 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.

Una vez generados los datos de las alternativas para todas las observaciones, para cada escenario se generan las elecciones mediante un modelo logit multinomial. La especificación de la componente sistemática de la utilidad es la que se presenta en la Ecuación 2. Donde  $th_x$  corresponde al parámetro de la variable X, el subíndice i corresponde a la alternativa y el subíndice n corresponde al individuo.

$$V_{in} = th_{tiempo} \cdot tiempo_{in} + th_{costo} \cdot costo_{in} \quad (2)$$

### 2.1 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:

- 
- Aleatorio (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.
  - Todas las alternativas (Todas): 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.
  - Eliminación por aspecto (EBA): En esta heurística se crea el conjunto de consideración verdadero mediante un algoritmo que se asemeja a eliminación por aspecto (Tversky 1972), eliminando las alternativas cuyo costo o tiempo es sobre 0.8 desviaciones estándar sobre el promedio. 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.
  - Modelo logit binario de consideración (CBL): La última regla consiste en utilizar un modelo logit binario, donde para cada alternativa a considerar se calcula una probabilidad basada en los parámetros de tiempo y costo del modelo de elección original, de manera que simulando números aleatorios uniformes entre 0 y 1 se obtiene la predicción de que una alternativa sea o no considerada por cada individuo. La especificación de la utilidad para este modelo se puede encontrar en Villalobos (2018).

## 2.2 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 es 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.

- 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.
- 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 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. 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.

- 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. Este método se utiliza ya que se acerca a la realidad donde el modelador no conoce los parámetros reales y debe utilizar una función de costo generalizado con parámetros arbitrarios.
- Conjunto de consideración con logit binario (CLB): Por último, este método es el único que pertenece a la familia de los métodos estocásticos. Consiste en estimar un modelo logit binario, donde a partir del conjunto de consideración verdadero, se utiliza cada alternativa como una observación. Luego, las elecciones corresponden al hecho de que una alternativa sea considerada 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. Este método es de esperar que tienda a representar a la heurística que está detrás de la generación del conjunto de consideración verdadero.
- 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 y que por ende pertenecen al conjunto de consideración verdadero, generando de esta manera un subconjunto del conjunto de consideración real. Es posible comparar este método a lo que se puede obtener de una base de datos de serie de tiempo. Este método muestra ventajas en cuanto la factibilidad de su aplicación práctica. Las alternativas observadas pueden ser obtenidas mediante datos pasivos como GPS en el caso de elecciones de ruta o con el historial de compras de una tarjeta bancaria en el caso de cualquier producto. Una vez que se estima el modelo de consideración, se utiliza el mismo modelo en modalidad predictiva 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, si la probabilidad de ser elegida de una alternativa (número entre 0 y 1) es mayor que el número aleatorio simulado (número entre 0 y 1) entonces dicha alternativa se agrega al conjunto de consideración supuesto del individuo.

### 2.3 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 bajo los 4 escenarios de generación del conjunto de consideración verdadero. En la Tabla 1 se puede observar el P-Valor de recuperación de los parámetros, donde un valor cercano al bajo el 5% se interpreta como un mal método de estimación para el escenario mostrado, se omite fila para CV por ser un método de chequeo únicamente.



**Tabla 1** P-Valor de recuperación de parámetros de la población

Método/Escenario	I. Todas	II. Aleatorio	III. EBA	IV. CLB
i. CC	91%	~100%	~0%	~0%
ii. KRM_V	65%	86%	~0%	~0%
iii. KRM_E	~0%	~0%	~0%	~0%
iv. CLB	91%	~100%	~0%	55%
v. CE	88%	90%	88%	90%
Bajo P-Value (<5%)				

Con lo anterior, 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.

### **3. ANÁLISIS CRÍTICO DE TRES MÉTODOS PARA RECOLECTAR DATOS SOBRE EL CONJUNTO DE CONSIDERACIÓN**

Con el fin de realizar un estudio con datos empíricos, en esta sección 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. Todos estos experimentos tienen cabida en Santiago, Chile, por lo que se encuentran sujetos a posibles factores locales. A continuación 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.

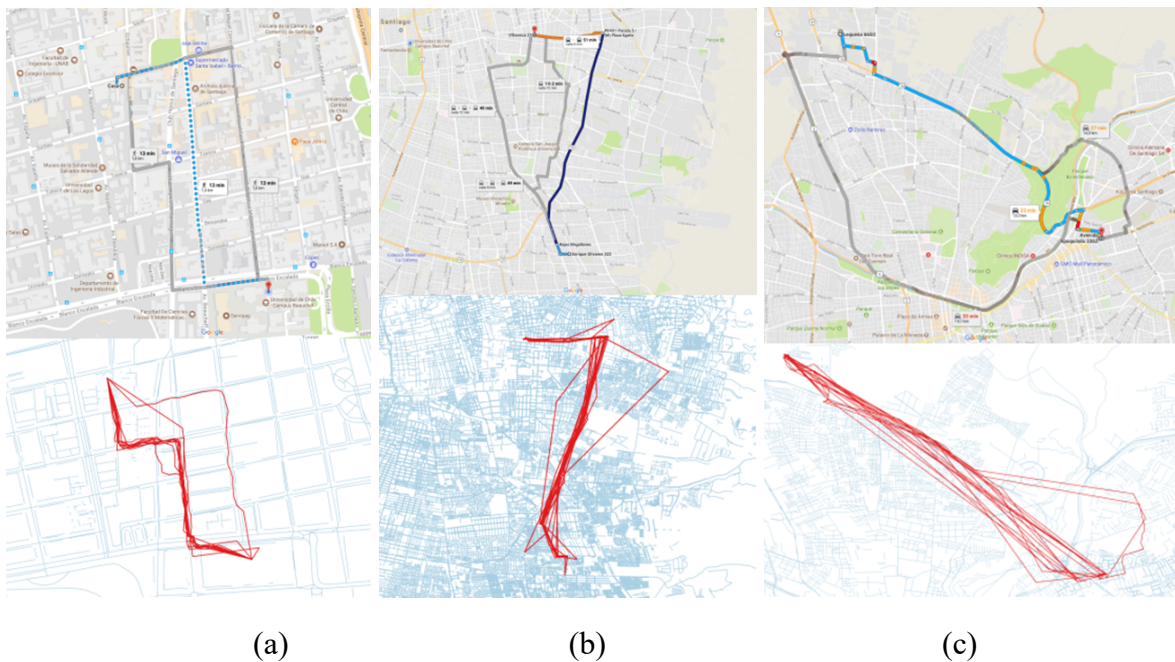
#### **3.1 Datos pasivos**

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. Los pasos seguidos para obtener los datos se detallan en Villalobos (2018).

En la Figura 1(a) 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 sólo posee un trazo erróneo. En segundo lugar, se debe notar que aunque haya rutas que aparentemente son iguales, ningún elemento vectorial es igual a otro, esto tiene que ver directamente con 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.



**Figura 1** Datos Pasivos (a) Caminata, (b) Transporte Público (c) Auto

Por otro lado, en la Figura 1(b) se presenta al usuario transporte público, en este caso la mayor parte de las rutas se realizan en una combinación de metro y bus 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 1 (a) 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 1(c) 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 práctica 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 conseguir 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 anteriormente. 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. A pesar de lo anterior, no es posible desarrollar esta herramienta en el alcance de este trabajo 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.

### **3.2 Encuesta de preferencias declaradas online**

Como segundo método de recolección de datos, se estudió 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.

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 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 2(a) se presenta el ejemplo de una ruta preguntada.

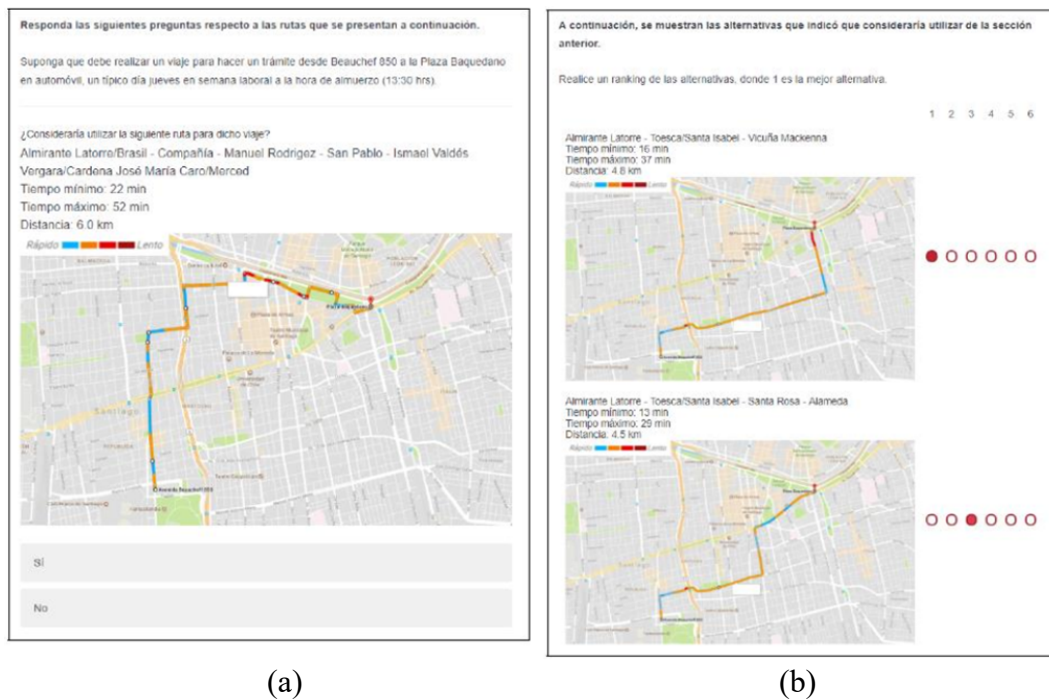


Figura 2 Ejemplo de (a) ruta preguntada en encuesta online (b) ejercicio de elección con ranking de alternativas

---

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 2(b) se muestra un ejemplo del sistema de ranking de alternativas implementado.

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 como las que se utilizan para calcular rutas mínimas en el artículo de Bekhor et al. (2006), entre las cuales se tienen por ejemplo la consideración de rutas basados en un menor tiempo de viaje, una menor distancia, una menor cantidad de virajes, entre otros. Más detalles sobre este ejercicio puede ser encontrado en Villalobos (2018).

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 sólo a las personas que consideraron una 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 fue necesaria otra forma de filtrar los datos. Por otro lado, es de importancia destacar que la población encuestada mediante este método posee una alta proporción de personas del grupo etario entre 18 a 24 años dado el contexto universitario para la distribución de la encuesta.

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.

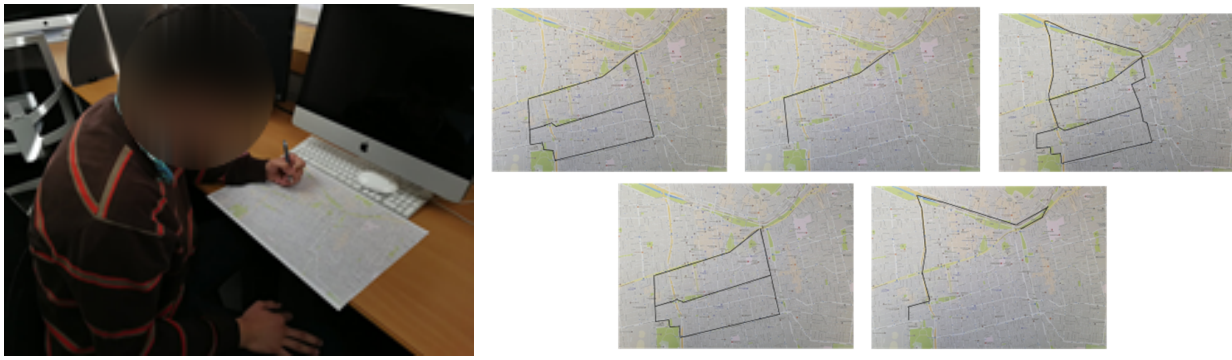
En el caso del experimento de consideración, se debe notar que otra opción es mostrar todas las alternativas disponibles de una vez y no preguntarlas una a una, ya que de esta manera se representaría mejor la realidad. El método actual con el que se presentan los datos puede causar un sesgo de orden o lexicográfico.

### **3.3 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 escenario como el de la Figura 3(a), 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 jueves normal, a las 13:30 hrs. con propósito trámite.



(a) Ejemplo de escenario de encuesta sobre mapa

(b) Respuestas de encuesta sobre 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 3(b). 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 obtenidos mediante la recolección 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.

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, conseguir esta pequeña muestra requiere de mucho más esfuerzo en cuanto a conseguir a los encuestados para participar de manera presencial. La encuesta 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 requerirían horas de trabajo.

#### 4. CARACTERIZACIÓN DEL PROCESO DE CONSIDERACIÓN MEDIANTE DATOS DE PREFERENCIAS DECLARADAS

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 este trabajo) 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 esta sección se 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 y del experimento de consideración realizado en la encuesta sobre mapa, donde se busca saber si existe un único tamaño o si depende, por ejemplo, del grupo socio-económico. Luego, se estima 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 realizado en la simulación de Monte Carlo. Esto busca corroborar con los datos de la encuesta online las conclusiones obtenidas anteriormente.

#### **4.1 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.

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 aproximadamente la mitad de las personas se les mostraron 10 alternativas y a la otra mitad 15 alternativas, 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, 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. Los detalles del test se pueden ver en Villalobos (2018), donde se obtiene que los valores son estadísticamente distintos.

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.

Utilizando encuesta de preferencias declaradas sobre mapa se estudia el tamaño del conjunto con los datos obtenidos con la encuesta sobre mapa. 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 anteriormente, 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, 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 y con muchas más posibilidades de rutas.

## 4.2 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, donde se utiliza cada pregunta del ejercicio de consideración de la encuesta online, 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 puede encontrar en Villalobos (2018) junto con más detalles de los resultados obtenidos. 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.

Por otro lado, en la Tabla 2 se muestra el modelo de consideración definitivo, donde todas las variables son significativas. Entre estas variables se encuentran “Rojo” que corresponde a la

proporción de tramo de color rojo que se presenta en el mapa de la ruta, lo que indica velocidades bajas, número de virajes a la derecha y a la izquierda, el tiempo mínimo que se puede demorar en completar la ruta, y variables de la persona sobre si trabaja y utiliza el automóvil para viajar al trabajo o estudio al menos 1 vez a la semana. Además de estas variables, se presenta una corrección a causa del diseño experimento, representada por la variable Orden >10, debido al sesgo lexicográfico existente ya que las alternativas son presentadas una a una y las limitaciones cognitivas o cansancio de las personas resultan en una menor probabilidad de considerar alternativas más allá de la décima alternativa presentada.

**Tabla 2** Modelo definitivo de consideración  
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										
$\rho^2$ ajustado	0.237										
Log-verosimilitud	-1244.843										
Códigos de significancia:	0	****	0.001	***	0.01	**	0.05	*	0.1	.	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. 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 Villalobos (2018) se puede encontrar un análisis más en detalle de los pesos de cada parámetro en el modelo.

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. Mostrando que las variables de virajes son estadísticamente significativas y variables como el tiempo mínimo se ratifican en este análisis.

#### 4.3 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 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 décima alternativa presentada podría corresponder a la primera alternativa en el modelo, sin embargo, para la persona 2, la misma alternativa podría ser la quinta alternativa 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. Lo anterior es posible ya que lo que se busca no es encontrar parámetros asociados a características específicas de cada alternativa, sino que los parámetros asociados a variables presentes en todas ellas.

Como primera instancia, se intentaron utilizar las mismas variables que el modelo de consideración, pero la variable sobre el orden no tiene sentido en este modelo (dado que la variable hace 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.

Los resultados de la estimación del mejor modelo de elección se presentan en la Tabla 3, 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 4. 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  $\rho^2$  ajustado, pero no se debe comparar debido a que el número de observaciones es diferente.

**Tabla 3** Modelo de elección utilizando ranking de alternativas

Logit Multinomial de Elección Utilizando Ranking					
	Parámetro	Error Estándar	t-valor	Prob(>t)	
Rajo	-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				
$\rho^2$ ajustado	0.097				
Log-verosimilitud	-1106.265				
Códigos de significancia:	0	****	0.001	***	0.01 ** 0.05 . 0.1 ' ' 1

**Tabla 4** Modelo de elección sin utilizar ranking de alternativas

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				
$\rho^2$ ajustado	0.231				
Log-verosimilitud	-239.388				
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 simultáneo una de estas escalas es estimable. En los modelos estimados esta escala se ve presentada como  $\mu$ eleccion.

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). Y el segundo modelo donde los 4 parámetros comunes son iguales (modelo restringido). Mayor detalle sobre estos modelos puede ser encontrado en Villalobos (2018).

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. Dicho test se muestra a continuación (Ecuación 3) e indica que se rechaza la hipótesis nula y por lo tanto los parámetros son estadísticamente distintos.

$$\begin{aligned}
 -2(-L(\text{restringido}) - L(\text{sin restringir})) &= -2(-2365,812 - -2351,108) & (3) \\
 &= 11,208 > \chi^2(4,95\%) = 9,488
 \end{aligned}$$

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, ya que si las personas tuviesen la capacidad cognitiva de evaluar cada alternativa como si fuera una elección, entonces no existiría la necesidad de generar un conjunto de consideración en primer lugar, 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.

#### 4.4 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, 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.

En esta última etapa, se comparan los modelos recién mostrados 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. Más detalles sobre este análisis comparativo puede ser encontrado en Villalobos (2018). En la Tabla 5 se puede ver el P-Valor de recuperación de parámetros, lo que indica si es posible rechazar alguno de los métodos aplicados.

**Tabla 5** P-Valor de recuperación de parámetros de la población

Método/Configuración	(10 o 15)	(20)
Completo	72%	67%
K-ésimas rutas mínimas, betas verdaderos	63%	-
K-ésimas rutas mínimas, betas errador	71%	-
Modelo logit binario de consideración	83%	86%
Experimentado	infactible	infactible

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 (10 o 15), 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 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 con signos 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 con 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.

## 5. CONCLUSIONES

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 encuesta 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, indicando 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. 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.

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 cómo leer un mapa cuando se enfrenta a él, sin 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 este trabajo.

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.

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 la sección 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.

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, 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.

**AGRADECIMIENTOS** Esta publicación fue financiada por CONICYT, FONDECYT 1191104 y por el Instituto de Sistemas de Ingeniería Complejos (CONICYT - PIA - FB0816). Todos los modelos y análisis de datos se desarrollaron utilizando el software de código abierto R (R Development Core Team, 2008).

## REFERENCIAS

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*, 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*. 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. Recuperado el 11 de junio 2016 en: <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*. 11-13 octubre 2010, Glasgow, Reino Unido.

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. Recuperado el 5 de diciembre 2017 en: [http://www.qgistutorials.com/es/docs/downloading\\_osm\\_data.html](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]. Recuperado el 18 de marzo 2021 en: <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. Recuperado el 18 de marzo 2021 en: <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. Recuperado el 18 de marzo 2021 en: <http://www.sectra.gob.cl/biblioteca/detalle1.asp?mf=325>
- 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.
- Tversky, A. (1972) Elimination by aspects: A theory of choice. *Psychological review*, 79(4), 281.

---

Villalobos, G.N. (2018) Caracterización del conjunto de consideración en elección de ruta. Memoria Ingeniería Civil, Universidad de Chile.

Wickham, H. (2009) ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. Recuperado el 18 de marzo 2021 en: <http://ggplot2.org>