



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

SEGREGACIÓN ESPACIAL EN LA REGIÓN METROPOLITANA

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN ECONOMÍA APLICADA

MATÍAS EZEQUIEL ARRIOLA ARAVENA

PROFESORA GUÍA:
SOFÍA CORREA DEISLER

PROFESOR CO-GUÍA:
MARCEL GOIC FIGUEROA

COMISIÓN:
RAIMUNDO UNDURRAGA RIESCO

Este trabajo ha sido parcialmente financiado por:
Proyecto FONDECYT n° 1221711 y
MIPP (Instituto Milenio Imperfecciones de Mercado y Política Públicas)

SANTIAGO DE CHILE
2023

SEGREGACIÓN ESPACIAL EN LA REGIÓN METROPOLITANA

En este trabajo, estudiamos el nivel de segregación que experimentan las personas de diferentes orígenes socioeconómicos a lo largo de sus días. Utilizando datos detallados de teléfonos celulares para la ciudad de Santiago, construimos medidas de segregación experiencial y caracterizamos su evolución en el tiempo. Específicamente, introducimos dos medidas novedosas para cada zona censal: (i) *Segregación Experiencial de Entrada*, que captura la segregación en las interacciones que ocurren dentro de la zona censal, y (ii) *Segregación Experiencial de Salida*, que cuantifica la segregación experimentada por los residentes de la zona censal dada en sus interacciones a lo largo del día. La distinción es crucial ya que nos permite separar el entorno que rodea a una determinada zona censal de las personas que viven en ella.

Calculamos cada uno de estos índices diariamente entre marzo de 2020 y julio de 2022 y estudiamos su evolución. Ambas medidas han permanecido relativamente constantes a lo largo del tiempo, con la *Segregación Experiencial de Entrada* mostrando una variabilidad ligeramente mayor. Al compararlas con la *Segregación Residencial*, encontramos que ambas medidas experienciales dan índices de segregación más bajos que la *Segregación Residencial*, siendo la *Segregación Experiencial de Salida* la más baja. Esto se traduce en que dada una zona, en promedio las personas en la RM tienden a interactuar con sujetos más diversos (o distintos a sí mismos) en sus lugares de destino que a los que podrían encontrar en sus lugares de habitación.

A su vez, las medidas experienciales presentan heterogeneidad entre distintos grupos socioeconómicos, donde los grupos de altos ingresos segregan al rededor de cuatro veces más que los grupos de ingresos medios y bajos. Las causas subyacentes a este resultado están fuera del alcance de esta tesis, no obstante, este resultado se puede deber a que sujetos de altos ingresos poseen pocos o nulos incentivos a dirigirse a otros sectores de un nivel de ingreso más bajo, dado que poseen todos los servicios necesarios para vivir dentro de sectores de su mismo nivel y/o por preferencias respecto al nivel socioeconómico de las personas con quienes interactúan.

Para comprender mejor las dinámicas que afectan las interacciones en la ciudad, también investigamos el papel de la infraestructura crítica en cada una de estas medidas de segregación. Para hacerlo, primero identificamos varias categorías de infraestructura que podrían facilitar las interacciones de las personas (parques recreativos, centros comerciales, estaciones de metro) y estudiamos el efecto de cada una de estas categorías en la segregación experiencial. Luego, utilizando datos de panel, obtenemos que el metro es un factor importante para integrar la ciudad, que combinado con malls y parques, sirve para disminuir tanto la *Segregación Experiencial de Salida* como la *Segregación Experiencial de Entrada* entre zonas de altos ingresos que de otra manera no se integrarían.

La distinción entre la interacción en el origen y el destino, así como la caracterización de la infraestructura de la ciudad, proporcionan nuevos y valiosos conocimientos sobre las dinámicas en las que las personas participan en sus actividades diarias.

A Dios y mi familia

Agradecimientos

Quiero agradecer primero a mi familia por el apoyo incondicional durante todos mis años de estudio. A mis viej@s por su apoyo y dedicación, y a mis herman@s por ser mis mejores amig@s.

Quiero agradecer a la profesora Sofía primero por haber accedido a ser mi profesora guía en un tema que era nuevo para mí, pero dado su conocimiento me permitieron llevar a cabo este trabajo. Muchas gracias por la buena onda, por los consejos y críticas tanto constructivas como propositivas que sin duda permitieron mejorar muchísimo esta tesis.

También quiero agradecer al profesor Marcel por su buena onda, por las críticas y propuestas realizadas a este trabajo, y sus preguntas que siempre invitaban a pensar. Todo esto en su conjunto contribuyeron mucho al desarrollo de esta tesis. Muchas gracias profesor.

Al profesor Raimundo por acceder a ser miembro de la comisión, y por haber dado sus consejos sobre ciertos puntos sobre este trabajo, los cuales pueden orientar futuros caminos a desarrollar. Muchas gracias profesor.

Por último también agradecer a las señoras Olga y Fernanda por todo el apoyo durante mi estancia en el programa. Muchas gracias a ambas.

Tabla de Contenido

1. Introducción	1
2. Revisión de literatura	3
3. Metodología para medir segregación	5
3.1. Datos	5
3.2. Medida de Segregación	6
3.3. Índice de Entropía e índices de segregación	10
4. Segregación en la RM	13
4.1. Análisis descriptivo	13
4.2. Patrones geográficos de índices de segregación	14
4.3. Comparación entre índices de segregación	17
4.4. Contribución y análisis por grupo socioeconómico	22
4.5. Evolución temporal	24
4.6. Segregación, nivel de ingresos y periferia	25
5. Determinantes de la segregación	27
5.1. Resultados regresiones	28
6. Conclusiones	31
Bibliografía	33
Anexos	36
A. Otros Resultados	36
A.1. Relación entre movilidad y segregación	36
A.2. Resultados desagregados por grupo socioeconómico	36
A.2.1. Contribución por grupo a cada zona en las distribuciones Q_{in} y Q_{out}	37
A.2.2. Estadísticos descriptivos de la segregación por grupo socio- económico	40
A.2.3. Contribución por grupo a la entropía de cada zona	42
A.2.4. Estadísticos descriptivos por grupo socioeconómico de algunas variables	50
B. Comparación de formulaciones de medidas de segregación experiencial	53
B.1. Resultados de casos evaluados	55
C. Sobre los datos	59

C.1.	Datos	59
	C.1.1. Representatividad de datos GPS	59
C.2.	Selección de datos para caracterización de individuos y zonas censales	60

Índice de Tablas

4.1.	Estadísticos descriptivos de índices usando $Q_{base} = \bar{Q}$	13
4.2.	Estadísticos descriptivos de índices usando $Q_{base} = Q_j$	14
5.1.	Resultados modelos que analizan relación entre infraestructura y segregación. .	29
5.2.	Resultados modelos que analizan relación entre ingreso como mediador de infra- estructura	29
5.3.	Resultados modelos que analizan relación entre ingreso y metro, como mediador de infraestructura	30
A.1.	Resultados de modelo A.1. Se obtuvo un $r^2 = 0.07$	36
A.2.	Estadísticos descriptivos de índices de segregación por grupo socioeconómico (con $Q_{base} = \bar{Q}$).	42
A.3.	Estadísticos descriptivos por nivel socioeconómico de % de individuos que sale de su comuna y % de individuos que se dirige a zonas de un nivel socioeconómico distinto.	49
A.4.	Estadísticos descriptivos para grupo <i>ABC1</i>	51
A.5.	Estadísticos descriptivos para grupo <i>C2</i>	52
A.6.	Estadísticos descriptivos para grupo <i>C3</i>	52
A.7.	Estadísticos descriptivos para grupo <i>D</i>	52
A.8.	Estadísticos descriptivos para grupo <i>E</i>	52
B.1.	Casos considerados para comparar distintas formas en que planteamos el índice de entropía.	55
C.1.	Estadísticos descriptivos de los datos disponibles a nivel de zona censal.	62

Índice de Ilustraciones

3.1.	Ilustración del cálculo de $Q_{t,j}^{in}$. Por ejemplo, el primer componente de la distribución para la zona censal E está dado por: $Q_{E1}^{in} = \frac{10 \cdot 0.1 + 15 \cdot 0.1 + 5 \cdot 0.3}{10 + 15 + 5} = 0.13$	8
3.2.	Ilustración del cálculo de $Q_{t,j}^{out}$. Por ejemplo, el primer componente de la distribución para la zona censal A está dado por: $Q_{A1}^{out} = \frac{25 \cdot 0.4 + 42 \cdot 0.2 + 35 \cdot 0.4 + 18 \cdot 0.2}{25 + 42 + 35 + 18} = 0.3$	9
4.1.	Mapa de la RM con resultados de la medición de los índices EI^{out} y EI^{in} para distintas entropías base ($h(\bar{Q})$ y $h(Q_j)$).	15
4.2.	Mapa de la RM con resultados de la medición de segregación residencial RI , % de altos ingresos, años de escolaridad y movilidad por zona censal.	16
4.3.	Mapa de la RM con número de visitas por zona que realizan individuos de altos y bajos ingresos (promedio). Se muestra además, cuando se eliminan aquellos individuos que se quedan en una misma zona ($i \neq j$).	17
4.4.	Segregación experiencial outbound (EI^{out}) vs inbound (EI^{in}), utilizando como distribución base para la entropía el promedio regional de ingresos ($Q_{base} = \bar{Q}$) y la distribución de ingresos de cada zona ($Q_{base} = Q_j$).	18
4.5.	Segregación experiencial de salida (EI^{out}) vs residencial (RI), utilizando como distribución base para la entropía el promedio regional de ingresos (\bar{Q}) y la distribución de ingresos de cada zona (Q_j).	19
4.6.	Histograma de entropías de distribuciones agregadas de salida y de entrada diferenciadas por nivel de ingreso, donde altos ingresos se consideran aquellas zonas con $\%ABC1 > 70\%$, y bajos ingresos $\%D + E > 70\%$ sobre el número de hogares. En línea punteada verde se observa la entropía de la distribución promedio regional.	20
4.7.	Mapa de la RM con resultados de la medición de los índices EI^{out} y EI^{in} menos la segregación residencial en las distintas zonas censales.	21
4.8.	Evolución de los índices durante el periodo analizado.	24
4.9.	Curva segregación vs ingresos.	25
4.10.	Curva segregación vs Distancia al centro en kilómetros. El gráfico superior posee un $r^2 = -0.01$ y el inferior $r^2 = 0.04$	26
A.1.	Contribución de grupos $ABC1$ y $C2$ a la distribución agregada de llegada por zona.	38
A.2.	Contribución de grupos $C3$ y D a la distribución agregada de llegada por zona.	38
A.3.	Contribución de grupo E a la distribución agregada de llegada por zona.	39
A.4.	Contribución de grupos $ABC1$ y $C2$ a la distribución agregada de salida por zona.	39
A.5.	Contribución de grupos $C3$ y D a la distribución agregada de salida por zona.	40
A.6.	Contribución de grupo E a la distribución agregada de salida por zona.	40
A.7.	Distribución de ingresos a nivel de hogar en la RM.	41
A.8.	Resultado de aplicar la función $-p \cdot \log_2(p)$ a diferentes proporciones de grupo p . En línea punteada se visualiza el máximo que se alcanza en $p = \frac{1}{e}$	43

A.9.	Distribuciones Q_{out} y Q_{in} por grupo socioeconómico, utilizando la clasificación socioeconómica para cada zona de acuerdo al punto <i>b</i>).	44
A.10.	Distribución regional del aporte de los grupos socioeconómicos $ABC1$ y $C2$ a h_{out} en cada zona censal.	45
A.11.	Distribución regional del aporte de los grupos socioeconómicos $C3$ y D a h_{out} en cada zona censal.	46
A.12.	Distribución regional del aporte del grupo socioeconómico E a h_{out} en cada zona censal.	46
A.13.	Distribución regional del aporte de los grupos socioeconómicos $ABC1$ y $C2$ a h_{in} en cada zona censal.	47
A.14.	Distribución regional del aporte de los grupos socioeconómicos $C3$ y D a h_{in} en cada zona censal.	47
A.15.	Distribución regional del aporte del grupo socioeconómico E a h_{in} en cada zona censal.	48
A.16.	Porcentaje de individuos que sale de una zona a otra zona de una comuna distinta a la suya, y porcentaje de individuos que sale de una zona de un determinado nivel de ingresos a otra zona de un nivel de ingresos distinto al propio.	50
B.2.	Segregación experiencial en su forma agregada y desagregada frente a la segregación residencial utilizando como entropía base A , la entropía de \bar{Q}	56
B.1.	Segregación experiencial en su forma agregada y desagregada frente a la segregación residencial utilizando como entropía base A , la entropía de \bar{Q}	57
B.3.	Distribución de entropías.	57
B.4.	Segregación experiencial outbound e inbound en sus forma desagregada versus agregada.	58
B.5.	Segregación experiencial inbound vs segregación outbound, agregada y desagregada.	58
B.6.	Evolución temporal de los índices planteados de forma agregada y desagregada.	59

Capítulo 1

Introducción

Las barreras entre distintos grupos sociales merman las opciones de acceder a igualdad de oportunidades (Lamont et al., 2014). A menudo, la falta de interacciones entre diferentes grupos socioeconómicos se convierte tanto en una causa como en una consecuencia de la persistente desigualdad económica (Reardon y Bischoff, 2011; Boar y Giannone, 2023).

Este fenómeno no se limita únicamente a espacios fijos, como las residencias, las escuelas y los lugares de trabajo (Massey y Denton, 1988; Cutler et al., 1999; Echenique y Fryer, 2007; Frankel y Volij, 2011), sino que las interacciones cotidianas (Athey et al., 2021) complementan y en ciertas circunstancias pueden empeorar los resultados.

Dentro de los muchos lugares en que los individuos pueden interactuar en su diario vivir (colegios, trabajo, universidad, etc) hay lugares típicos que comúnmente individuos de distintas características pueden encontrarse. Hay trabajos que han sugerido que parques en áreas urbanas y céntricas disminuyen la segregación (Gobster P., 1998), así como también, otros sugieren que los malls son un lugar de encuentro para individuos de distintos estratos socioeconómicos (Beiró et al, 2018). Por otro lado, otros han encontrado que las interacciones dentro un carril de metro son un reflejo de la segregación residencial (Swyngedouw, 2013). De estos últimos puntos mencionados¹, se vuelve interesante evaluar como segregan los individuos en lugares que visitan de manera cotidiana.

En este trabajo específicamente, definimos dos medidas de segregación experiencial. Luego utilizamos estas medidas para evaluar si lugares típicos como malls, parques y estaciones de metro son factores integradores, esto es, que permiten disminuir la segregación.

A priori uno esperaría que por su carácter público, que la segregación tanto en parques como malls disminuya, ya que son lugares donde típicamente las personas se encuentran, ya sea para consumir o simplemente pasar el rato. A su vez, el tener cerca una estación de metro puede ser un factor que facilita la integración, funcionando como aliciente para dirigirse a otros lugares lejanos a sus lugares de habitación, ya sea por ocio o por trabajo.

La importancia estudiar el rol de estos lugares como facilitadores de integración socioeconómica, es que puede orientar futuras políticas públicas. Por ejemplo, proyectar estaciones de metro cerca de parques con el fin de aumentar la integración, y/o servicios de transporte público que faciliten el acceso a parques, o replicar ciertas características de los malls en espacios públicos no necesariamente orientados al comercio.

Para implementar estas medidas obtenemos el patrón de movimiento de los individuos en la región, para lo cual usamos datos de movilidad que corresponden a registros de movilidad

¹ Notar que estos lugares son los que en este trabajo consideramos, no obstante, pueden existir otros distintos a los mencionados y que también incentiven (o desincentiven) a que individuos diversos se encuentren.

de personas donde se captura tanto el inicio como el destino usual de una persona. El poseer información tanto de inicio como destino diario de una persona, permite medir segregación de una manera distinta al enfoque usual (residencial, escolar u otro) dado que permite capturar las interacciones efectivas que un individuo posee en su diario vivir. De esta manera, podemos capturar si un individuo pese a vivir en un lugar muy segregado, se integra (o no) en sus lugares de destino habituales, disminuyendo así la segregación en su conjunto.

Los datos, corresponden a datos de teléfonos, los cuales utilizamos para identificar los lugares más frecuentes donde un individuo se encuentra. A partir de esta información, caracterizamos cuántos individuos de cada zona se encuentran con personas de otras zonas censales en la RM. Utilizando esta información y teniendo en cuenta el nivel socioeconómico basado en el hogar de cada individuo, construimos dos índices de segregación experiencial: *Segregación Experiencial de Entrada* (EI_{in}) y *Segregación Experiencial de Salida* (EI_{out}).

Los resultados muestran que EI_{in} en promedio es siempre mayor a EI_{out} , y ambas son menores a RI . Además, obtenemos que para los índices experienciales existe heterogeneidad por ingresos siendo sujetos de altos ingresos los que más segregan en cada uno de los índices, y también, dado un nivel de ingreso existen zonas segregan más que otras².

Luego, utilizando estos índices, investigamos si lugares comunes dentro de la ciudad, como parques, centros comerciales y estaciones de metro, contribuyen a que individuos de diversos estratos se encuentren en su vida diaria.

Utilizando datos de panel, encontramos que la segregación es creciente en el nivel de ingresos de una zona, por lo que individuos de mayores ingresos son los que más segregan. A su vez, ambos índices de segregación experiencial, poseen una alta correlación con el índice de segregación residencial, aunque ambos son menores a esta. Por otro lado, en cuanto a la infraestructura estudiada, el metro se posiciona como un factor integrador importante en la ciudad. Este logra integrar por sí solo distintas zonas de Santiago, con excepción de las zonas de altos ingresos. No obstante, en dichas zonas si la presencia de metro es acompañada por algún lugar público que posea ciertas características, como malls y parques, estos en conjunto logran disminuir la segregación en estas zonas que de otra manera estarían segregadas.

Nuestro trabajo aporta principalmente en dos puntos: uno es contribuir con medir segregación en distintos ámbitos en que un individuo segrega, tomando en cuenta tanto el lugar donde un individuo vive, interactúa y/o se encuentra diariamente, lo cual creemos que dan una mirada más amplia a la segregación a como se ha realizado hasta la actualidad. El segundo punto, es estudiar si lugares típicos de una ciudad, contribuyen a disminuir la segregación.

En lo que viene el documento se estructura de la siguiente manera: primero revisaremos literatura asociada a nuestro trabajo y la contribución que realizamos a ella. Luego, planteamos la metodología propuesta en este documento donde explicaremos los datos de georeferenciación que utilizamos y las medidas de segregación que proponemos. A partir de los resultados de la medición de segregación, planteamos distintos modelos econométricos para obtener la correlación de la presencia de estos lugares de interés sobre la segregación, y finalmente hablaremos sobre las conclusiones.

² Sujetos ABC1 por ejemplo, donde hay zonas que alcanzan la segregación más alta medida, mientras que otros poseen valores de segregación menores o iguales a los de otros grupos socioeconómicos.

Capítulo 2

Revision de literatura

Algunos de los primeros trabajos que realizaron una contribución tanto en la definición así como en la propuesta de una medida de segregación, comprenden los trabajos de Duncan (1955), Taeuber & Taeuber (1985), White (1983), y Massey & Denton (1988). En estos trabajos proponen y realizan un análisis de distintos índices de segregación y sus respectivas características para medir la segregación.

Una forma tradicional de medir segregación es fijar un lugar de estudio, y luego por medio de una métrica (bi-grupal o multigrupal) medir la segregación que hay en ese determinado lugar: dentro de este enfoque se puede mencionar los trabajos que han medido segregación residencial o segregación escolar. La segregación residencial consiste en medir la segregación tomando en cuenta el lugar que viven las personas y variables demográficas de estos (Massey & Denton 1988 ; Cutler et al. 1999 ; Echenique & Fryer 2007). A su vez, la segregación escolar asigna un índice de segregación a una zona, a partir de la segregación que se produce en los colegios que se encuentran dentro de ella tomando en cuenta variables demográficas de los alumnos que asisten a dichos colegios (Frankel & Volij, 2011). Nuestro trabajo contribuye a los anteriores, en considerar una medida de segregación multigrupo que varía a través del tiempo, donde por medio de los datos de movilidad, podemos capturar la segregación que se realiza por interacciones efectivas (vemos donde efectivamente se encuentra en promedio un individuo a determinada hora del día). Si bien nosotros consideramos segregación socioeconómica, las medidas propuestas son fácilmente extensibles a otras características como las que se han considerado en los trabajos mencionados en este párrafo.

Un enfoque novedoso para medir segregación, consiste en medir segregación en los espacios de interacción de individuos (Cagney et al. 2020)³. Para encontrar los espacios de interacción algunos han utilizado datos obtenidos de entrevistas (Wong & Shaw, 2011; Jones & Pebley, 2014; Browning et al, 2017) donde se les ha preguntado a dichos individuos los lugares que comúnmente frecuentan, mientras que otros han utilizado distintos métodos para obtener datos de geolocalización: datos de geolocalización por medio de Twitter (Shelton et al., 2015), solicitando a un grupo de individuos utilizar un dispositivo GPS durante un periodo de tiempo (Zhang et al., 2019), datos de GPS de movimiento de individuos recolectados en páginas web (Lee et al., 2021), o por medio por medio de GPS de celulares utilizando *pings* de localización asociados a un individuo (Athey et al., 2021; Yabe, T. 2023). Nuestro trabajo es similar a los que han utilizado geolocalización por medio de GPS, y contribuye en que consideramos una medida multigrupo que se puede comprimir (o extender) a una medida que mida segregación bi-grupal o multigrupal fácilmente según se requiera.

³ Una completa revisión bibliográfica de espacio de interacción se puede encontrar en Cagney et al. 2020

Este trabajo está también relacionado con aquellos que han medido segregación socioeconómica (Reardon & Bischoff, 2011) y en particular aquellos que han medido segregación socioeconómica en malls utilizando datos georreferenciados usando celulares (Beiró et al, 2018). Por medio de los datos de movilidad que usamos, nuestro trabajo contribuye a los anteriores en considerar una medida que permite obtener una medida diaria de la segregación y donde además capturamos interacciones efectivas de los individuos (Reardon & Bischoff, 2011), y además, nuestro trabajo mide segregación en todas las zonas de la región no limitándonos sólo a malls (Beiró et al, 2018).

Finalmente, los trabajos mencionados se han abocado a medir segregación en un ámbito más acotado, esto es, han considerado una medida bi-grupal o a su vez, solo han medido segregación en espacios específicos como lo son colegios, malls y/o lugares de habitación. Nuestro trabajo contribuye proponiendo para una determinada zona distintos enfoques para medir segregación de manera granular (medimos diariamente y a nivel de zona), con lo que se cubre un enfoque mucho más amplio de la segregación que un individuo puede experimentar en su día a día: residencial, por el hecho de vivir en una zona, así como también la segregación que puedan experimentar en sus destinos diarios. Por lo demás, si bien los índices en este trabajo miden segregación socioeconómica, estos son fácilmente extensibles a otras variables demográficas, tanto para un enfoque multi-grupal como bi-grupal.

Capítulo 3

Metodología para medir segregación

En esta sección planteamos la metodología usada en este trabajo. Comenzamos describiendo los datos de movilidad de individuos y la manera en cómo imputamos las variables demográficas a nivel de individuo. Como se verá en la siguiente sección, estos datos nos ayudan a caracterizar el movimiento que poseen las personas en las distintas zonas censales y saber cuántos individuos de una determinada zona fueron a otra zona específica. Con esto obtenemos una caracterización de los distintos encuentros que ocurren en la región durante un día.

Luego, una vez introducida la notación a utilizar en el resto del documento, planteamos las ecuaciones para obtener los índices de segregación experiencial: *Segregación Experiencial de Entrada* y *Segregación Experiencial de Salida*, los que como ya se mencionó previamente, miden la segregación de un determinado lugar a partir de los encuentros que allí ocurren y la segregación de un lugar a partir de las interacciones que poseen los individuos que allí habitan en sus lugares de destino respectivamente.

A su vez, planteamos la ecuación para obtener la segregación residencial de un lugar, esto dado que nos permitirá comparar la segregación socioeconómica que posee un individuo en su lugar de habitación, con la que posee en los distintos lugares que éste interacciona.

3.1. Datos

Los datos de movilidad utilizados corresponden a registros de movilidad de personas donde se captura tanto el inicio como el destino usual de una persona. Utilizando una georreferenciación por medio de teléfonos móviles a través de una triangulación de conexiones a antenas (Drane et al. 1998), cada móvil registra el movimiento de una persona de manera anónima. A cada móvil se le asoció una *home zone* que corresponde a la zona censal más frecuente en la que estuvo dicho móvil durante la 9-11 pm del mes de observación. A su vez, a cada aparato móvil se le asoció durante los intervalos de horario AM (10 am-1 pm) y PM (2 pm-5 pm) una *destination zone* que corresponde a la zona censal que más frecuentemente fue hallado⁴.

A partir del *home zone* y *destination zone*, caracterizamos las distintas interacciones que ocurren en las distintas zonas censales: obtenemos cuantos individuos de una determinada zona van a otra, donde tanto el nivel socioeconómico como otras variables demográficas son imputadas a un individuo a partir del *home zone* que se le asignó a dicho individuo. Luego,

⁴ Dado los horarios comprendidos en los datos de movilidad, no se realiza distinción entre los dos rangos de horarios mañana (AM) y tarde (PM), dado que ambos comprenden horarios de jornada laboral por lo que no existen diferencias importantes en la medición de segregación.

los índices de segregación se plantean sobre las interacciones que se producen en las distintas zonas. De esto, construimos la matriz de flujo f_{ijt} de individuos que va de la zona censal i a la zona j un día t (Carranza et al., 2020).

Para la construcción de f_{ijt} , consideramos que individuos de una misma zona censal pueden permanecer en su misma zona, y por lo tanto, pueden interactuar entre ellos y/o con individuos que van a ese lugar⁵. Además, consideramos el supuesto de que individuos que se encuentran en una determinada zona efectivamente interactúan. Esto es un supuesto un tanto fuerte, dado el tamaño promedio de las zonas censales. Para nosotros dos sujetos ubicados a una distancia de dos calles, efectivamente interactúan.

Otro punto relevante, es que del total de zonas que venían en lo datos (1867) se consideran 1800, dado que son las que cumplen con el criterio de exclusión por representatividad (Ver Anexo C), donde consideramos un mínimo de personas que deben venir representadas por cada zona censal. Si dicho mínimo no se cumple, entonces la zona no se considera.

3.2. Medida de Segregación

En la literatura existen variados índices de segregación, no obstante, el índice de entropía corresponde a un índice que cumple con varios criterios necesarios para ser un índice de segregación⁶ (Reardon 2002; Frankel & Volij, 2011). Además, al ser un enfoque basado en distribuciones, puede ser fácilmente extensible a otras variables demográficas, tanto para enfoques multigrupales como bi-grupales (cuando se trabaja con segregación racial por ejemplo).

Antes de definir los índices, a continuación describimos la notación a utilizar el resto del documento y las distribuciones *Distribución Agregada de Llegada* y la *Distribución Agregada de Encuentro* que nos serán útiles para definir los índices. De esta manera, la notación a utilizar es la siguiente:

- Σ : Espacio de características sociales. En el caso del nivel socioeconómico, $\Sigma = \{ABC1, C2, C3, D, E\}$.
- J el conjunto de zonas censales (por simplicidad de notación, también la usamos para denotar la cardinalidad del conjunto)
- $Q_i \in \Delta(\Sigma)$: Distribución de características de las personas con *home zone* en la zona censal i . Para el caso de nivel socioeconómico, Q_i corresponde a un vector en que cada elemento indica la proporción de personas perteneciente a cada nivel.
- $\bar{Q} \in \Delta(\Sigma)$: Distribución de características a nivel de población ($P \in \Delta(\Sigma)$). Esto puede ser la región o comuna dependiendo de la comparación que sea pertinente.
- F_t es una $J \times J$ matriz con los flujos entre las zonas censales para un determinado tiempo t donde el elemento (i, j) (los cuales denotamos por f_{ijt}) corresponde al número de personas de la zona censal i que visita j al tiempo t .

⁵ En anexos se encuentran los resultados obtenidos al considerar solo aquellos movimientos en que individuos se mueven fuera de su zona censal, esto es, para un día t elementos de la diagonal de f_{ij} son imputados por cero

⁶ Ver anexo B

Luego, definimos la *Distribución Agregada de Llegada* como la distribución promedio ponderada de individuos que llega a una zona en un día⁷. Esta distribución captura la distribución de ingresos representativa de individuos que existe en cada zona en un día, por lo que si un sujeto se dirige a dicha zona, interactúa con sujetos cuya distribución de ingresos es la *Distribución Agregada de Llegada*. Su ecuación es:

$$Q_{t,j}^{in} = \frac{\sum_{k \in J} f_{kjt} \cdot Q_k}{\sum_{k \in J} f_{kjt}} \quad (3.1)$$

donde f_{kjt} corresponde al flujo de individuos de va de k a j en el día t , y Q_k es la distribución de ingresos (censal) residencial de k (ver fig. 3.1 para un ejemplo).

A su vez, definimos *Distribución Agregada de Encuentro* como la distribución agregada de individuos con los que los individuos de una zona se encuentran en un día. Esta distribución la definimos a partir de la *Distribución Agregada de Llegada*, y corresponde a la distribución promedio representativa con la que sujetos de una determinada zona interactúan en un día. Su ecuación es:

$$Q_{t,j}^{out} = \frac{\sum_{k \in J} f_{jkt} \cdot Q_{t,k}^{in}}{\sum_{k \in J} f_{jkt}} \quad (3.2)$$

donde $Q_{t,k}^{in}$ corresponde a la *Distribución Agregada de Llegada* de la zona k , y f_{jkt} al flujo de individuos de va de j a k el día t (ver fig. 3.2 para un ejemplo).

⁷ Para cada día, estamos considerando que el flujo viene dado por el máximo del horario AM y PM, esto es, $f_{ijt} = \text{Max}(f_{ijt}^{AM}, f_{ijt}^{PM})$

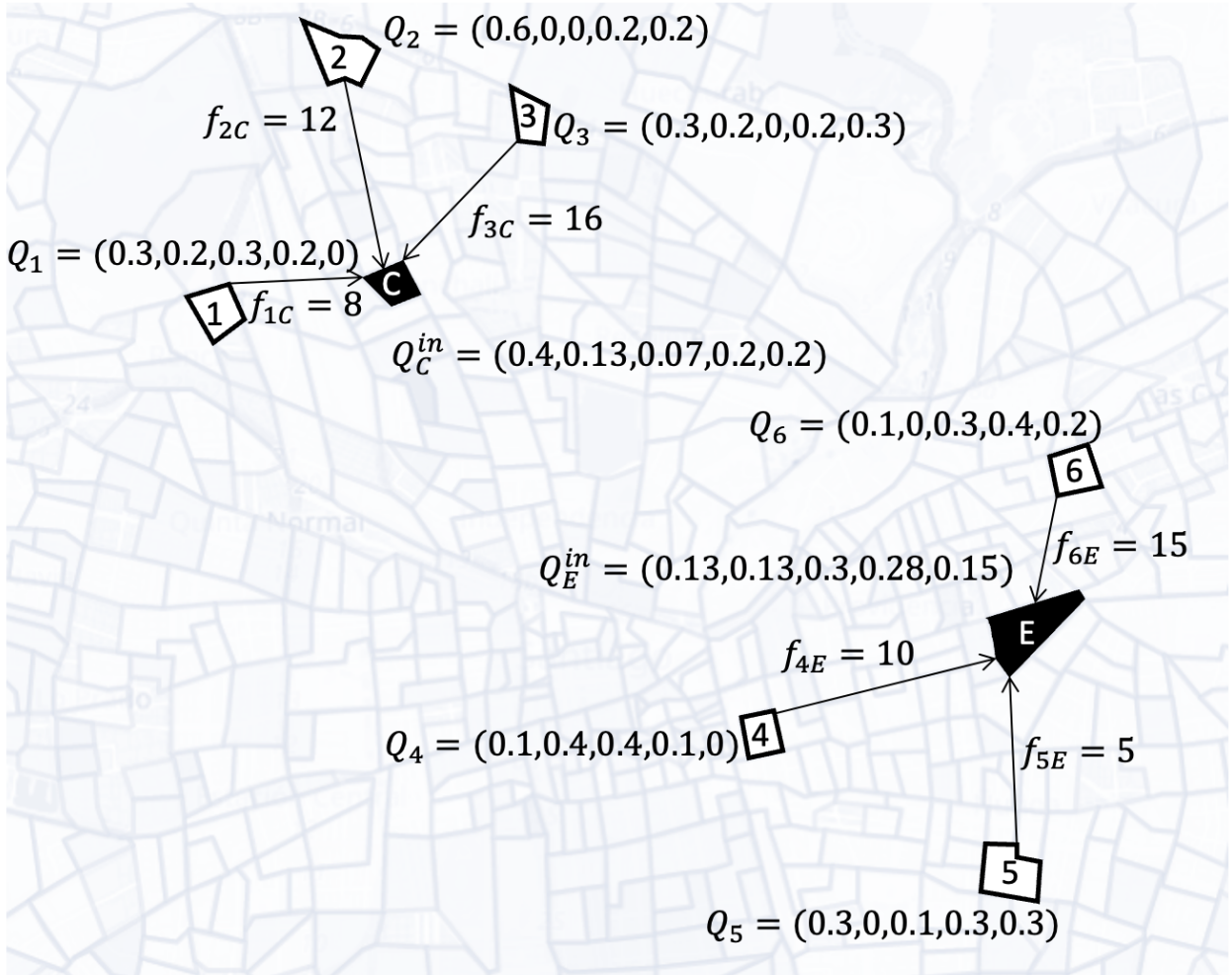


Figura 3.1: Ilustración del cálculo de $Q_{t,j}^{in}$. Por ejemplo, el primer componente de la distribución para la zona censal E está dado por: $Q_{E1}^{in} = \frac{10 \cdot 0.1 + 15 \cdot 0.1 + 5 \cdot 0.3}{10 + 15 + 5} = 0.13$.

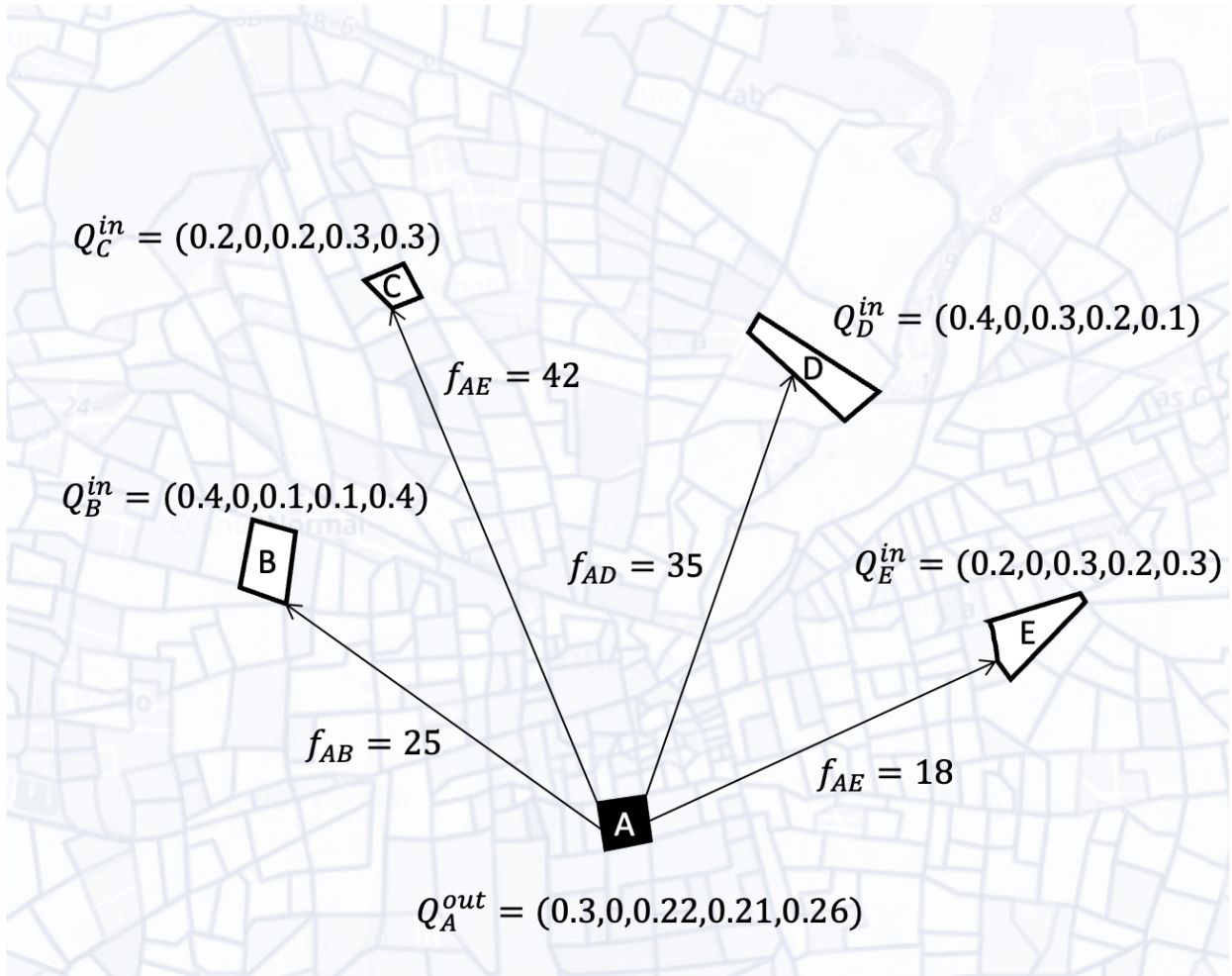


Figura 3.2: Ilustración del cálculo de $Q_{t,j}^{out}$. Por ejemplo, el primer componente de la distribución para la zona censal A está dado por: $Q_{A1}^{out} = \frac{25 \cdot 0.4 + 42 \cdot 0.2 + 35 \cdot 0.4 + 18 \cdot 0.2}{25 + 42 + 35 + 18} = 0.3$.

3.3. Índice de Entropía e índices de segregación

La entropía es una medida que proviene de la teoría de la información, la cual mide el nivel promedio de incerteza de un evento. La entropía es mayor, si el proceso es más incierto, y es menor si hay más certeza sobre una posible salida de dicho evento (Cover, 1991). Para una unidad de análisis i (en nuestro caso zona censal) se define como

$$h(Q_i) = - \sum_{g \in \Sigma} Q_{ig} \cdot \log_2(Q_{ig}) \quad (3.3)$$

donde Q_{ig} corresponde a la proporción del grupo g en la zona i y Σ al espacio de características sociales. La entropía toma un valor máximo cuando la distribución en un área es uniforme, y un valor mínimo, cuando posee un solo grupo. De esta manera, un valor elevado de h indica más diversidad, mientras que un valor más pequeño más homogeneidad.

La entropía ha sido usada para definir el *índice de entropía* (Thiel, 1985), el cual se ha demostrado que posee ciertas propiedades adecuadas para ser un índice de segregación (Reardon, 2002; Frankel & Volij, 2011). No obstante, la forma en que este se debe plantear posee cierta estructura que en este trabajo no son adecuadas (ver Anexo B).

No obstante, más allá de las propiedades, el utilizar la entropía como componente de un índice de segregación, nos permite evaluar fácilmente la diversidad de la distribución que se está evaluando, y en base a una distribución *ideal* con la que se desea comparar, obtener un índice de segregación: segregación en nuestro caso será que tanto se aleja la diversidad de un grupo de personas frente a una diversidad base considerada. Mientras más lejana está la diversidad de dicho grupo de la diversidad de la distribución *ideal*, diremos que un lugar es más segregado, mientras que, si dichas diversidades son parecidas, diremos que un lugar es menos segregado.

De esta manera, a partir de la entropía definimos nuestras medidas de segregación:

- Definimos **Segregación Experiencial de Salida**⁸ (EI_{out}) para una zona j como la segregación que personas experimentan en los lugares que comunmente frecuentan como:

$$EI_{tj}^{out} = h(Q_{base}) - h(Q_{tj}^{out}) \quad (3.4)$$

donde $h(Q_{base})$ corresponde a la entropía de una distribución base o ideal con la se quiere comparar, y $h(Q_{tj}^{out})$ corresponde a la entropía de la *Distribución Agregada de Encuentro* para la zona j el día t .

- Definimos **Segregación Experiencial de Entrada**⁹ (EI_{in}) para una zona j el día t , como la segregación que experimentan individuos que van a esa zona.

$$EI_{tj}^{in} = h(Q_{base}) - h(Q_{tj}^{in}) \quad (3.5)$$

donde $h(Q_{base})$ corresponde a la entropía de una distribución base o ideal con la se quiere comparar, y $h(Q_{tj}^{in})$ corresponde a la entropía de la *Distribución Agregada de Llegada* para la zona j el día t .

⁸ En algunas partes del documento hacemos referencia a esta medida simplemente como segregación de salida.

⁹ En algunas partes del documento hacemos referencia a esta medida simplemente como segregación de entrada.

- Definimos **Segregación Residencial** (RI) para una zona j como la segregación que los sujetos experimentan al vivir en dicha zona,

$$RI_j = h(Q_{base}) - h(Q_j) \quad (3.6)$$

donde $h(Q_{base})$ corresponde a la entropía de la distribución base considerada, y $h(Q_j)$ a la entropía de la distribución residencial de j .

Para las distribuciones base Q_{base} consideramos dos casos: 1) tomamos la distribución de ingresos promedio regional ($Q_{base} = \bar{Q}$), y 2) la distribución de ingresos propia de cada zona ($Q_{base} = Q_j$). Con el primer caso queremos determinar que tanto se segregan los individuos de las distintas zonas respecto del promedio regional, y con el segundo que tanto segregan respecto a sí mismos¹⁰.

A su vez, las ecuaciones para $h(Q_{tj}^{in})$ y $h(Q_{tj}^{out})$ están descritas sobre una distribución ponderada, no obstante, estas se pueden plantear de una manera desagregada (Thiel, 1985). En este documento las planteamos de una manera agregada por los siguientes motivos: 1) dado que ayuda a la simplicidad de notación, 2) por que al plantear de manera desagregada el índice de *Segregación de Entrada* no logra capturar cambios de diversidad en el tiempo. Además, experimentalmente comprobamos que los resultados son similares (ver Anexo B).

En cuanto a las distintas medidas de segregación, tenemos que tanto RI , EI_{out} como EI_{in} nos permiten caracterizar los distintos tipos de segregación que individuos de una zona podrían experimentar: la segregación que experimentan por habitar en ese lugar (RI), por interactuar con individuos que van a esa zona (EI_{in}) y la segregación que experimentan en sus lugares de destino (EI_{out}).

A partir de esto podemos analizar si hay zonas que poseen ciertas características que los hace ser lugares de encuentro entre sujetos de distintos sectores. Por ejemplo EI_{in} nos será útil para realizar el análisis sobre si parques y malls son lugares de encuentro, o si el tener presencia de metros cerca de ellos es un aliciente al encuentro de sujetos diversos.

Al mismo tiempo, por medio de EI_{out} podemos analizar si el vivir cerca de un metro, es una facilitador para personas que habitan cerca de estaciones se integren más, o en su defecto, personas que habitan lejos de estaciones, se integren menos.

Por último el obtener RI nos permite realizar un análisis de integración. Esto es que si bien los individuos pueden vivir en lugares altamente segregados, en sus lugares de destino puede que se integren, y/o que dicho lugar sea un lugar donde personas diversas se integren, transformando así a ese lugar como un lugar de encuentro.

Con el propósito de comparar los distintos índices entre sí, y a su vez, evaluar las diferencias en las distintas zonas dentro de un mismo índice (si existe heterogeneidad u otro) obtenemos los siguientes puntos:

- Obtenemos estadísticos descriptivos de cada métrica.
- Una caracterización espacial de los distintos índices para visualizar patrones en las distintas zonas de la RM.

¹⁰ En el primer caso 1), una zona será menos segregada que otra si su entropía de la distribución *Distribución Agregada de Llegada*, *Distribución Agregada de Encuentro* o residencial (según corresponda) esta más cercana a la entropía de la distribución regional. Mientras que en el segundo caso 2), una zona será menos segregada que otra si la entropía propia esta más alejada de la entropía de las distribuciones *Distribución Agregada de Llegada* y *Distribución Agregada de Encuentro*.

- La evolución temporal de los índices EI_{out} y EI_{in} .
- La relación entre la *Segregación de Salida*, *Segregación de Entrada* y residencial.
- Desagregamos la entropía medida por zona, a modo de obtener cuanto contribuye cada grupo socioeconómico a la segregación en las distintas zonas censales.

Capítulo 4

Segregación en la RM

4.1. Análisis descriptivo

En esta sección analizamos los resultados de la segregación medida en la RM de acuerdo a los indicado en la sección anterior. De esta manera, de los estadísticos descriptivos (ver tabla 4.1) para cuando se toma como base la distribución regional, se tiene que la *Segregación de Salida* tiende a ser menor que la *Segregación de Entrada*, siendo en promedio la mitad de esta última. A su vez, la segregación residencial en promedio es el doble que la *Segregación de Entrada* y cuatro veces mayor a la *Segregación de Salida*. Esto sugiere, que efectivamente los individuos en promedio se integran más en sus lugares de destino que en sus lugares de habitación

Tabla 4.1: Estadísticos descriptivos de índices usando $Q_{base} = \bar{Q}$.

Índice	Media	std	min	25 %	50 %	75 %	max
Poblacional							
EI_{out}	0.04	0.05	-0.02	0.01	0.03	0.06	0.30
EI_{in}	0.08	0.07	-0.02	0.03	0.06	0.12	0.42
RI	0.16	0.11	-0.05	0.08	0.14	0.21	0.65
Altos ingresos							
EI_{out}	0.16	0.05	0.04	0.14	0.16	0.18	0.30
EI_{in}	0.24	0.06	0.08	0.20	0.23	0.27	0.42
RI	0.44	0.06	0.34	0.40	0.43	0.48	0.60
Bajos ingresos							
EI_{out}	0.03	0.02	-0.02	0.01	0.02	0.04	0.17
EI_{in}	0.06	0.05	-0.02	0.03	0.06	0.10	0.21
RI	0.13	0.08	-0.05	0.06	0.12	0.18	0.65

Si se obtiene lo anterior pero separando por nivel de ingreso (alto ingresos y bajos ingresos), tenemos que dichas proporciones se mantienen relativamente constantes, siendo la *Segregación de Salida* la menor en cada caso. No obstante, si se compara entre grupos socioeconómicos, surge una diferencia importante, donde se tiene que grupos de altos ingresos segregan casi cuatro veces más en cada uno de los índices, esto es, grupos de personas de altos ingresos segregan cuatro veces más tanto residencial, como experiencialmente.

A su vez, los valores de la desviación estándar, así como los valores mínimo y máximo de los índices de segregación, sugieren la existencia de heterogeneidad dentro de un mismo

grupo socioeconómico. En la siguiente sección, estudiaremos esta heterogeneidad por medio de regresiones de datos de panel que nos permitan evaluar si parte de esta heterogeneidad es explicada por la presencia de malls, estaciones de metro o áreas verdes.

En cuanto a los resultados para cuando la entropía base se obtiene sobre la distribución de cada zona, se tiene que la menor segregación sigue siendo la de salida por sobre la de entrada. Esto quiere decir que la distancia entre la diversidad de los sujetos con los que individuos se encuentran y su diversidad propia, es mayor que la diversidad que existe en los lugares donde estos habitan. Un punto que destaca es que la segregación es aún menor para grupos de altos ingresos por sobre bajos ingresos. Esto se podría mal interpretar como que aquellos grupos de altos ingresos se integran más que aquellos de bajos ingresos, pero la razón de lo anterior es que simplemente la distancia relativa es mayor para las zonas de altos ingresos: su diversidad propia ($h(Q_j)$) es lo suficientemente homogénea como para estar lo suficientemente lejos de la distribución con la que se encuentran.

Tabla 4.2: Estadísticos descriptivos de índices usando $Q_{base} = Q_j$.

Índice	Media	std	min	25 %	50 %	75 %	max
Poblacional							
EI_{out}	-0.11	0.08	-0.66	-0.16	-0.11	-0.06	0.07
EI_{in}	-0.08	0.06	-0.64	-0.11	-0.07	-0.03	0.06
Altos ingresos							
EI_{out}	-0.27	0.05	-0.39	-0.30	-0.27	-0.24	-0.20
EI_{in}	-0.20	0.05	-0.33	-0.23	-0.20	-0.16	-0.12
Bajos ingresos							
EI_{out}	-0.10	0.06	-0.66	-0.14	-0.09	-0.05	0.06
EI_{in}	-0.06	0.05	-0.64	-0.08	-0.05	-0.03	0.05

4.2. Patrones geográficos de índices de segregación

Si observamos patrones geográficos en la RM, tenemos (figuras 4.1 y 4.2) que existe una clara correlación entre ingresos y segregación, esto es que personas con mayores ingresos son las que más segregan, tanto en sus interacciones (EI_{out}) como en las respectivas zonas de altos ingresos (EI_{in}). Esto sugiere que personas que habitan en lugares de altos ingresos, suelen moverse principalmente en lugares de altos ingresos y relacionarse con personas que, a su vez, poseen altos ingresos.

Segregación experiencial en la RM

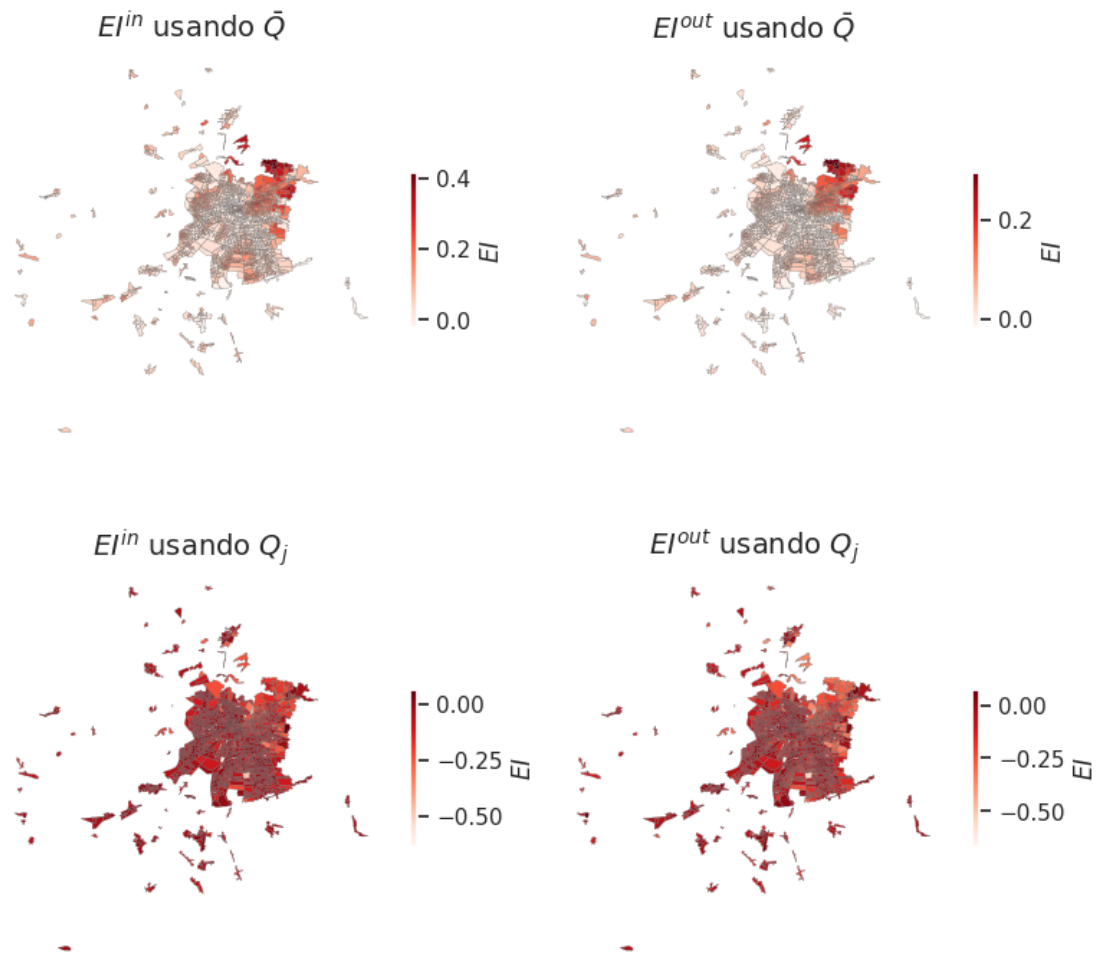


Figura 4.1: Mapa de la RM con resultados de la medición de los índices EI^{out} y EI^{in} para distintas entropías base ($h(\bar{Q})$ y $h(Q_j)$).

RI y % de altos ingresos por zona

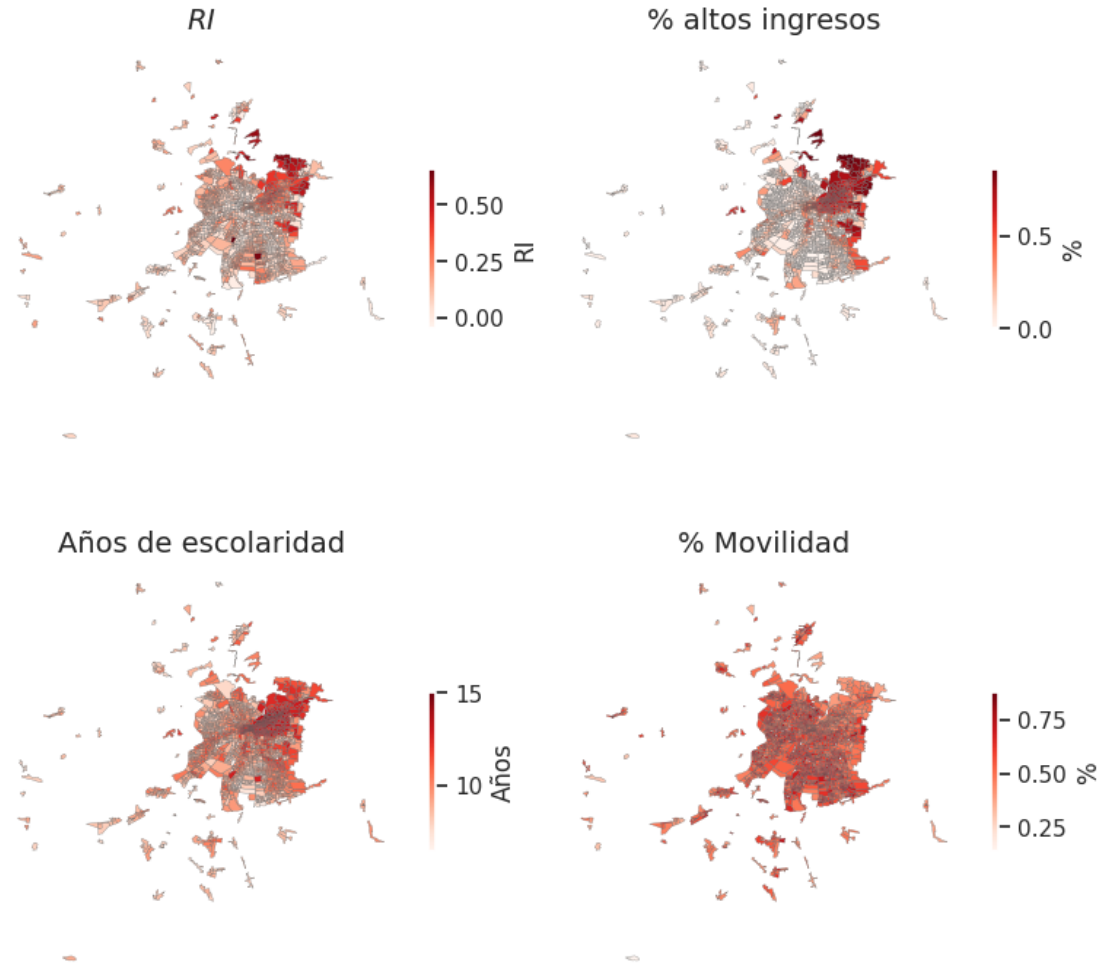


Figura 4.2: Mapa de la RM con resultados de la medición de segregación residencial RI , % de altos ingresos, años de escolaridad y movilidad por zona censal.

Del nivel de movilidad (Carranza et al, 2020)¹¹, se puede apreciar (fig. 4.2) que las zonas de altos ingresos tienden a moverse menos que aquellas personas de bajos ingresos. De lo anterior, tenemos que personas de altos ingresos segregan más y a su vez se mueven menos a zonas donde pueden interactuar con otro tipo de individuo, quedándose dentro de sectores de su mismo nivel socioeconómico.

Esto se sustenta también en la figura 4.3, donde al diferenciar entre las zonas que visitan en promedio individuos de altos ingresos de aquellas que visitan individuos de bajos ingresos, se tiene que todos tienden en promedio a visitar lugares de su mismo estrato socioeconómico, siendo el caso más dramático para las personas de altos ingresos, los cuales salen muy poco a otras zonas de otros niveles¹². Esto se puede deber a múltiples motivos, como falta de

¹¹ La movilidad se define a partir de la matriz f_{ijt} donde se obtiene el promedio de la fracción de individuos que salen de la zona i a otras zonas distintas de i ($i \neq j$) un día t .

¹² Esto es para ambos caso que se consideran, en que en uno alguien puede quedarse en un zona e interactuar con gente que allí va f_{ijt} o cuando consideramos solo el movimiento fuera de la zona $f_{ijt}^{i \neq j}$.

incentivos para hacerlo (por poseer todos los servicios necesarios dentro de su comuna y/o zona), trabajar dentro de estas zonas y/o simplemente por discriminación. No obstante, el análisis de estos motivos está fuera del alcance de este trabajo.

Visitas por grupos de análisis

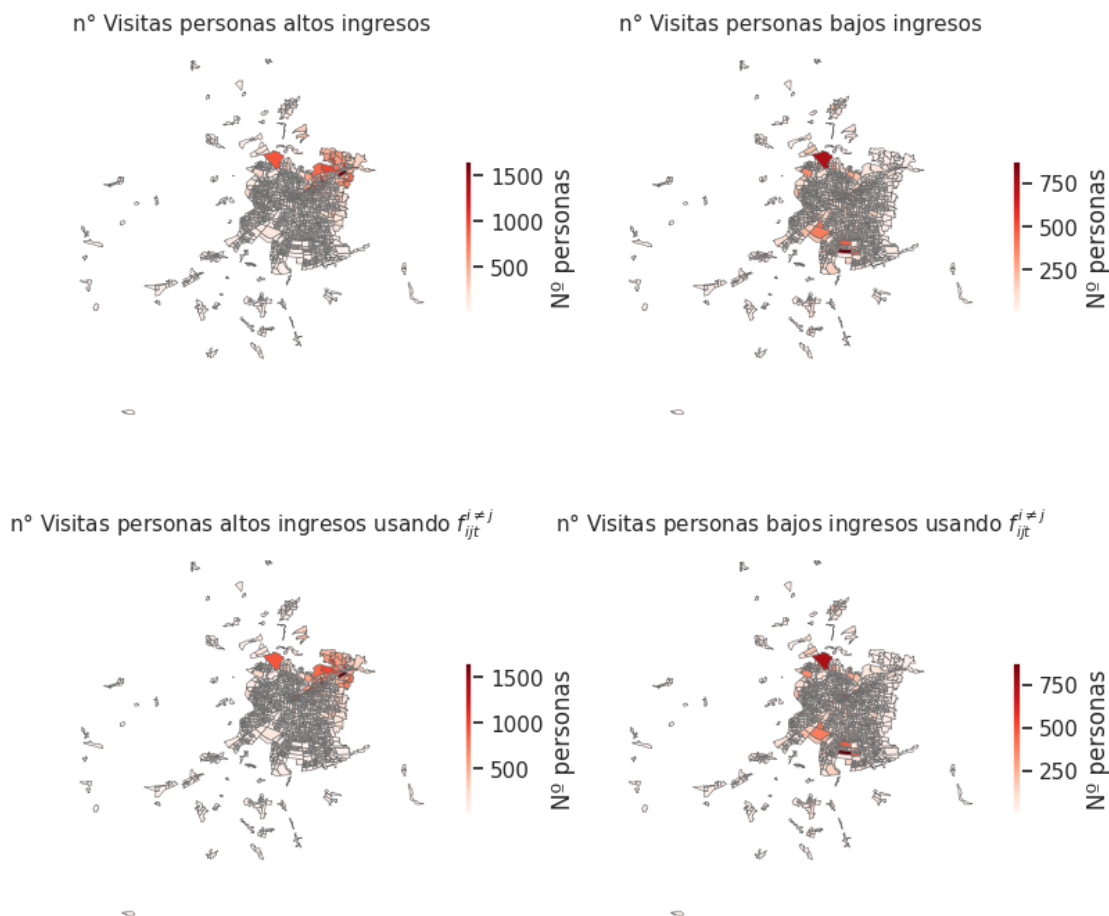


Figura 4.3: Mapa de la RM con número de visitas por zona que realizan individuos de altos y bajos ingresos (promedio). Se muestra además, cuando se eliminan aquellos individuos que se quedan en una misma zona ($i \neq j$).

4.3. Comparación entre índices de segregación

Mencionamos previamente que en promedio EI_{out} es menor a EI_{in} , tanto en el agregado poblacional como en la diferenciación por nivel de ingresos (altos y bajos ingresos). Lo anterior, queda reflejado también en la figura 4.4 donde independiente de la distribución base considerada ($Q_{base} = \bar{Q}$ o $Q_{base} = Q_j$) la *Segregación de Entrada* prácticamente es siempre mayor a la *Segregación de Salida*, por lo que para un individuo determinado el permanecer en la zona en que habita debería ser motivo para experimentar una mayor segregación (que interactúe con sujetos más homogéneos o más parecidos a el mismo) que si sale de su lugar de habitación (lo cual es algo lógico).

EI^{in} vs EI^{out} por zona censal

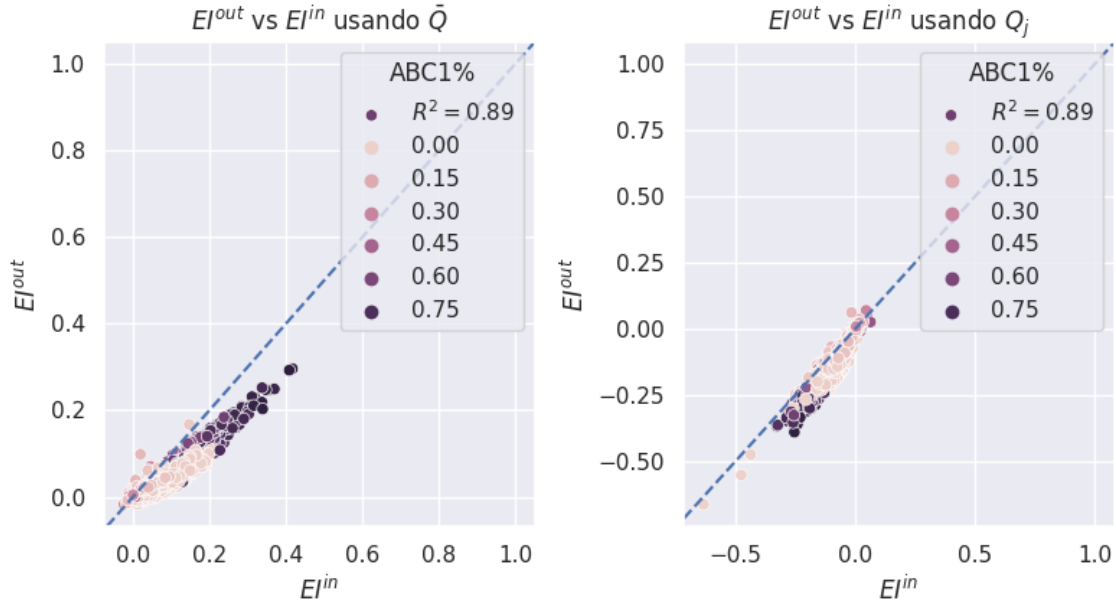


Figura 4.4: Segregación experiencial outbound (EI^{out}) vs inbound (EI^{in}), utilizando como distribución base para la entropía el promedio regional de ingresos ($Q_{base} = \bar{Q}$) y la distribución de ingresos de cada zona ($Q_{base} = Q_j$).

El hecho de que los índices de segregación experiencial de salida y entrada se vean similares se debe a lo ya mencionado, de que si bien hay personas de sectores que interactúan con personas de otras zonas, las interacciones que se experimentan son generalmente con personas parecidas a ellos mismos, lo cual es aún más marcado para zonas de altos ingresos, donde a partir de los resultados, mayoritariamente se produce menos movilidad a una zona censal externa, y cuando se produce es para moverse a otra zona censal muy parecida a la propia (alguien de Vitacura que viaja a su trabajo en Las Condes por ejemplo), lo cual ocasiona que estas zonas sean las que más segregan.

En cuanto a la relación de la segregación experiencial y residencial, podemos observar que la segregación residencial está altamente correlacionada con la experiencial, ya sea EI_{out} o EI_{in} . La segregación residencial salvo algunos casos, es siempre mayor a la experiencial (fig. 4.5), lo que significa que los individuos tienden a integrarse a lo largo de su día, encontrándose con sujetos más heterogéneos que a los que encuentran en donde habitan.

De la figura 4.5, observamos además, que dependiendo con la distribución base con la se compare las pendientes son diferentes: al considerar la entropía de \bar{Q} como base, tenemos que la pendiente es positiva, y que sectores de altos ingresos segregan más. Lo anterior es porque la diferencia entre las entropías de $h_{out,in}$ y $h(\bar{Q})$ es mayor para sectores de altos ingresos (la entropía de sectores de altos ingresos es más lejana para sectores de altos ingresos). Luego, como los sectores de altos ingresos son los que más segregan residencialmente, da como resultado una relación positiva entre la segregación residencial y experiencial cuando la distribución base es \bar{Q} .

Por otro lado, cuando la distribución base es la distribución propia de cada zona Q_j , la relación se invierte (ver fig. 4.5) dado que los sectores de altos ingresos son los que más segregan residencialmente, y al mismo tiempo, su distancia entre la entropía de la distribución

propia ($h(Q_j)$) y las entropías $h_{out,in}$ es mayor a la que tienen sectores de bajos ingresos. Como se verá más adelante, esto se debe a que la distancia relativa entre la segregación propia con la distribución de personas que las que interactúan, es mayor para individuos de altos ingresos: las distribuciones propias Q_j son lo suficientemente sesgadas a su mismo grupo, para hacer que la distancia sea mayor.

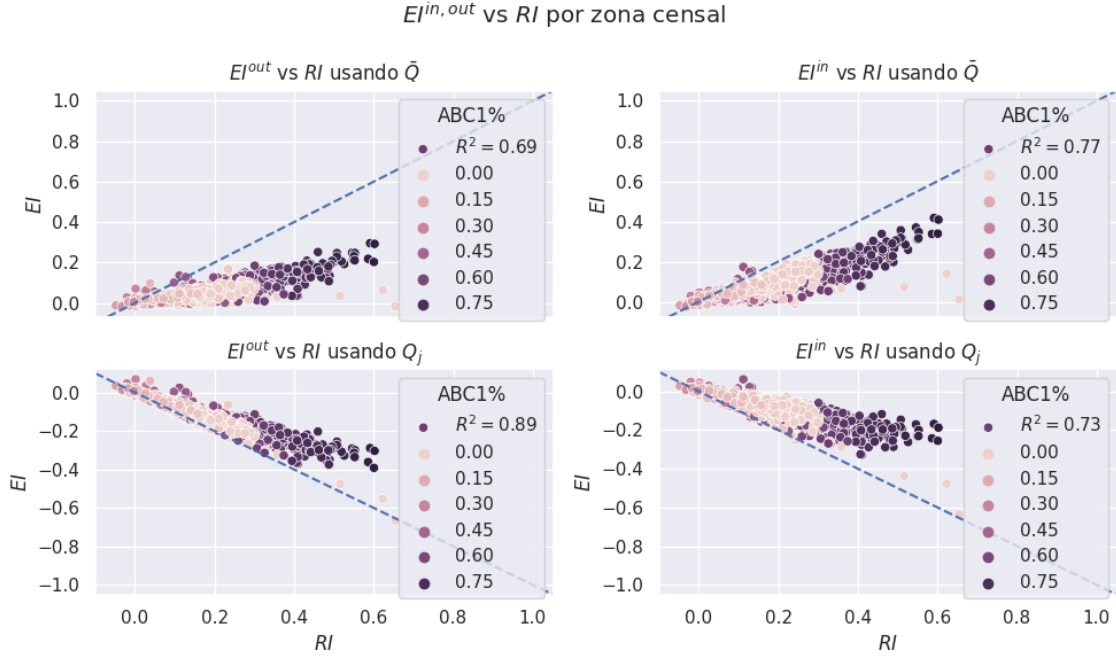


Figura 4.5: Segregación experiencial de salida (EI^{out}) vs residencial (RI), utilizando como distribución base para la entropía el promedio regional de ingresos (\bar{Q}) y la distribución de ingresos de cada zona (Q_j).

Independiente de la distribución base con la que se compare, individuos de mayores ingresos que poseen una elevada segregación residencial, son los que más segregan experiencialmente. Si se toma como base la distribución promedio regional (\bar{Q}) la distancia entre la entropía regional y las entropías de las distribuciones *Agregada de Llegada* y *Agregada de Encuentro*, es mayor para individuos de altos ingresos (fig. 4.6)¹³.

Ahora, cuando se toma como base la distribución propia (Q_j), a priori uno esperaría que la distancia entre la entropía propia y las entropías de las distribuciones *Agregada de Llegada* y *Agregada de Encuentro*, fuera menor para individuos de mayores ingresos, no obstante, la distancia promedio es mayor para individuos de altos ingresos (ver fig. 4.6)¹⁴.

Lo anterior se debe a que la diferencia es relativa a sí mismo: en los sectores de altos ingresos, su distribución residencial (Q_j) es lo suficientemente sesgada hacia su mismo grupo para que su entropía ($h(Q_j)$) este muy por debajo de las distribuciones experienciales de individuos de altos ingresos, mientras que en sectores de bajos ingresos su distribución resi-

¹³ Sobre las distribuciones de la figura 4.6, se obtiene en promedio para altos ingresos $h_{out} - h(\bar{Q}) = -0.38$, mientras que para bajos ingresos $h_{out} - h(\bar{Q}) = -0.1$. A su vez, para altos ingresos la distancia promedio entre $h_{in} - h(\bar{Q}) = -0.55$ y para bajos $h_{in} - h(\bar{Q}) = -0.24$

¹⁴ Sobre las distribuciones de la figura 4.6, se obtiene en promedio para altos ingresos $h_{out} - h(Q_j) = 0.64$, mientras que para bajos ingresos $h_{out} - h(Q_j) = 0.42$. A su vez, para altos ingresos la distancia promedio entre $h_{in} - h(Q_j) = 0.47$ y para bajos $h_{in} - h(Q_j) = 0.28$

dencial es más homogénea, lo cual hace que en promedio la distribución de $h(Q_j)$ no están tan lejana de la distribuciones experiencial (ver fig.4.6).

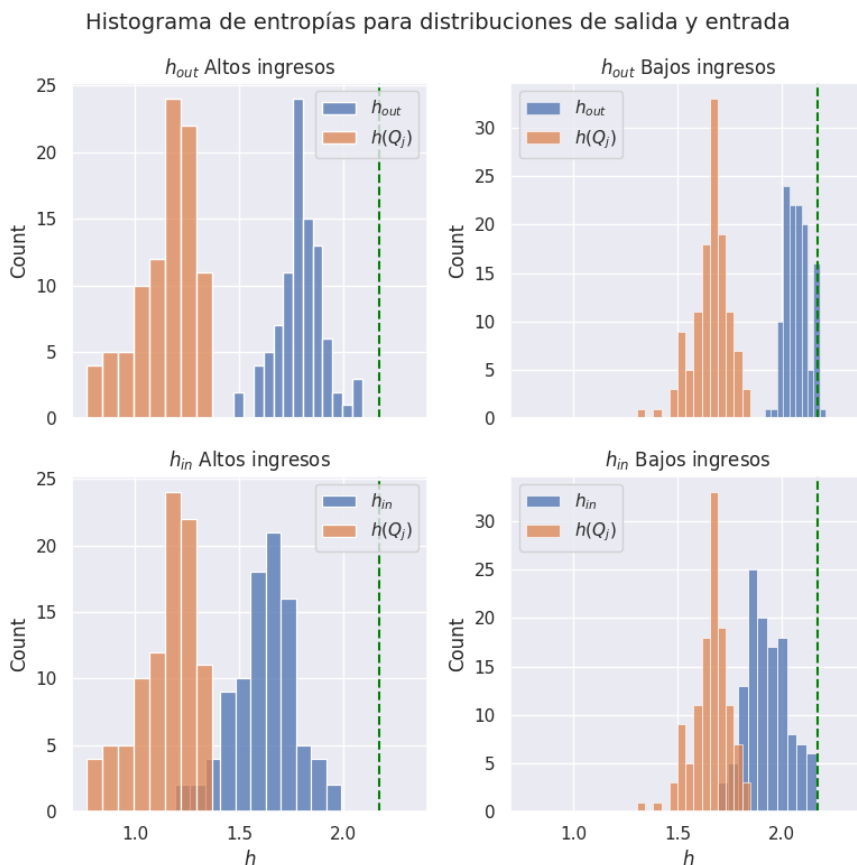


Figura 4.6: Histograma de entropías de distribuciones agregadas de salida y de entrada diferenciadas por nivel de ingreso, donde altos ingresos se consideran aquellas zonas con $\%ABC1 > 70\%$, y bajos ingresos $\%D + E > 70\%$ sobre el número de hogares. En línea punteada verde se observa la entropía de la distribución promedio regional.

Una medida interesante y que está relacionado con esto último, es comparar como varia la segregación experiencial en relación a la segregación residencial, esto es, tomar $EI_{in,out} - RI$. Notar que esto a priori nos da una medida de *integración* de las distintas zonas, dado que podemos ver cuánto se integra una persona en relación a su segregación residencial. A priori un mayor valor de esta diferencia, debería reflejar una mayor integración, no obstante, como veremos a continuación esto es relativo y es equivalente a tomar el análisis que acabamos de hacer sobre la figura 4.6.

Si observamos los gráficos que reflejan *integración* (fig. 4.7), aquellas zonas que poseen una mayor diferencia entre su segregación residencial y la experiencial, son sectores de altos ingresos. Esto quiere decir que dada su segregación residencial, estas zonas son las que más se integran. No obstante, lo anterior se debe a que en estas zonas su segregación residencial es tan elevada, que ocasiona que la diferencia en cada caso sea mayor.

Segregación residencial menos experiencial $RI - EI$

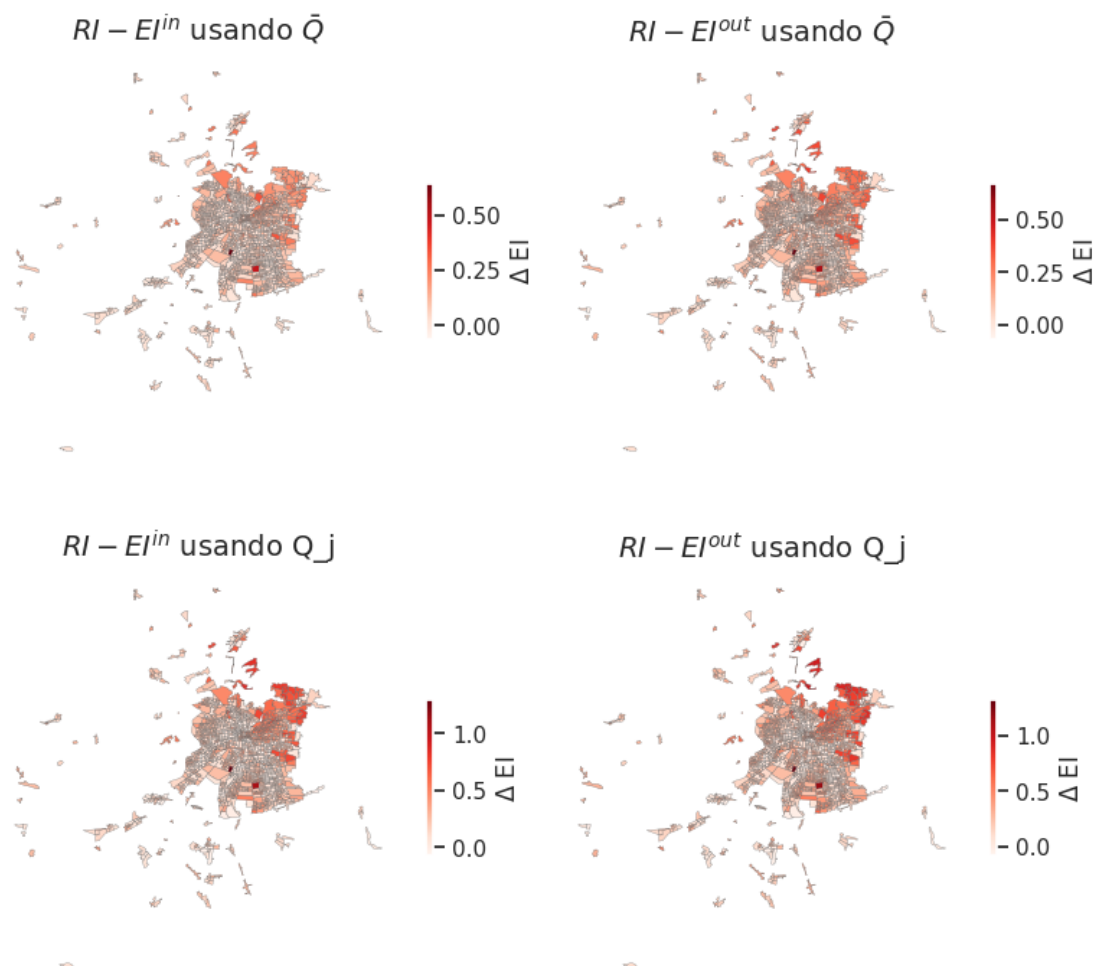


Figura 4.7: Mapa de la RM con resultados de la medición de los índices EI^{out} y EI^{in} menos la segregación residencial en las distintas zonas censales.

Recordar de la sección anterior, y simplificando notación tenemos que $EI - RI$ equivale a¹⁵

$$EI - RI = h(\bar{Q}) - h^{out,in} - (h(\bar{Q}) - h^{res}) = h^{res} - h^{out,in} \quad (4.1)$$

De esta manera, la diferencia entre las segregaciones residencial y experiencial, es la diferencia entre las entropías de salida (o de entrada) y la entropía de las distribuciones residenciales, donde ya vimos y explicamos previamente la razón por la cual dicha diferencia es mayor en sectores de altos ingresos.

¹⁵ Esto es para el caso en que la distribución base es constante e igual a \bar{Q} . De esto, en este caso se tiene $EI - RI = h^{Q_j} - h^{out,in} - (h^{\bar{Q}} - h^{res}) = (h^{res} - h^{out,in}) + (h^{res} - h^{\bar{Q}}) = \Delta_2 + (h^{res} - h^{\bar{Q}})$, ya que $h^{Q_j} = h^{res}$. De esto, y de la figura ?? podemos ver que en este caso también sectores de altos ingresos poseen una diferencia mayor.

4.4. Contribución y análisis por grupo socioeconómico

Un punto importante de nuestra medida de segregación es que dada la concavidad de la entropía, asignamos un mismo nivel de segregación a individuos de distintos estratos socioeconómicos que poseen distribuciones de encuentro (Q_{out} o Q_{in} según corresponda) similares (igualmente sesgadas), independiente del nivel socioeconómico del grupo en referencia.

De esta manera, no consideramos casos donde por ejemplo uno quisiera medir como segregan los individuos de altos ingresos tomándolos como grupo de referencia (similar al enfoque racial usado en Athey et al, 2020), dado que se puede argumentar que este grupo es el que realmente tiene la oportunidad real de elegir con quien relacionarse, donde trabajar o que lugares visitar, mientras que grupos de escasos recursos sus oportunidades de elegir son escasas.

Para tener una idea de esto y entender como segregan los distintos grupos socioeconómicos, realizamos tres ejercicios: *i*) Graficamos la contribución por grupo socioeconómico a cada zona en las distribuciones Q_{in} y Q_{out} (ver Anexo A.2.1). *ii*) Obtenemos estadísticos descriptivos de la segregación por grupo socioeconómico (ver Anexo A.2.2). *iii*) Desagregamos la entropía y obtenemos cuánto contribuye cada grupo a la suma de la entropía medida en una zona (ver Anexo A.2.3).

En cuanto al punto *i*) obtenemos que tanto para Q_{out} , al igual que para Q_{in} , predominan interacciones que reflejan que las personas tienden a interactuar más en promedio con personas de su mismo nivel socioeconómico. No obstante, se puede observar una contribución similar de los grupos $C2$, $C3$ y D a zonas de ingresos medios y bajos, pero también un aporte a zonas $ABC1$, lo cual se debe a que personas de distintos sectores logran interactuar con personas $ABC1$ (se integran más). Por otro lado, en promedio sectores de altos ingresos reflejan interacciones con sujetos mayoritariamente de su mismo grupo, donde si bien existe un aporte de otros grupos, este no es suficiente para cambiar lo que en promedio aportan sujetos también $ABC1$.

En cuanto al punto *ii*), obtenemos que los grupos que menos segregan son individuos de ingresos medios ($C2$ y $C3$), mientras que aquellos que más segregan son los ubicados en los extremos de la distribución de ingresos ($ABC1$, D y E). Las causas de esto último pueden ser muy diferentes entre ambos grupos, dado que aquellos de altos ingresos puede ser por elección (y/o por poseer todos los servicios necesarios en zonas de su mismo nivel), mientras que aquellos de bajos ingresos quizá no pueden integrarse dado que no poseen los medios para ello: bajo nivel de escolaridad, baja empleabilidad u otro, lo cual dificulta que salgan de su zona de origen.

Lo anterior se sustenta por medio de un análisis descriptivo de algunas variables demográficas (sección A.2.4) desglosadas por grupo socioeconómico, donde grupos de altos ingresos poseen un menor desempleo y una mayor escolaridad. Ahora, claramente estas variables no son suficientes para explicar el porqué grupos de altos y bajos ingresos, segregan más que ingresos medios, pero si pueden explicar parte de estos resultados.

Otro punto relevante, es la desviación estándar que presenta la segregación dentro de cada grupo, donde se puede observar que la desviación tiende a ser mayor en el grupo $ABC1$, cuyo valor mínimo de segregación dentro de las zonas $ABC1$ es cercano al de otros grupos, mientras que su valor máximo, mayor al de todos los demás grupos. Esto sugiere que en zonas de altos ingresos, existen algunas zonas que por alguna razón logran integrar a la comunidad, ya sea porque poseen parques, malls u otra estructura que sea un facilitador e encuentros diversos.

En cuanto al punto *iii*), obtenemos que si bien el desagregar la entropía por grupo da una aproximación a entender cómo contribuyen los distintos grupos a la segregación de una zona, el análisis no es independiente por grupo, dado que es una distribución, las proporciones de los distintos grupos no son independientes entre si (si un grupo posee más peso en la distribución, otros tendrán menos y viceversa), y además, un grupo puede contribuir una misma cantidad a la entropía poseyendo una proporción baja o alta (por concavidad), de esta manera, esto puede dar lugar a que se malinterprete la contribución de un grupo como si dicho grupo es el que más ayuda la integración¹⁶.

Además de lo anterior, obtenemos que para cada grupo socioeconómico, las distribuciones Q_{in} y Q_{out} , son bastante similares, lo cual es la causa de que EI_{in} y EI_{out} , sean a su vez muy similares. El que ambas distribuciones sean muy similares se entiende de que en promedio las relaciones que establecen los individuos son con sujetos parecidos a los que llegan a la zona donde habitan: una persona del centro de Santiago digamos en un rango de ingresos C2 se mueve a otra zona donde interactúa con otras que en promedio son C2, y al mismo tiempo, otros individuos también C2 van a la zona de donde salió (habita) la primera persona.

Lo anterior es claro para los grupos $ABC1$, $C3$ y D , los cuales en ambas distribuciones (Q_{in} y Q_{out}) interactúan mayoritariamente con personas de su mismo nivel (ver fig. A.9). En particular para el grupo E este interactúa más con personas del grupo D , y en segundo lugar con personas de su mismo grupo, mientras que para personas del grupo $C2$ su mayor interacción es con personas del grupo $C3$, D y luego $C2 - ABC1$.

En relación a lo anterior, al analizar los destinos de las personas, tenemos que en promedio solo un 35% sale de su comuna, siendo este último relativamente constante a través de los distintos grupos socioeconómicos. Por otro lado, esto cambia al tomar el porcentaje de individuos que se dirige a una zona de un nivel distinto al propio, donde hay mucha más varianza entre grupos, siendo el que menos se dirige a otras zonas de un nivel distintos el grupo $ABC1$ (19%) y el que más el grupo $C2$ (63.5%). Lo cual claramente explica el porqué el grupo $ABC1$ es el que más segrega mientras que el $C2$ el que menos.

El que estos dos grupos sean el que más ($ABC1$) y menos segregue ($C2$), puede resultar curioso dada su cercanía en la distribución de ingresos. Si observamos características demográficas para cada uno de estos grupos (sección A.2.4), dentro de las diferencias que existen entre estos, es que el grupo $ABC1$ posee en promedio una mayor escolaridad, vive más cerca al centro de Santiago y posee una mayor cantidad de áreas verdes. De esta manera, parte de las diferencias en la segregación de estos dos grupos, se puede explicar debido a estas variables: al poseer una mayor escolaridad pueden acceder a trabajos en los que se encuentran mayoritariamente con personas de su mismo nivel socioeconómico ($ABC1$), mientras que al poseer un menor nivel de escolaridad ($C2$), se ven forzados de cierta manera a trabajar en otras zonas, o trabajos en que a su vez, se encuentran con personas con un nivel de escolaridad menor (grupos $C2$, $C3$ o D).

A su vez, al poseer menos áreas verdes ($C2$), quizá sea un aliciente a moverse a otras zonas para poder disfrutar de un rato de esparcimiento. Mientras que al poseer más áreas verdes y en óptimas condiciones ($ABC1$), no poseen dicha necesidad, quedándose más en zonas de su mismo nivel socioeconómico.

En cuanto a la distancia al centro de Santiago, este también puede ser una causa de que individuos $C2$ deban moverse más (dado que habitan más lejos de sus trabajos), y de esta manera, interactuar con sujetos más diversos.

¹⁶ Se podría tener una alta (pequeña) entropía con una contribución pequeña (alta) a Q_{in} o Q_{out} .

4.5. Evolución temporal

En cuanto a la evolución de los índices, podemos observar que estos fluctúan de una manera muy parecida, y se condicen siempre con la movilidad de los individuos: a menor movilidad mayor segregación y viceversa (ver fig. 4.8). Luego, este mismo factor es uno de los que incide en que la segregación sea mayor durante los fines de semana, dado que al tener menos movilidad, las personas interactúan más con las personas de los lugares donde viven. Además de esto, otro motivo que puede también afectar que la segregación sea mayor los fines de semana, es que las interacciones que surgen entre individuos pueden reflejar elecciones personales (juntarse con amigos, etc) más que relaciones *forzosas* que son cuando deben ir a trabajar, y por ende, integrarse forzosamente quizá.

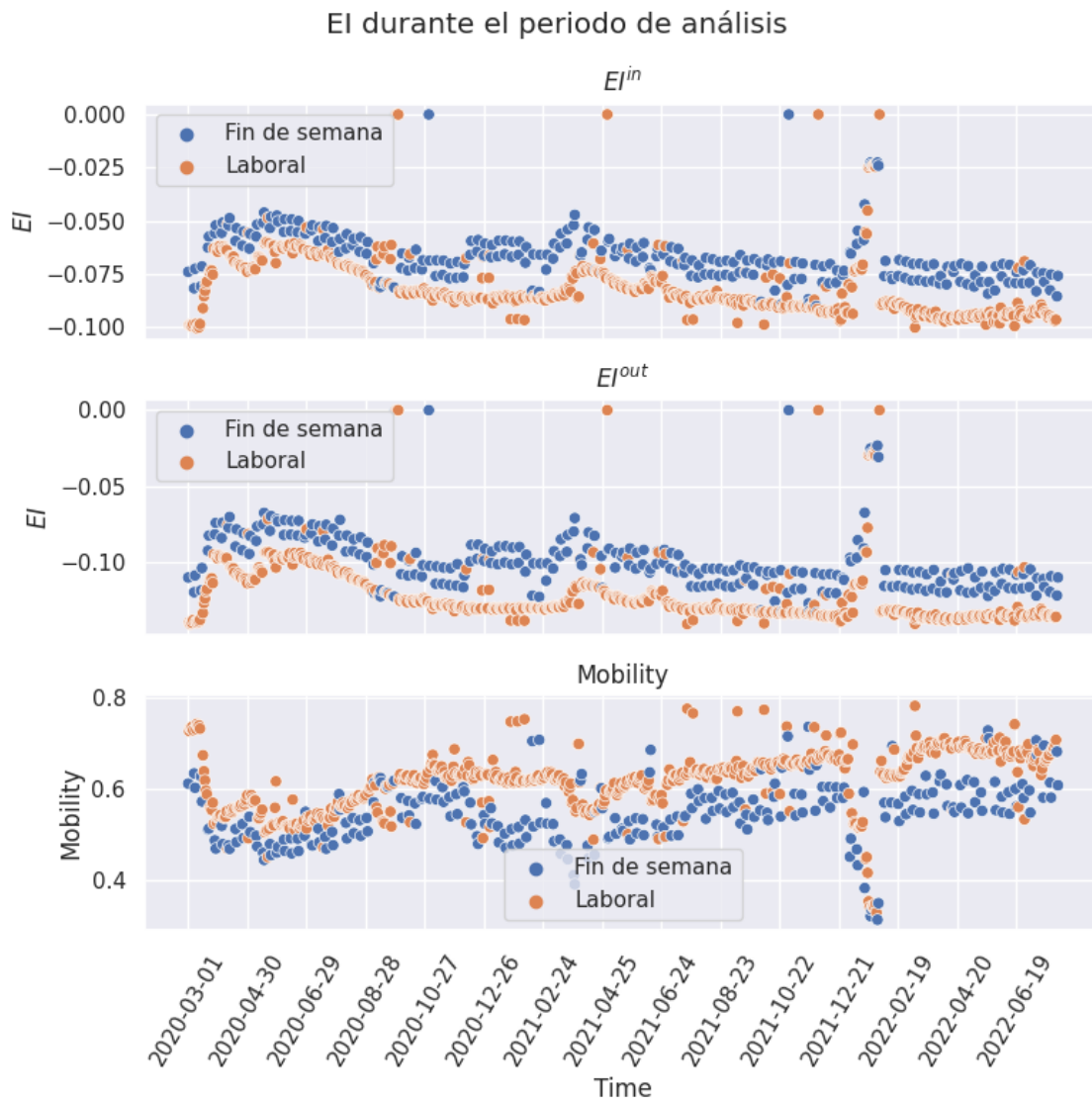


Figura 4.8: Evolución de los índices durante el periodo analizado.

4.6. Segregación, nivel de ingresos y periferia

En esta parte analizamos la relación de la segregación con el nivel de ingresos¹⁷ y la relación de la segregación con la cercanía al centro de la RM. Con esto último queremos estudiar si es que sectores segreguen solo por el hecho de vivir más lejos, o en su defecto, se integran más por estar en un sector más céntrico.

La primera variable que observamos su relación con la segregación, y de la cual ya hemos hablado previamente, son los ingresos. Podemos observar que el obtener el nivel de ingresos (entendido aquí como %ABC1) versus la segregación de la zona, la curva presenta una relación no lineal, mostrando un comportamiento creciente y decreciente según los ingresos: para sectores de bajos ingresos (inferior a 20% ABC1) se observa un comportamiento decreciente, esto es, mientras más ingresos menos se segrega, y a su vez, para sectores de medios y altos ingresos (superior a 20% ABC1), a medida que aumenta el ingreso aumenta la segregación (ver. fig. 4.9) experiencial.

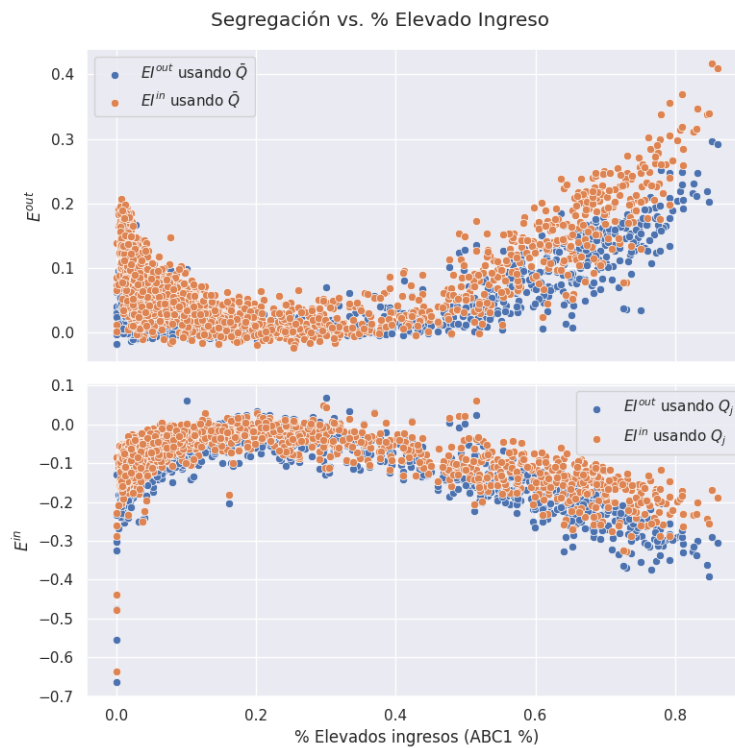


Figura 4.9: Curva segregación vs ingresos.

De esto, al igual que en secciones anteriores, sectores medios son los que más se integran, saliendo a zonas de un nivel distintos al suyo, lo cual se refleja en distribuciones agregadas más homogéneas, tanto para Q_{in} como Q_{out} .

Por otro lado, en cuanto a la relación de segregación y lejanía al centro de la ciudad, podemos mencionar que no existe una diferencia en la segregación al considerar si una zona es más periférica que otra. Sólo se observa una concentración en distancias menores, pero no una relación o incidencia de la distancia en la segregación en general (ver fig. 4.10).

¹⁷ Que ya hemos realizado a lo largo del informe y acá sólo complementamos.

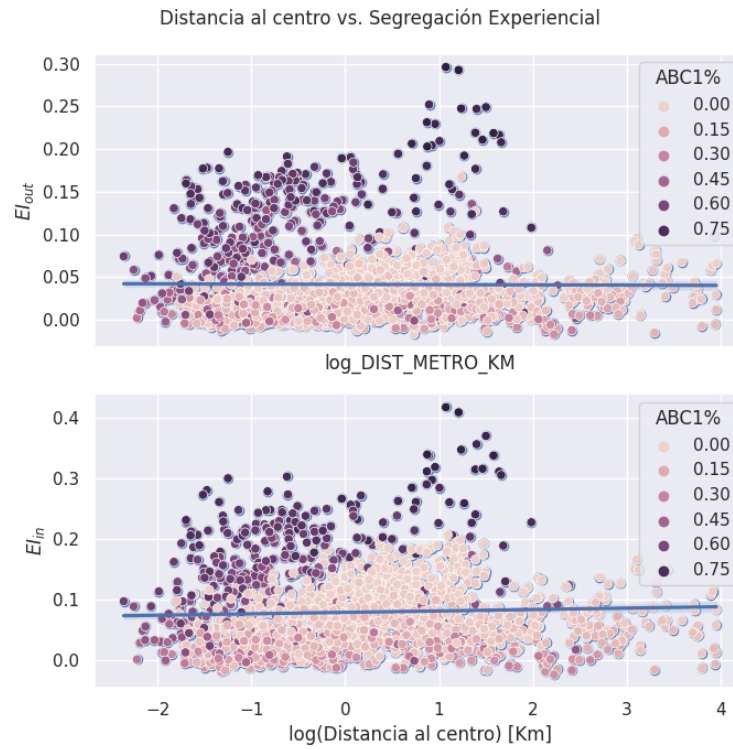


Figura 4.10: Curva segregación vs Distancia al centro en kilómetros. El gráfico superior posee un $r^2 = -0.01$ y el inferior $r^2 = 0.04$.

Capítulo 5

Determinantes de la segregación

En la sección anterior pudimos notar que existe heterogeneidad dentro de algunas zonas. En esta parte, buscaremos responder a parte de dicha heterogeneidad, estudiando en particular la relación de segregación y la infraestructura crítica que estamos considerando, esto es, metro, malls y parques.

Como se mencionó en la sección previa, es probable que la segregación y las relaciones planteadas en las regresiones anteriores sean diferentes entre fines de semana y días laborales, esto dado que en la semana las interacciones pueden representar mayoritariamente el trabajo de las personas, mientras que los fines de semana, pueden representar elecciones propias de los individuos sobre con quién relacionarse o a que lugares ir. Para estudiar esto, las regresiones, se diferenciarán para el caso de días laborales, y para fines de semana

Primero comenzamos estudiando la relación de la segregación con cada una de las infraestructuras anteriormente mencionadas, donde además, hacemos una diferenciación durante semana y fines de semana agregando una dummy de control ($\mathbf{1}_{weekend}$) que indica si el día respectivo es fin de semana o no. Luego, obtenemos lo siguiente

$$EI_i^{in,out} = X_{1i} + \beta_0 \quad (5.1)$$

donde $X_{1i} \in \{Mall_i + \mathbf{1}_{weekend}, Park_i + \mathbf{1}_{weekend}, subway_i + \mathbf{1}_{weekend}, Mall_i \cdot Park_i + Mall_i \cdot subway_i + Park_i \cdot subway_i + \mathbf{1}_{weekend} + Park_i + Mall_i + subway_i\}$, donde $Mall_i$, $Park_i$ ¹⁸ y $subway_i$ ¹⁹ corresponden a dummies que indican si en la zona i existe la infraestructura correspondiente.

Luego, para evaluar si existe una diferencia por nivel socioeconómico, generamos una interacción a lo anterior por el nivel socioeconómico de la zona, esto es, el porcentaje de hogares de altos ingresos que hay en ella, el cual denotamos por $income_i$. De esta manera, consideramos los siguientes modelos

$$EI_i^{in,out} = X_{2i} + \beta_0 \quad (5.2)$$

donde $X_{2i} \in \{Income_i \cdot Mall_i + Income_i + Mall_i + \mathbf{1}_{weekend}, Income_i \cdot Park_i + Income_i + Park_i + \mathbf{1}_{weekend}, Income_i \cdot subway_i + Income_i + subway_i + \mathbf{1}_{weekend}\}$, e $Income_i$ corresponde

¹⁸ Se seleccionó un valor umbral de 0.2, sobre el número de áreas verdes *buenas* que existe por comuna. Por ejemplo, si en una zona existía una razón de áreas verdes por persona de 10, y la razón de áreas verdes buenas en su comuna era de 0.8, la variable sería $park_i = 10 * 0.8 = 8$, y luego sobre esto se aplica el umbral.

¹⁹ Se toma como valor de soporte 500 metros, esto es, si la distancia es menor a 500 metros, entonces la dummy es uno, y cero en caso contrario.

a la proporción de hogares de altos ingresos en la zona i .

Ahora, para analizar si en lo anterior existe una diferenciación por la presencia de metro, esto es por ejemplo, que en lugares de altos ingresos en parques se integre más al estar más cerca de un metro que aquellos parques en zonas de bajos o medios ingresos, agregamos a la regresión anterior una interacción con ingresos de la respectiva zona la proporción de hogares de altos ingresos en la zona i .

$$EI_i^{in,out} = X_{3i} + \beta_0 \quad (5.3)$$

con $X_{3i} \in \{Income_i \cdot Mall_i \cdot subway_i + Income_i + Mall_i + subway_i + \mathbf{1}_{weekend}, Income_i \cdot Park_i \cdot subway_i + Income_i + Park_i + subway_i + \mathbf{1}_{weekend}\}$.

5.1. Resultados regresiones

Un punto común a todos los modelos considerados, es la variable indicativa de si el día correspondiente es fin de semana o no. A partir de los resultados, obtenemos que la segregación es mayor durante los fines de semana, lo cual se condice con el análisis visual de secciones previas (fig. 4.8). El que la segregación sea mayor durante los fines de semana, como ya se mencionó, en parte se debe a la reducción de movilidad, y en parte a las elecciones de los individuos.

En cuanto a la infraestructura de interés, para malls podemos observar que por sí solo no es significativo en promedio sobre la segregación, no obstante, si lo es su interacción con otras variables, donde obtenemos que cuando hay un mall y un estación de metro cerca la segregación disminuye en promedio (tabla 5.1).

Ahora, cuando se adiciona un control por ingresos (tabla 5.2) obtenemos que los malls en promedio disminuyen la segregación con excepción de aquellos ubicados en zonas de altos ingresos. Esto sugiere que las limitaciones de acceso, u otros motivos, inhiben que en promedio personas de otros sectores puedan ir a esos malls.

Pese a esto, viendo la correlación cuando se controla por estación de metro (tabla 5.3) obtenemos que en promedio la presencia de metro y mall no son un factor integrador en toda la ciudad, pero sí lo son en aquellas zonas de altos ingresos: zonas de altos ingresos con mall y estación de metro cerca.

Más aún, en promedio los malls ubicados en zonas de altos ingresos son visitados mayoritariamente por sujetos de un nivel socioeconómico alto, pero cuando hay una estación de metro cerca, esto facilita que personas de otras partes puedan moverse a dicho lugar, y con esto, disminuir la segregación.

En cuanto a la presencia de parques en la ciudad, tenemos que en promedio los parques tienden a aumentar la segregación: personas parecidas van a los mismos parques (tabla 5.1). Esto se debe quizá a que las personas prefieren normalmente aquellas áreas verdes que le son más cercanas, lo cual hace que interactúe con sujetos más parecidos a ellos mismos. Otro motivo, puede ser que esto se deba a que las zonas de altos ingresos poseen más áreas verdes en relación a su población, lo cual hace que en parques en promedio se segregue.

Para analizar lo anterior, agregamos un control por ingresos. Cuando controlamos por ingresos (tabla 5.2) tenemos que los parques en promedio integran, con excepción de aquellos de zonas de altos recursos, los cuales segregan. Luego, al observar los resultados cuando involucramos la presencia del metro (tabla 5.3) obtenemos zonas con parques y metro en promedio segregan, pero cuando se interactúa con ingresos, esta relación se invierte. Esto es que, aquellos parques ubicados en zonas de altos ingresos y que poseen un parque cercano a

una estación de metro, disminuyen la segregación.

Ahora, en cuanto al metro por sí solo, tenemos que esta logra la integración de individuos diversos (tabla 5.1), pero cuando se crea una interacción con el nivel socioeconómico, esta relación se invierte. Lo que quiere decir que en promedio, el metro ayuda a la integración en Santiago, con excepción de zonas que poseen elevados recursos (en promedio). No obstante, como ya se mencionó previamente, en lugares específicos como malls y parques, y que son de elevados ingresos, el metro es en efecto un factor integrador.

Lo anterior, sugiere que si bien hay movimientos que pueden producirse a zonas de altos ingresos, estos no son suficientes para que las distribuciones que allí se generen del agregado de personas que llegan a esos lugares, sean más heterogéneas. A pesar de esto, cuando hay zonas que poseen infraestructuras que son llamativas para personas de distintos lados, y que al mismo tiempo, tengan elementos que les faciliten el acceso a dicha infraestructura, esto permite que sujetos de distintas zonas vayan y se encuentren en dichos espacios, disminuyendo con esto la segregación.

Tabla 5.1: Resultados modelos que analizan relación entre infraestructura y segregación.

Variable	EI_{out}				EI_{in}			
	1)	2)	3)	4)	1)	2)	3)	4)
constante	0.04***(0.00)	0.04***(0.00)	0.03***(0.00)	0.03***(0.00)	0.07***(0.00)	0.08***(0.00)	0.07***(0.00)	0.07***(0.00)
$1_{weekend}$	0.02***(0.02)	0.02***(0.02)	0.02***(0.00)	0.02***(0.00)	0.01***(0.00)	0.01***(0.00)	0.01***(0.00)	0.01***(0.00)
$mall_i$	0.01 (0.01)			0.01 (0.00)	-0.00 (0.00)			-0.01 (0.00)
$subway_i$		-0.00 (0.00)		0.00 (0.00)		-0.02***(0.00)		-0.01***(0.00)
$park_i$			0.02***(0.00)				0.02***(0.00)	0.02***(0.00)
$park_i \cdot mall_i$				0.04***(0.00)				0.05*(0.01)
$mall_i \cdot subway_i$				-0.04***(0.00)				-0.04*(0.01)
$park_i \cdot subway_i$				-0.02****(0.01)				-0.02***(0.01)
Observaciones	$r^2 = 0.30$	$r^2 = 0.32$	$r^2 = 0.34$	$r^2 = 0.17$				

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Entre paréntesis se observa la desviación estándar.

Tabla 5.2: Resultados modelos que analizan relación entre ingreso como mediador de infraestructura

Variable	EI_{out}			EI_{in}		
	1)	2)	3)	1)	2)	3)
constante	0.01***(0.00)	0.01***(0.00)	0.02***(0.00)	0.05***(0.00)	0.05***(0.00)	0.06***(0.00)
$1_{weekend}$	0.02***(0.00)	0.02***(0.00)	0.02***(0.00)	0.01***(0.00)	0.01***(0.00)	0.01***(0.00)
$income_i$	0.11***(0.00)	0.01***(0.00)	0.12***(0.00)	0.11***(0.00)	0.08***(0.00)	0.12***(0.00)
$mall_i$	-0.05***(0.00)			-0.10***(0.00)		
$park_i$		-0.01***(0.00)			-0.01***(0.00)	
$subway_i$			-0.03***(0.00)			-0.06***(0.00)
$mall_i \cdot income_i$	0.01***(0.00)			0.17***(0.00)		
$park_i \cdot income_i$		0.07***(0.00)			0.10***(0.00)	
$subway_i \cdot income_i$			0.02***(0.00)			0.05***(0.00)
Observaciones	$r^2 = 0.30$	$r^2 = 0.32$	$r^2 = 0.34$	$r^2 = 0.17$	$r^2 = 0.18$	$r^2 = 0.22$

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Entre paréntesis se observa la desviación estándar.

Tabla 5.3: Resultados modelos que analizan relación entre ingreso y metro, como mediador de infraestructura

Variable	EI_{out}		EI_{in}	
	1)	2)	1)	2)
constante	0.02*** (0.00)	0.02*** (0.00)	0.06*** (0.00)	0.07*** (0.00)
$1_{weekend}$	0.02*** (0.00)	0.02*** (0.00)	0.01*** (0.00)	0.01*** (0.00)
$subway_i$	-0.03*** (0.01)	-0.04*** (0.01)	-0.06*** (0.01)	-0.07*** (0.01)
$park_i$		-0.01*** (0.00)		-0.01** (0.00)
$mall_i$	-0.06*** (0.01)		-0.12*** (0.01)	
$income_i$	0.22*** (0.02)	0.01*** (0.00)	0.12*** (0.00)	0.01*** (0.00)
$subway_i \cdot mall_i \cdot income_i$	-0.15*** (0.00)		-0.25*** (0.00)	
$subway_i \cdot park_i \cdot income_i$		-0.09*** (0.00)		-0.13*** (0.00)
$subway_i \cdot park_i$		0.02*** (0.00)		0.02** (0.00)
$subway_i \cdot mall_i$	-0.05*** (0.00)		0.10*** (0.00)	
$subway_i \cdot income_i$	0.03** (0.00)	0.07*** (0.01)	0.06*** (0.01)	0.12*** (0.01)
$mall_i \cdot income_i$	0.11** (0.02)	0.01*** (0.00)	0.21*** (0.01)	0.13*** (0.01)
Observaciones	$r^2 = 0.30$	$r^2 = 0.37$	$r^2 = 0.23$	$r^2 = 0.25$

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Entre paréntesis se observa la desviación estándar.

Capítulo 6

Conclusiones

De las secciones anteriores obtuvimos que la segregación está positivamente correlacionada tanto la segregación residencial como con los ingresos, de esta manera individuos de alto recursos son los que más segregan tanto en sus interacciones como en los lugares donde viven.

A partir del análisis por grupo socioeconómico, los grupos ubicados en los extremos de la distribución de ingresos ($ABC1$, D y E) son los grupos que más segregan, siendo el que más segrega el grupo $ABC1$, el cual muestra distribuciones de encuentro más sesgadas a su mismo grupo socioeconómico y al mismo tiempo el que menos interactúa con individuos de otros grupos. Las razones de estos grupos sean los que más segreguen pueden ser diversas, y diferentes para aquellos de altos ingresos y bajos: para grupos de altos ingresos pueden quizá representar elecciones, mientras que para individuos de bajos ingresos falta de oportunidades o incapacidad de integrarse (no poseer el nivel educacional área trabajar en otras áreas, vivir lejos, etc.).

Por otro lado, grupos de ingresos medios son los que menos segregan ($C2$ y $C3$) siendo el que menos segrega el grupo $C2$, el cual muestra distribuciones agregadas de encuentro y de llegada bastante homogéneas, y a su vez, interactúa mucho más con otros grupos socioeconómicos (por sobre el 60%).

En este mismo sentido, la segregación experiencial está íntimamente relacionada con la movilidad de las personas, donde a mayor movilidad menor segregación, siendo este uno de los motivos del porque durante los fines de semana la segregación aumenta (individuos se mueven menos). Al observar hacia dónde van los individuos por zona y grupo, los grupos muestran tener una proporción de individuos similar que sale de su comuna ($\sim 35\%$), mientras que existe una gran varianza a través de grupos cuando se considera si estos se mueven a una zona de distinto nivel socioeconómico (el mínimo es 19% para $ABC1$ y máximo 63% para $C2$).

De esta manera, individuos de altos ingresos se mueven en promedio por comunas de altos ingresos, y con un nivel socioeconómico igual al de ellos. Mientras que grupos de ingresos medios, son los que más interactúan con individuos distintos a ellos.

En cuanto a la relación entre las medidas EI^{out} y EI^{in} , éstas poseen una elevada correlación entre sí, y a su vez, una elevada correlación con la segregación residencial. Esto muestra que las personas tienden a relacionarse con sujetos parecidos a ellos desde un punto de vista socioeconómico, y tienden a visitar en promedio lugares de su mismo estrato, lo cual causa que estas métricas sean parecidas. No obstante, ambas mediciones experienciales son menores a la residencial, por lo que a pesar de la correlación mencionada, efectivamente sí existe

integración en la ciudad.

En cuanto a la infraestructura de interés, tenemos que tantos parques, como malls son factores integradores en promedio, siendo relevante en zonas de altos ingresos donde por medio del metro se vuelven zonas que integran la ciudad, y que de otra manera si no hubiera metro, estas serían zonas más segregadas.

Como se menciona en un inicio del documento, estos resultados pueden orientar el posicionamiento de futuras estaciones de metro, encontrando zonas donde a partir de las características descritas de un alto ingreso, metro cercano y lugar de interés público, puedan con esto disminuir la segregación en la RM.

Bibliografía

- Athey, S. et al. 2020. Experienced Segregation, NBER Working Papers 27572, National Bureau of Economic Research, Inc. [<http://www.nber.org/papers/w27572>].
- Beiró, M., et al. 2018. Shopping mall attraction and social mixing at a city scale. EPJ Data Sci. 7, 28. [<https://doi.org/10.1140/epjds/s13688-018-0157-5>]
- Boar, C., Giannone, E. 2023. Consumption Segregation. NBER Working Paper No. w31133, Available at SSRN: [<https://ssrn.com/abstract=4420548>]
- Browning, CR., et al. 2017. Ecological networks and neighborhood social organization. American Journal of Sociology. 122:1939–1988. [<https://dx.doi.org/10.1086%2F691261>].
- Cagney, K. E., et al. 2020. Urban mobility and activity space. Annu. Rev. Sociol. 46, 623–648. [<https://doi.org/10.1146/annurev-soc-121919-054848>]
- Carranza A., et al. 2020. The Social Divide of Social Distancing: Shelter-in-Place Behavior in Santiago during the Covid-19 Pandemic. [<https://doi.org/10.48550/arXiv.2103.13292>].
- Cowgill, D., Cowgill, M. 1951. An Index of Segregation Based on Block Statistics. American Sociological Review 16:825-31. [<https://doi.org/10.2307/2087511>].
- Cutler, D., Glaeser E., Vigdor, J. 1999. The Rise and Decline of the American Ghetto. The Journal of Political Economy 107(3):455–506. [<http://dx.doi.org/10.1086/250069>].
- Cover T.M., Thomas J.A., Elements of Information Theory, Wiley and Sons, New York, NY, 1991.
- Duncan, D. B. 1955. Multiple range and multiple F tests. Biometrics, 11, 1–41. [<https://doi.org/10.2307/3001478>]
- Drane, C., Macnaughtan, M., Scott, C. 1998. Positioning gsm telephones. IEEE Communications Magazine 36(4):46–54. [<https://doi.org/10.1109/35.667413>].
- Departamento de demografía y censos, Instituto Nacional de Estadísticas (INE). Manual de usuario de la base de datos del censo de población y vivienda 2017, Septiembre 2018.
- Echenique, F., Fryer, R. 2007. A Measure of Segregation Based on Social Interactions. The Quarterly Journal of Economics 122(2):441–85. [<http://www.its.caltech.edu/~fede/published/echenique-fryer-qje.pdf>].
- Endres D. M., Schindelin J. E. 2003. A new metric for probability distributions, in IEEE Transactions on Information Theory, vol. 49, no. 7, pp. 1858-1860. [<https://doi.or/TIT..813506>].
- Frankel, D., Volij, O. 2011. Measuring school segregation, Journal of Economic Theory 146(1), 1-38. [<https://doi.org/10.1016/j.jet.2010.10.008>]

- Gobster, P. 1998. Urban parks as green walls or green magnets? Interracial relations in neighborhood boundary parks, *Landscape and Urban Planning*, Volume 41, Issue 1, Pages 43-55, ISSN 0169-2046, [[https://doi.org/10.1016/S0169-2046\(98\)00045-0](https://doi.org/10.1016/S0169-2046(98)00045-0)].
- Gentzkow, M., Shapiro J. 2011. Ideological Segregation Online and Offline. *The Quarterly Journal of Economics* 126(4):1799–1839. [<http://www.nber.org/papers/w15916>].
- Jahn, J. A. 1950. The Measurement of Ecological Segregation: Derivation of an Index Based on the Criterion of Reproducibility. *American Sociological Review* 15:101-104. [<http://dx.doi.org/10.2307/2086407>].
- Jones, M., Pebley, A. 2014. Redefining neighborhoods using common destinations: Social characteristics of activity spaces and home census tracts compared. *Demography* 51:727–752. [<https://dx.doi.org/10.1007%2Fs13524-014-0283-z>].
- Lee, LJZ., Shu. En, Shen, L. 2021. Segregation across neighborhoods in a small city. *Asia Compet. Inst. Res. Pap. Ser.* [https://lkyspp.nus.edu.sg/docs/default-source/aci/acirp202107.pdf?sfvrsn=a862240a_2]
- Lamont, M., Beljean, S., Clair, M. 2014. What is missing? cultural processes and causal pathways to inequality, *Socio-Economic Review* 12(3), 573-608.[<https://doi.org/10.1093/ser/mwu011>].
- Massey, D., Denton, N. 1988. The Dimensions of Residential Segregation. *Social Forces. A Scientific Medium of Social Study and Interpretation* 67(2):281–315. [<https://dx.doi.org/10.1093%2Fsf%2Fsos118>].
- Reardon, S., Bischoff, K. 2011. Income Inequality and Income Segregation. *American Journal of Sociology*, 116(4), 1092–1153. [<https://doi.org/10.1086/657114>]
- Reardon, S. F., Firebaugh, G. 2002. Measures of Multigroup Segregation. *Sociological Methodology*, 32, 33–67.[<http://www.jstor.org/stable/3186153>]
- Gobster, P. 1998. Urban parks as green walls or green magnets? Interracial relations in neighborhood boundary parks, *Landscape and Urban Planning*. [[https://doi.org/10.1016/S0169-2046\(98\)00045-0](https://doi.org/10.1016/S0169-2046(98)00045-0)]
- Phillips, N., Levy, B., Sampson, R., Small, M., Wang, R. 2021. The Social Integration of American Cities: Network Measures of Connectedness Based on Everyday Mobility across Neighborhoods. *Social Methods & Research* 50:1110-1149. [<https://doi.org/10.1177%2F0049124119852386>].
- Shelton, T., Poorthuis, A., Zook, M. 2015. Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information *Landsc. Urban Planning* 142:198–211. [<https://doi.org/10.1016/j.landurbplan.2015.02.020>].
- Swyngedouw, E. 2013. The Segregation of Social Interactions in the Red Line L-Train in Chicago. *Symbolic Interaction*, 36: 293-313.[<https://doi.org/10.1002/symb.64>]
- James, D. R., Taeuber, K. E. 1985. Measures of Segregation. *Sociological Methodology*, 15, 1–32. [<https://doi.org/10.2307/270845>]
- Theil, H. 1967. *Economics and Information Theory*, North-Holland, Amsterdam.
- Wong, DW., Shaw, S. 2011. Measuring segregation: An activity space approach. *Journal of Geographical Systems* 13:127–145. [<https://dx.doi.org/10.1007%2Fs10109-010-0112-x>].

- White, Michael J. 1983. The Measurement of Spatial Segregation. *American Journal of Sociology*. [<https://www.journals.uchicago.edu/doi/10.1086/227768#:~:text=https%3A//doi.org/10.1086/227768>]
- Yabe, T., et al. 2023. Behavioral changes during the COVID-19 pandemic decreased income diversity of urban encounters. *Nat Commun* 14, 2310. [<https://doi.org/10.1038/s41467-023-37913-y>]
- Zhang, X., Wang, J., Kwan, MP., Chai, Y. 2019. Reside nearby, behave apart? activity-space-based segregation among residents of various types of housing in Beijing, China. *Cities* 88:166–180. [<https://doi.org/10.1016/j.cities.2018.10.009>].

Anexos

Anexo A. Otros Resultados

A.1. Relación entre movilidad y segregación

Relación de movilidad y segregación Para estudiar los puntos enteramente mencionados, primero obtenemos la relación entre movilidad y EI_{out} . De la sección anterior, se aprecia una clara correlación negativa entre movilidad y segregación, dicha relación la formalizamos estudiando un modelo de datos de panel, donde controlaremos por α_i y ϕ_t de modo de inhibir aquellos factores no observables de cada zona i , y aquellos factores causados por restricciones temporales (esto es importante considerando las restricciones en movilidad de la pandemia). Luego, el modelo es el siguiente

$$EI_{it}^{out} = \beta_0 + \beta_i * Mob_{it} + \alpha_i + \phi_t + \epsilon_{it} \quad (A.1)$$

Tabla A.1: Resultados de modelo A.1. Se obtuvo un $r^2 = 0.07$

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Movilidad (Mob_{it})	-0.10	0.01	-12.13	0.00	-0.11	-0.08

De esta manera, obtenemos que efectivamente la movilidad posee una relación negativa con la segregación outbound, siendo este parámetro significativo estadísticamente (tabla A.1).

A.2. Resultados desagregados por grupo socioeconómico

Para entender mejor la segregación que experimenta cada grupo realizamos los siguientes ejercicios: *i*) Graficamos la contribución por grupo socioeconómico a cada zona en las distribuciones Q_{in} y Q_{out} . *ii*) Obtenemos estadísticos descriptivos de la segregación por grupo socioeconómico. *iii*) Desagregamos la entropía y obtenemos cuánto contribuye cada grupo a la suma de la entropía medida en una zona.

En las subsecciones siguientes realizamos distintas consideraciones sobre las distribuciones de cada zona:

- a) Consideramos la contribución de los distintos grupos a las distribuciones Q_{in} y Q_{out} , donde por ejemplo, para obtener como contribuye al grupo $ABC1$ a las distintas zonas de la RM, tomamos la proporción del grupo $ABC1$ en Q_{in} (o Q_{out} según corresponda) para cada zona, y luego obtenemos su representación geográfica.
- b) Asignamos un nivel socioeconómico a cada zona a partir del grupo que está más representado en una zona (el máximo) a nivel de hogares: por ejemplo, si una zona posee una

distribución socioeconómica de hogares $Q = (ABC1, C2, C3, D, E) = (12, 30, 5, 1, 0)$, dicha zona se considera como zona $C2$. En la subsección A.2.4, se presentan estadísticos descriptivos (tablas A.4, A.5, A.6, A.7, A.8) de acuerdo a esta segmentación para algunas variables de interés.

A.2.1. Contribución por grupo a cada zona en las distribuciones Q_{in} y Q_{out}

Utilizando las distribuciones Q_{in} y Q_{out} de cada zona, obtenemos la representación geográfica por grupo socioeconómico: por ejemplo para el grupo $C2$ (y $ABC1$), en la figura A.1 para cada zona se representa el peso de dicho grupo en la distribución Q_{in} .

El valor máximo teórico es de 1 y que significa (en la figura A.1) que dicha zona interactúa sólo con individuos $ABC1$ (figura izquierda en fig. A.1) o sólo $C2$ (figura derecha en fig. A.1). El valor mínimo es de cero que indica que el grupo en análisis no posee incidencia en la zona respectiva, mientras que el valor ideal es de 0.2, ya que indicaría que Q_{in} o Q_{out} corresponde a una distribución uniforme²⁰.

Observando la contribución que realiza cada grupo a las distintas zonas censales, tanto para la distribución agregada de llegada Q_{in} (figuras A.1, A.2, A.3) como de salida Q_{out} (figuras A.4, A.5, A.6), las personas del grupo $ABC1$ tienden a ir a zonas $ABC1$ e interactuar mayoritariamente por lo mismo con personas $ABC1$.

En cuanto a Q_{in} , obtenemos que los grupos de ingresos medios ($C2 - C3$) y bajo-medio (D) si bien contribuyen a sectores de altos ingresos (figuras A.1 y A.2) esta contribución no es lo suficientemente importante como para variar dramáticamente la contribución que realiza el mismo grupo $ABC1$ a dichas zonas (fig. A.1).

A su vez para Q_{out} , al igual que para Q_{in} , predominan interacciones que reflejan que las personas tienden a interactuar más en promedio con personas de su mismo nivel socioeconómico. No obstante, si bien se puede observar una contribución similar de los grupos $C2$, $C3$ y D a zonas de ingresos medios y bajos, también se aprecia un aporte del grupo $ABC1$, lo cual se debe a que personas de distintos sectores logran interactuar con personas $ABC1$ (se integran más).

²⁰ Recordar que nuestra distribución es de cardinalidad 5. Para una zona j cualquiera tenemos $Q_{in}^j = (ABC1, C2, C3, D, E)$, donde cada componente indica la proporción de dicho grupo en la distribución Q_{in}^j , de donde lo ideal es una distribución uniforme $Q_{in}^j = (0.2, 0.2, 0.2, 0.2, 0.2)$, ya que disminuye la segregación. Lo anterior es análogo para Q_{out} .

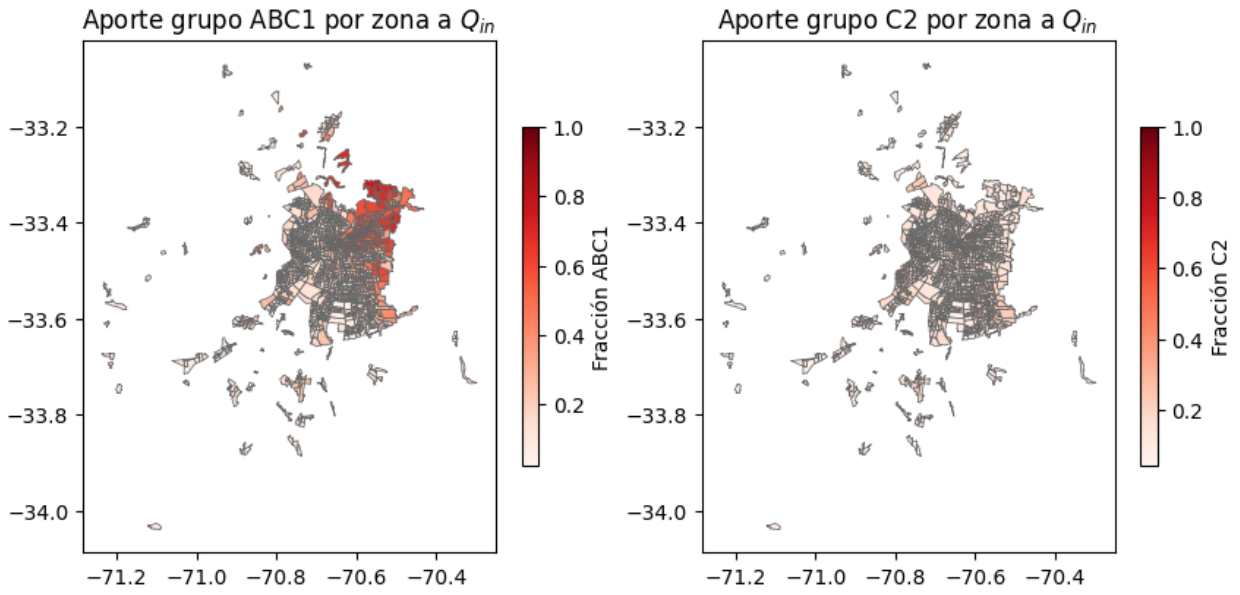


Figura A.1: Contribución de grupos $ABC1$ y $C2$ a la distribución agregada de llegada por zona.

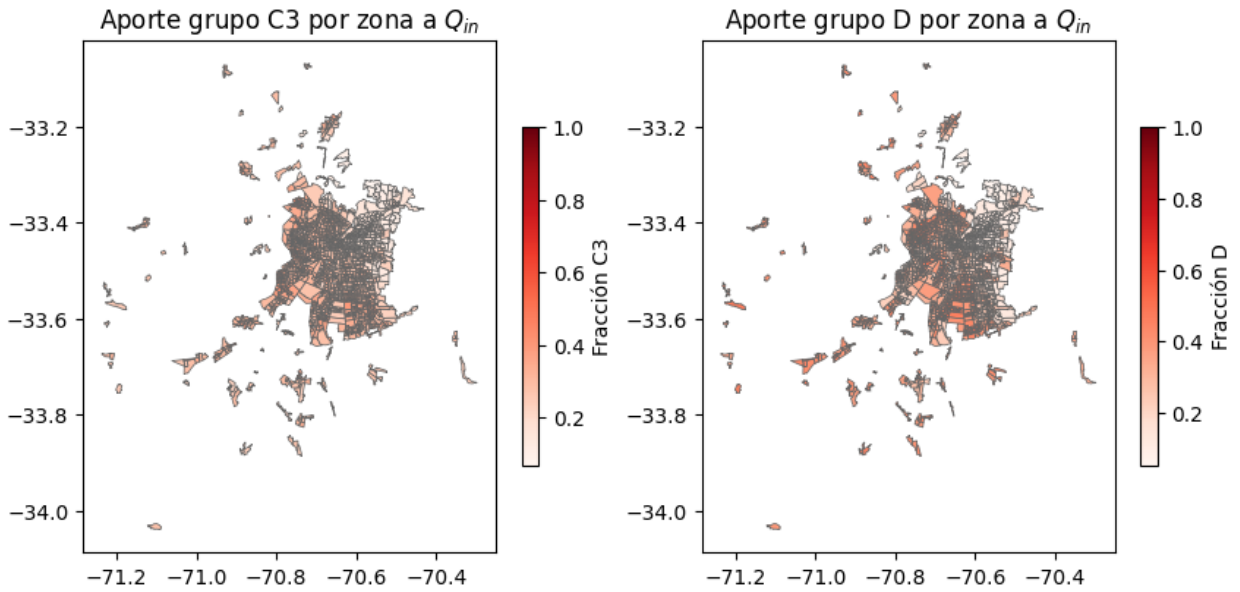


Figura A.2: Contribución de grupos $C3$ y D a la distribución agregada de llegada por zona.

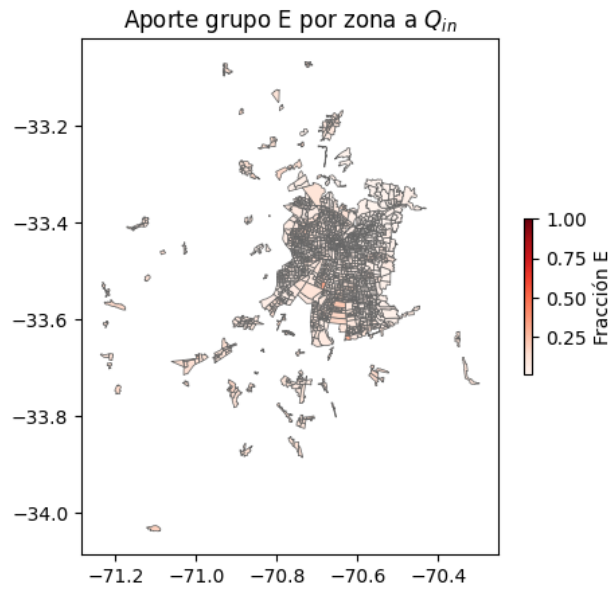


Figura A.3: Contribución de grupo E a la distribución agregada de llegada por zona.

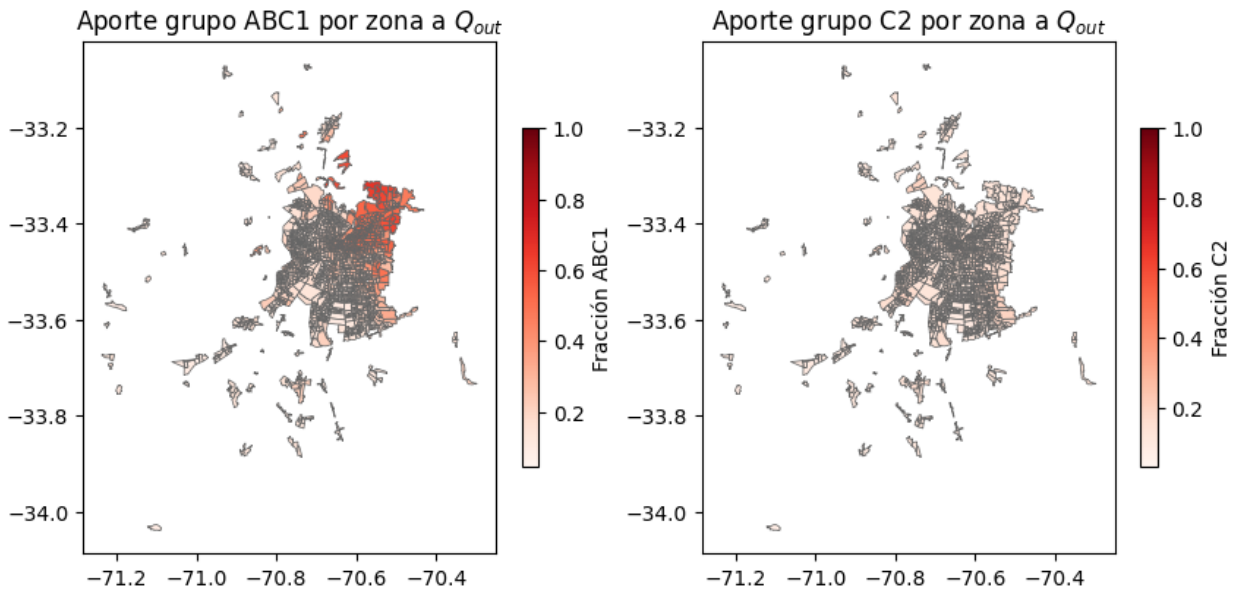


Figura A.4: Contribución de grupos $ABC1$ y $C2$ a la distribución agregada de salida por zona.

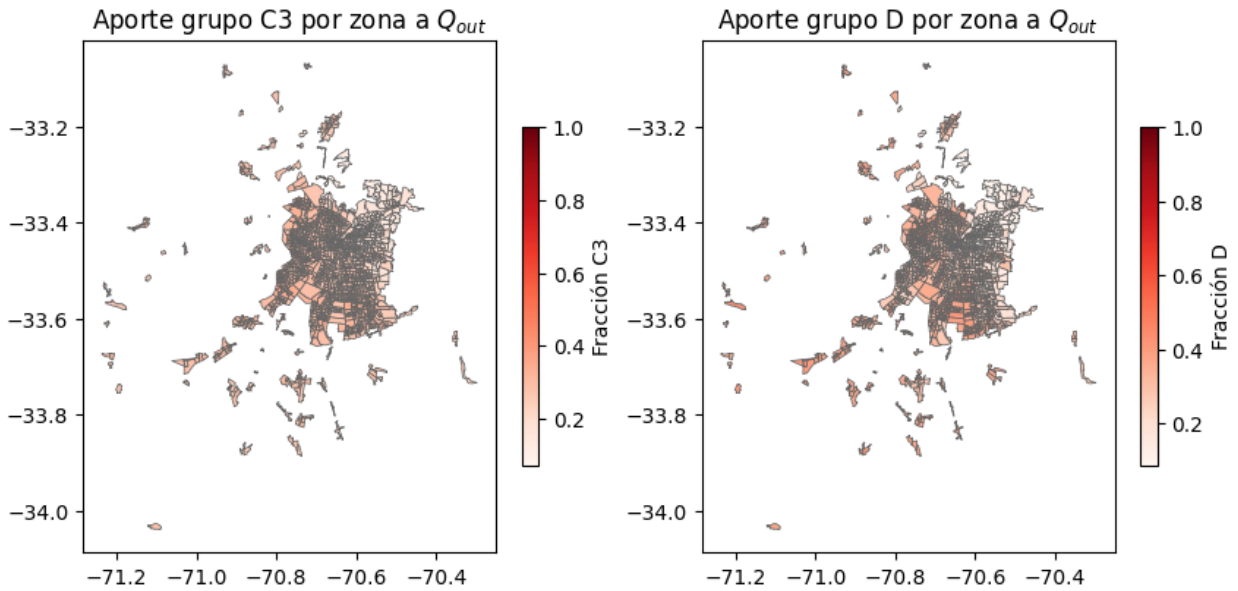


Figura A.5: Contribución de grupos $C3$ y D a la distribución agregada de salida por zona.

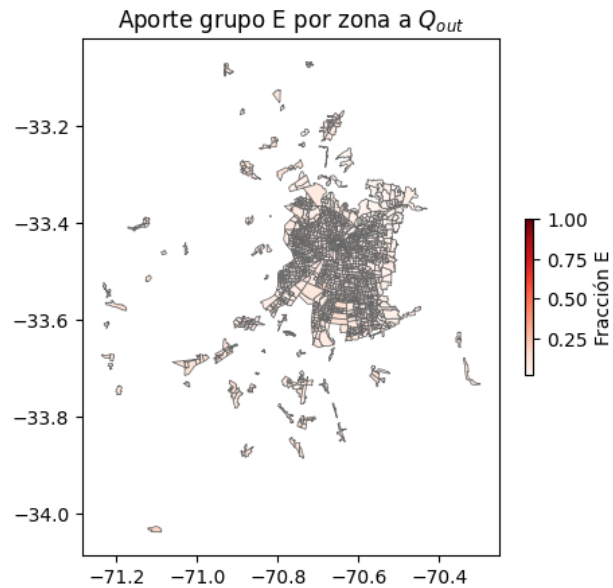


Figura A.6: Contribución de grupo E a la distribución agregada de salida por zona.

A.2.2. Estadísticos descriptivos de la segregación por grupo socioeconómico

Utilizando la clasificación por zona de acuerdo a al punto b) mencionado previamente, la mayoría de las zonas son clasificadas como $ABC1$, $C3$ y D , quedando menos representados los grupos $C2$ y E (ver tabla A.2). Esto es reflejo de la distribución de ingresos a nivel de hogares de la RM (ver fig. A.7).

Si bien es difícil concluir sobre los grupos con baja representación ($C2$ y E), dentro de los resultados que destacan es que los grupos que en promedio menos segregan (experiencial

y residencialmente), son los de ingresos medios $C2 - C3$, y aquellos que más segregan, los ubicados en los extremos de la distribución de ingresos ($ABC1$, D y E)²¹. Este resultado sugiere que tanto grupos de altos como bajos ingresos son los que más segregan. Ahora, las causas pueden ser muy diferentes entre ambos grupos, dado que aquellos de altos ingresos puede ser por elección (y/o por poseer todos los servicios necesarios en zonas de su mismo nivel), mientras que aquellos de bajos ingresos quizá no pueden integrarse dado que no poseen los medios para ello: bajo nivel de escolaridad, baja empleabilidad u otro, lo cual dificulta que salgan de su zona de origen.

Un resultado llamativo es que los grupos que más y menos segregan, sean respectivamente los grupos $ABC1$ y $C2$. Si bien el análisis de características específica a cada zona (fuera de malls, metro y parques) está fuera del alcance de esta tesis, en la subsección A.2.4 se presentan estadísticos descriptivos de algunas variables de interés para cada grupo socioeconómico. En particular para los grupos $ABC1$ y $C2$, presentan diferencias más importantes en cuanto a nivel de escolaridad, distancia al centro y superficie de áreas verdes por habitante, donde individuos $ABC1$ en promedio poseen una mayor escolaridad, viven más cerca del centro y poseen una elevada proporción de parques por habitantes en relación al grupo $C2$.

Otro punto relevante, es la desviación estándar que existe dentro de cada grupo. Se puede observar que la desviación tiende a ser mayor en el grupo $ABC1$, cuyo valor mínimo de segregación es cercano al de otros grupos, mientras que su valor máximo, mayor al de todos los demás grupos. Esto sugiere que en zonas de altos ingresos, existen algunas zonas que por algunas razón logran integrar a la comunidad, ya sea porque poseen parques, malls u otra estructura que sea un facilitador de encuentros diversos (ver tabla A.2).

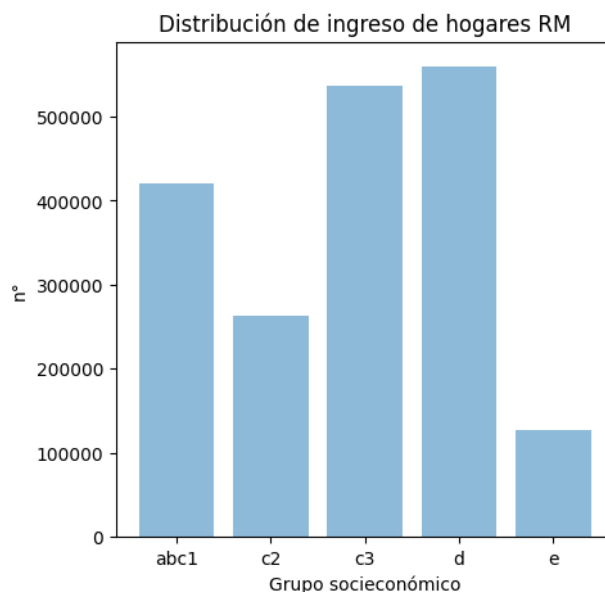


Figura A.7: Distribución de ingresos a nivel de hogar en la RM.

²¹ En relación al grupo E nos referimos residencialmente, dado que experiencialmente se integra, no obstante, este grupo posee una baja representatividad con apenas dos zonas

Tabla A.2: Estadísticos descriptivos de índices de segregación por grupo socioeconómico (con $Q_{base} = \bar{Q}$).

Índice	Media	std	min	25 %	50 %	75 %	max	n° zonas
<i>EI_{out}</i>								
<i>ABC1</i>	0.08	0.06	-0.01	0.02	0.07	0.13	0.30	466
<i>C2</i>	0.00	0.01	0.00	0.00	0.00	0.01	0.02	8
<i>C3</i>	0.02	0.02	-0.01	0.01	0.02	0.03	0.17	471
<i>D</i>	0.03	0.03	-0.02	0.01	0.03	0.05	0.13	853
<i>E</i>	0.00	0.03	-0.02	-0.01	0.00	0.01	0.02	2
<i>EI_{in}</i>								
<i>ABC1</i>	0.12	0.09	-0.02	0.04	0.11	0.19	0.42	466
<i>C2</i>	0.02	0.02	0.00	0.00	0.01	0.02	0.05	8
<i>C3</i>	0.05	0.04	-0.01	0.02	0.04	0.07	0.18	471
<i>D</i>	0.08	0.05	-0.02	0.04	0.07	0.11	0.21	853
<i>E</i>	0.04	0.04	0.01	0.03	0.04	0.05	0.07	2
<i>RI</i>								
<i>ABC1</i>	0.25	0.14	-0.03	0.13	0.24	0.35	0.60	466
<i>C2</i>	0.09	0.05	-0.01	0.08	0.11	0.13	0.13	8
<i>C3</i>	0.09	0.07	-0.02	0.05	0.08	0.12	0.62	471
<i>D</i>	0.15	0.07	-0.05	0.09	0.15	0.21	0.36	853
<i>E</i>	0.49	0.23	0.33	0.41	0.49	0.57	0.65	2

A.2.3. Contribución por grupo a la entropía de cada zona

Para cada grupo obtenemos su contribución tomando la proporción de dicho grupo en la distribución $Q_{out,in}$, y luego obtenemos su contribución a la entropía²². Si tomamos por ejemplo el grupo *C2*, se obtiene para una zona j y día t lo siguiente

$$C2_{t,j}^{in,out} = -Q_{t,j,C2}^{in,out} \cdot \log_2(Q_{t,j,C2}^{in,out}) \quad (\text{A.2})$$

que corresponde a la cantidad que el grupo *C2* contribuye a la entropía $h_{in,out}$ (tanto outbound como inbound) en la zona j y día t . En lo que viene presentamos el promedio de cada zona durante el período analizado.

Notamos de la ecuación A.2 que el máximo teórico que un grupo puede contribuir a la segregación de una zona (por concavidad de entropía no necesariamente un grupo contribuye más por poseer una proporción más grande en las distribuciones)²³ viene dada por:

$$\frac{\partial}{\partial P}(-P \cdot \log_2(P)) = 0 \rightarrow P^* = \frac{1}{e} \sim 0.37 \quad (\text{A.3})$$

con P la proporción del grupo. Lo que dice este resultado es que el grupo que posee una

²² Es importante notar esto, dado que simplemente estamos obteniendo $-p \cdot \log_2(p)$ y no $\sum -p \cdot \log_2(p)$. En el primer caso p es una proporción (número entre cero y uno) mientras que en el segundo la suma se realiza sobre toda la distribución.

²³ Por ejemplo $Q_{in} = (1, 0, 0, 0, 0)$ el grupo *abc1* contribuye cero ($1 \cdot \log_2(1)$) a la suma total, mientras que $Q_{in} = (0.8, 0.2, 0, 0, 0)$ el grupo *acb1* contribuye $-0.8 \cdot \log_2(0.8) > 0$.

mayor proporción, no necesariamente es el que más aporta a la entropía (suma) sino el que posea una proporción más cercana a P^* . De esto a su vez, obtenemos que lo máximo que un grupo puede contribuir a la suma en la entropía es de $\frac{1}{e} \cdot \log_2(\frac{1}{e}) \sim 0.53$.

Notar que este resultado no necesariamente es el óptimo (que disminuye la segregación): la entropía es máxima cuando la distribución es uniforme y mínima cuando la distribución posee un solo grupo. De esta manera, dada la cardinalidad de la distribución de ingresos considerada en este trabajo ($|Q| = 5$), en nuestro caso la proporción óptima que debe poseer cada grupo, con el fin de disminuir la segregación es de 0.2, cuya entropía es $-0.2 \cdot \log_2(0.2) = 0.46$. De esta manera, si bien un grupo por un lado podría estar contribuyendo más a la suma de la entropía al tener una proporción cercana a P^* , al mismo tiempo, esto afecta a lo que otro grupo contribuye, lo que ocasiona que no se alcance la distribución óptima que disminuye más la segregación (máxima suma de entropía).

Lo anterior queda reflejado en la figura A.8 donde se puede visualizar el resultado de aplicar la función $-p \cdot \log_2(p)$ a diferentes proporciones (p) de un grupo cualquiera. Es importante notar que la gráfica (fig. A.8) no corresponde a una entropía propiamente tal, dado que para ello p debe ser una distribución, cosa que en nuestro análisis no lo es, sino que corresponde simplemente a una proporción de un grupo de referencia.

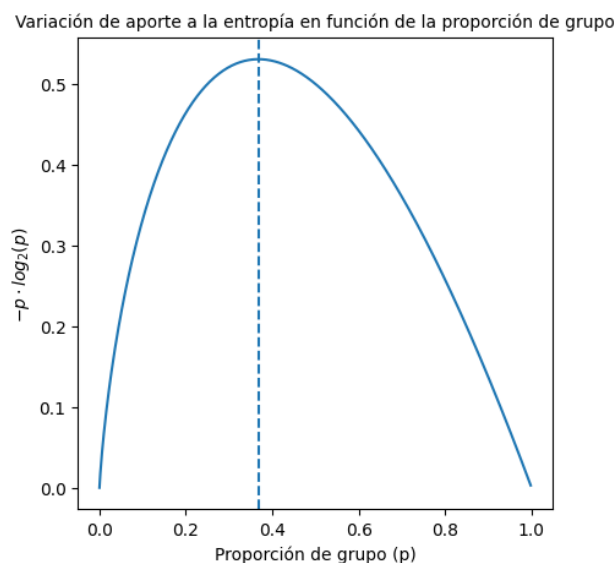


Figura A.8: Resultado de aplicar la función $-p \cdot \log_2(p)$ a diferentes proporciones de grupo p . En línea punteada se visualiza el máximo que se alcanza en $p = \frac{1}{e}$.

Un resultado importante que se visualiza de la figura anterior (fig. A.8), es que un grupo puede contribuir la misma cantidad a la suma de la entropía si es que posee una proporción baja o alta, digamos por ejemplo $p_{baja} \sim 0.2$ y alta $p_{alta} \sim 0.63$, que contribuyen lo mismo ~ 0.48 . De esta manera, en las distribuciones agregadas Q_{out} y Q_{in} , que un grupo represente al 20% de la distribución es lo mismo que otro represente al 60%, dado que ambos terminan pesando lo mismo en su contribución a la entropía. En otras palabras, que un sujeto se encuentre (Q_{out} por ejemplo) en promedio con un 20% con sujetos $ABC1$ tiene el mismo peso que se encuentre en un 63% con sujetos E . Esto quizá es una propiedad que quisiera evitarse, dado que se podría querer premiar ciertos encuentros entre sectores socioeconómicos más distintos entre sí ($ABC1$ y E por ejemplo), pero dicha aproximación queda fuera del

alcance de este trabajo.

Teniendo lo anterior en cuenta, podemos observar los resultados de desagregar las sumas de las entropías h_{out} y h_{in} , y ver su contribución por grupo a las distintas zonas de la RM (figuras A.10-A.15)²⁴. Primero comenzamos analizando la distribución geográfica del aporte de cada grupo socioeconómico a las distintas zonas tanto para h_{out} y h_{in} , y luego analizamos la semejanza que se obtiene entre los índices h_{out} y h_{in} (ver por ejemplo para el grupo de *ABC1* y *C2*, fig. A.10 y fig. A.13).

Para entender mejor las interacciones de los distintos grupos, asignamos el nivel socioeconómico de cada zona de acuerdo a la clasificación utilizada previamente (de acuerdo a *b*), con lo que obtenemos las distribuciones Q_{out} y Q_{in} por grupo (fig. A.9).

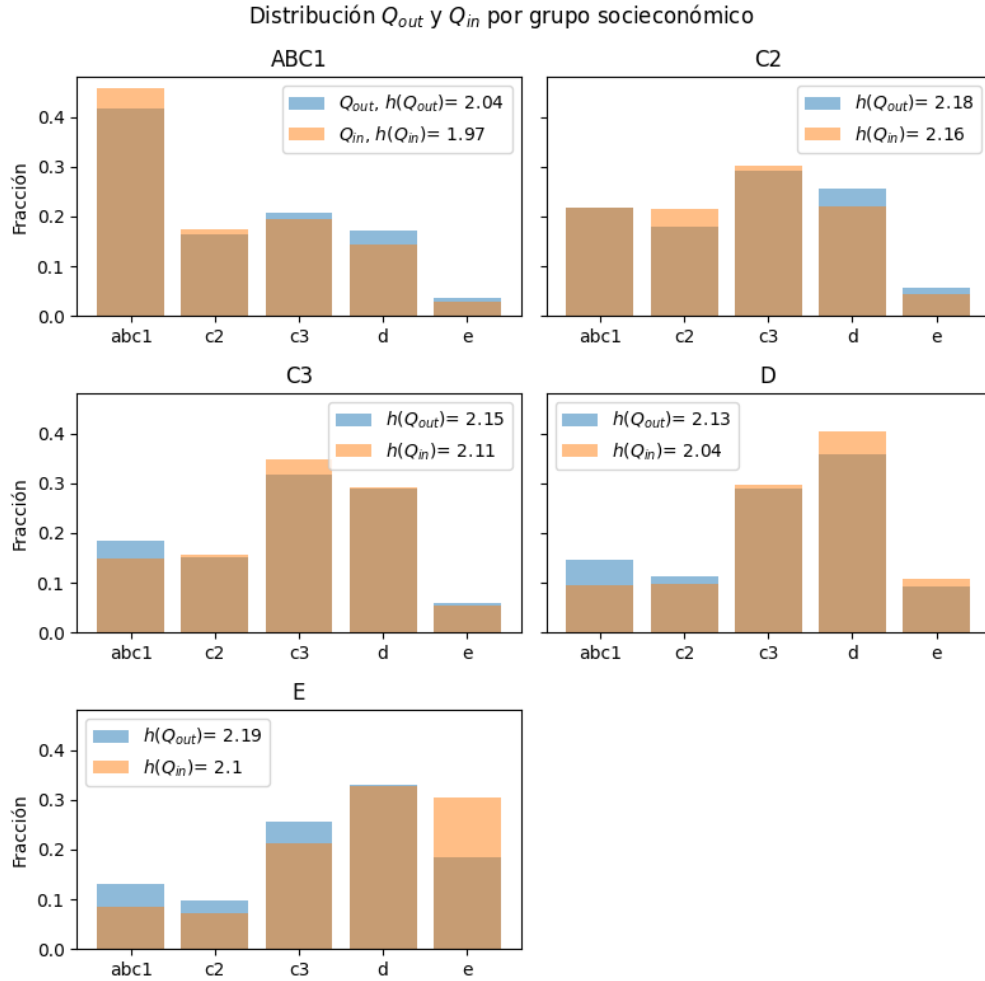


Figura A.9: Distribuciones Q_{out} y Q_{in} por grupo socioeconómico, utilizando la clasificación socioeconómica para cada zona de acuerdo al punto *b*).

Si observamos el aporte de los distintos grupos a las distintas zonas de la RM, lo que se visualiza es que tanto para h_{out} como h_{in} , cada grupo de sectores medios y bajos realiza la mayor contribución en promedio a sectores de su mismo nivel socioeconómico: medios (*C3–D*) a sectores medios y bajo (*E*) a sectores de bajos ingresos (ver figuras A.11, A.12, A.14

²⁴ Esto es análogo a lo realizado con las distribuciones Q_{out} y Q_{in} , donde obtenemos el aporte de cada grupo en $h_{out} = (h_{ABC1}, h_{C2}, h_{C3}, h_D, h_E)$ para cada zona y luego obtenemos su representación geográfica.

y A.15). La excepción a esta regla es el grupo $ABC1$ y $C2$ los cuales poseen distribuciones más uniforme a lo largo de la RM, contribuyendo a distintos lugares de manera más homogénea (ver figuras A.10 y A.13).

Previamente obtuvimos que la proporción que maximiza la contribución a la entropía es de $p = 0.37$ con la que se obtiene una contribución de 0.53. Luego, la razón por la que los grupos $C3$, D y E poseen una mayor densidad en sectores de ingresos medios y bajos, es debido a que interactúan en promedio en una proporción cercana a una 37% con dichos sectores (ver fig. A.9), mientras que su interacción con grupos $ABC1$ y $C2$ es bastante menor siendo inferior en ambos casos al 20%, lo cual equivale a una contribución máxima de $-0.2 \cdot \log_2(0.2) \sim 0.46$ en la entropía, y que es lo que se refleja en los mapas (figuras A.11, A.12, A.14 y A.15).

A su vez, para los grupos $ABC1$ y $C2$ se observa que si bien tienen una contribución mayor en sectores de su mismo nivel socioeconómico, también presentan cierta contribución en sectores de niveles socioeconómicos distintos a los propios. Para el caso del grupo $ABC1$, obtenemos que este grupo se relaciona mayoritariamente con personas de su mismo nivel, donde estas interacciones representan alrededor del 50% del total que tienen (ver fig. A.9). Al estar dicho porcentaje más alejado del óptimo que permite la máxima contribución (37%), esto ocasiona que en sectores de muy altos ingresos exista un cierto aclaramiento en los mapas (ver figuras A.10 y A.13).

Para el grupo $C2$, que corresponde al grupo con la contribución más homogénea de todos (fig. A.10 y fig. A.13), tenemos que esto se debe a que las interacciones de los individuos de este grupo son mucho más uniformes (ver fig. A.9), siendo el grupo que más interactúa con grupos distintos a sí mismo (de hecho interactuar más con otros grupos que consigo mismo).

Del análisis anterior tenemos que la función $-p \cdot \log_2(p)$ premia proporciones más mesuradas y cercanas al máximo teórico ($p = 0.37$), y si bien su análisis desglosado por grupo nos permite analizar cómo contribuye cada grupo, no indica por sí solo si un grupo integra o segregar más. Ejemplo de esto es el grupo $ABC1$ que si bien presenta una mayor contribución en áreas de altos ingresos, su contribución es relativamente homogénea a lo largo de la RM, no obstante, como se vio previamente dicho grupo es el que más segrega.

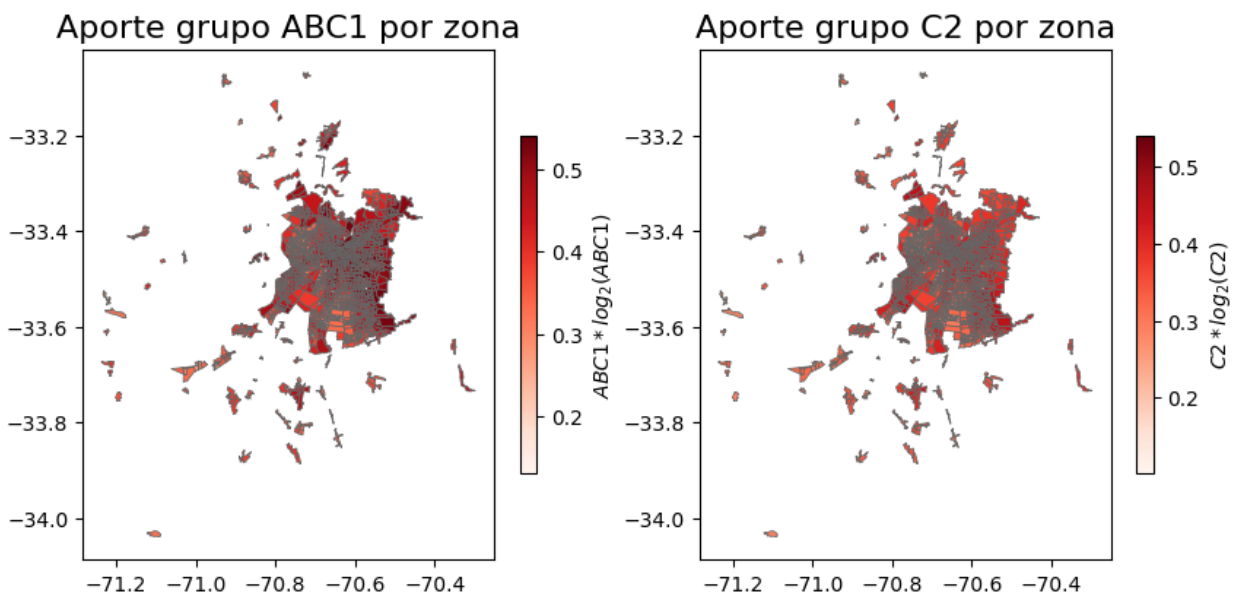


Figura A.10: Distribución regional del aporte de los grupos socioeconómicos $ABC1$ y $C2$ a h_{out} en cada zona censal.

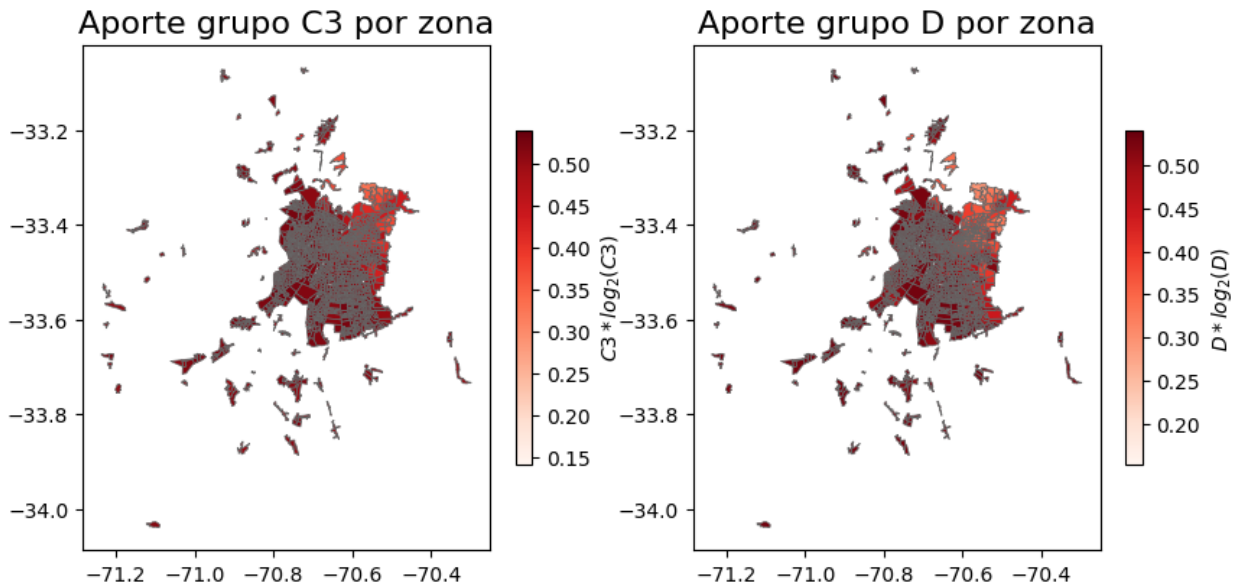


Figura A.11: Distribución regional del aporte de los grupos socioeconómicos $C3$ y D a h_{out} en cada zona censal.

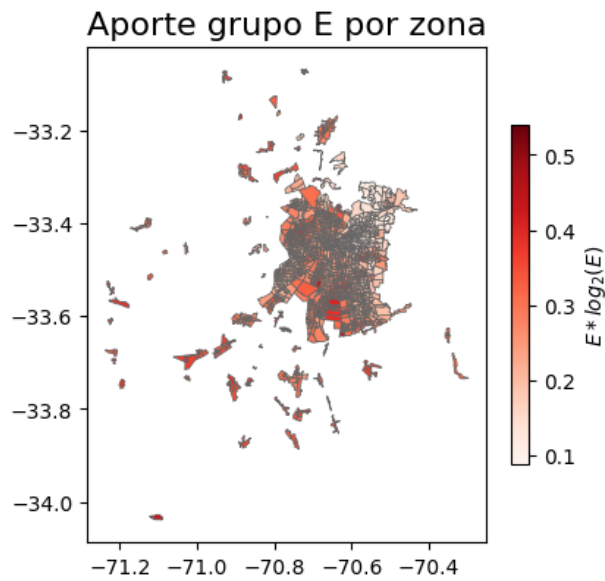


Figura A.12: Distribución regional del aporte del grupo socioeconómico E a h_{out} en cada zona censal.

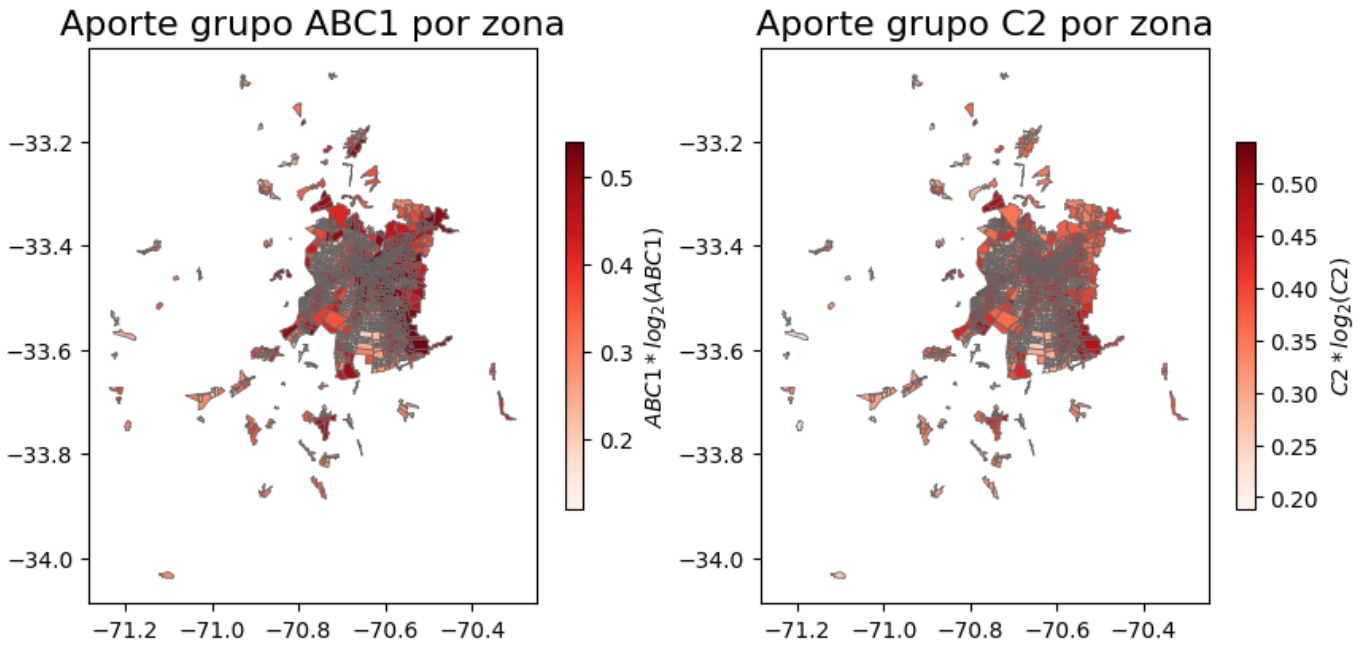


Figura A.13: Distribución regional del aporte de los grupos socioeconómicos $ABC1$ y $C2$ a h_{in} en cada zona censal.

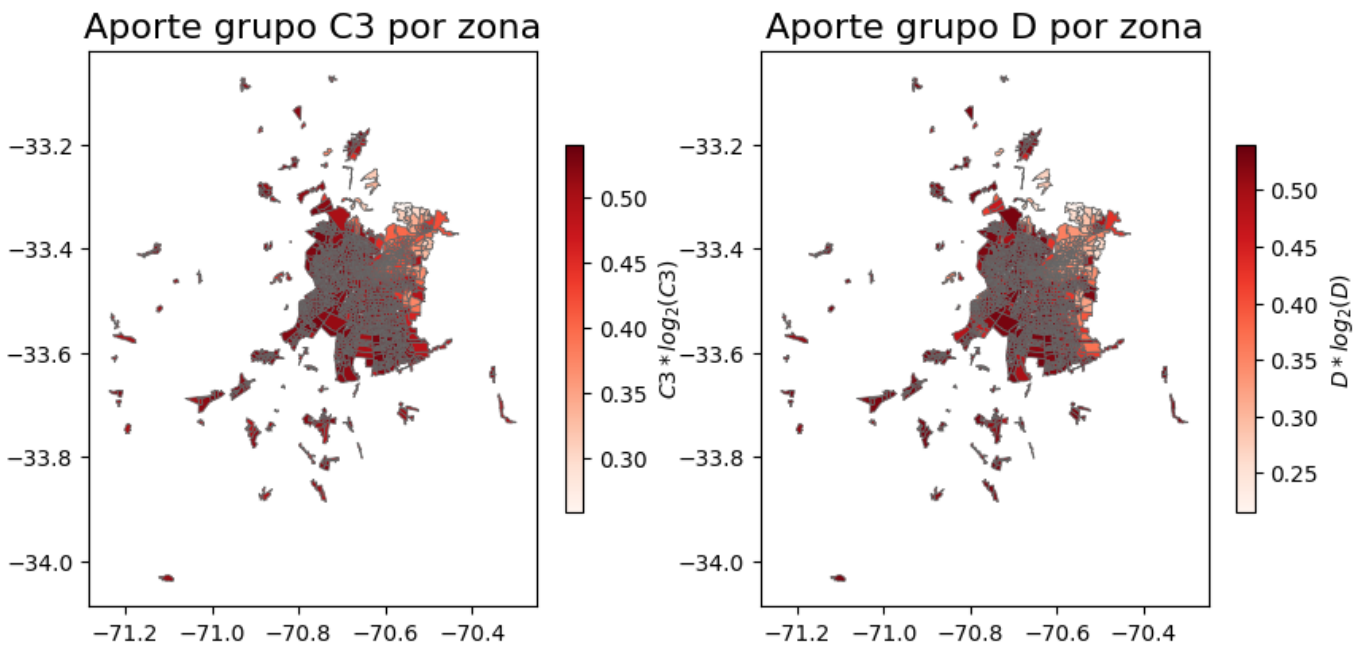


Figura A.14: Distribución regional del aporte de los grupos socioeconómicos $C3$ y D a h_{in} en cada zona censal.

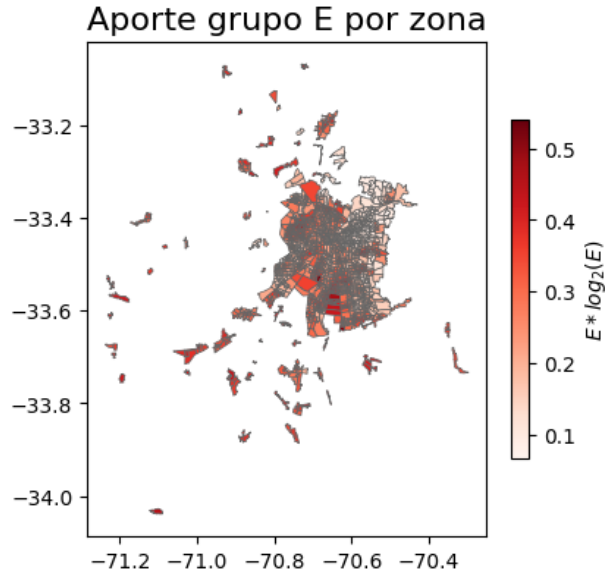


Figura A.15: Distribución regional del aporte del grupo socioeconómico E a h_{in} en cada zona censal.

Ahora analizamos la semejanza entre h_{out} y h_{in} . El que h_{out} y h_{in} sean similares, es la razón por la que los respectivos índices de segregación EI_{out} y EI_{in} sean a su vez, muy similares. El motivo de lo anterior lo podemos visualizar en la figura A.9, donde obtenemos la distribución promedio de llegada Q_{in} y de salida Q_{out} por grupo.

Lo que podemos visualizar en la figura A.9, es que las distribuciones agregadas Q_{in} y Q_{out} para cada zona y grupo socioeconómico son muy similares, siendo el caso en que menos se parecen en zonas del grupo E . El que ambas distribuciones sean muy similares se entiende de que en promedio las relaciones que establecen los individuos son con sujetos parecidos a los que llegan a la zona donde habitan: una persona del centro de Santiago digamos en un rango de ingresos $C2$ se mueve a otra zona donde interactúa con otras que en promedio son $C2$, y al mismo tiempo, otros individuos también $C2$ van a la zona de donde salió (habita) la primera persona.

Esto es claro para los grupos $ABC1$, $C3$ y D , los cuales en ambas distribuciones interactúan mayoritariamente con personas de su mismo nivel (ver fig. A.9). En particular para el grupo E este interactúa más con personas del grupo D , y en segundo lugar con personas de su mismo grupo, mientras que para personas del grupo $C2$ su mayor interacción es con personas del grupo $C3$, D y luego $C2 - ABC1$.

Lo anterior también se puede observar en la tabla A.3, donde se tiene que con excepción de los grupos $C2$ y E , el porcentaje promedio de individuos que visita una zona de un nivel socioeconómico distintos al propio está por debajo del 41%. La mayoría de las interacciones de individuos de los grupos $ABC1$, $C3$ y D son con sujetos de su mismo nivel socioeconómico, mientras que en el caso de los grupos $C2$ y E , las interacciones principales son con sujetos de un grupo socioeconómico distinto al de ellos.

Tenemos que de las distribuciones agregadas Q_{in} y Q_{out} (ver fig. A.9), y del porcentaje de individuos que sale a una zona de un nivel distinto al propio (tabla A.3 y fig. A.16) el grupo que más segrega es el grupo $ABC1$ mientras el que menos, es el grupo $C2$. El grupo $ABC1$ es el que menos se relaciona con sectores de un ingreso distinto (solo el 19% lo hace) y por lo mismo su distribuciones agregadas Q_{in} y Q_{out} son sesgadas a su mismo grupo, y de esto su

entropía es más baja. Caso contrario al grupo $C2$, donde este grupo es el que más se integra (se relaciona en un 63% con individuos de otros grupos) y sus distribuciones Q_{in} y Q_{out} son bastante más cercanas a una distribución uniforme y por lo mismo su entropía más elevada.

En cuanto a los grupos $C3$, D y E , tenemos que los grupos $C3$, D tienden a interactuar entre ellos (ver fig. A.9), donde en conjunto suman tanto para Q_{out} como Q_{in} más del 60% del total de interacciones que dichos grupos poseen: individuos $C3$ y D se relacionan en más de un 60% con individuos de dichos grupos (Q_{out}), y al mismo tiempo, a zonas de dichos grupos llegan más de un 60% individuos de estos grupos (Q_{in}). En referencia a los otros grupos, los grupos $C3$ y D , segregan menos que el grupo $ABC1$, pero más que el grupo $C2$ (ver tabla A.2), ya que poseen distribuciones más sesgadas que $C2$ (ver fig. A.9), y al mismo tiempo interactúan menos con individuos de otros niveles (tabla A.3).

Por último el caso del grupo E , dado que es un grupo con baja representatividad, los resultados que obtuvimos pueden estar sub-estimando la segregación que ahí se experimenta, en específico para EI_{out} , donde se obtienen un mínimo de -0.2 y un máximo de 0.2 , con lo cual se obtiene un promedio de cero.

Dado que nosotros capturamos movilidad y por ende, interacciones durante horario laboral, lo que puede estar detrás de esto es que los individuos ya sea por preferencias propias o por capacidades propias a cada grupo (nivel educacional, ingresos, facilidad/dificultad de movilidad, etc) forman cierto guetos por grupos socioeconómicos, los cuales agrupan ciertas características que son comunes y/o apetecibles al gueto que pertenecen. A priori podríamos proponer que estos guetos se componen por el grupo $abc1$ por sí solo, los grupos $c2 - c3 - d$ ya que poseen distribuciones y entropías parecidas y el grupo E que si bien se relaciona con grupos D y $C3$, su baja proporción para los otros grupo es clara (los demás grupos interactúan poco con personas del grupo E).

Tabla A.3: Estadísticos descriptivos por nivel socioeconómico de % de individuos que sale de su comuna y % de individuos que se dirige a zonas de un nivel socioeconómico distinto.

Índice	Media	std	min	25%	50%	75%	max	n° zonas
% Sale de comuna								
$ABC1$	33.82	8.13	2.36	28.6	33.12	38.29	63.11	466
$C2$	36.14	7.88	27.10	30.50	33.53	41.35	48.65	8
$C3$	35.11	9.16	8.47	29.28	34.52	40.85	70.89	471
D	33.55	10.90	0.00	26.39	33.36	39.96	75.12	853
E	39.56	9.04	33.17	36.36	39.56	42.75	45.95	2
% Sale de zona								
$ABC1$	19.04	11.26	1.98	10.26	15.39	24.40	59.05	466
$C2$	63.50	10.63	51.98	56.99	60.37	65.94	84.87	8
$C3$	40.10	10.42	12.81	32.70	39.59	46.33	75.59	471
D	29.28	10.24	2.63	22.65	28.10	36.24	71.06	853
E	56.59	0.76	56.05	56.32	56.59	56.86	57.13	2

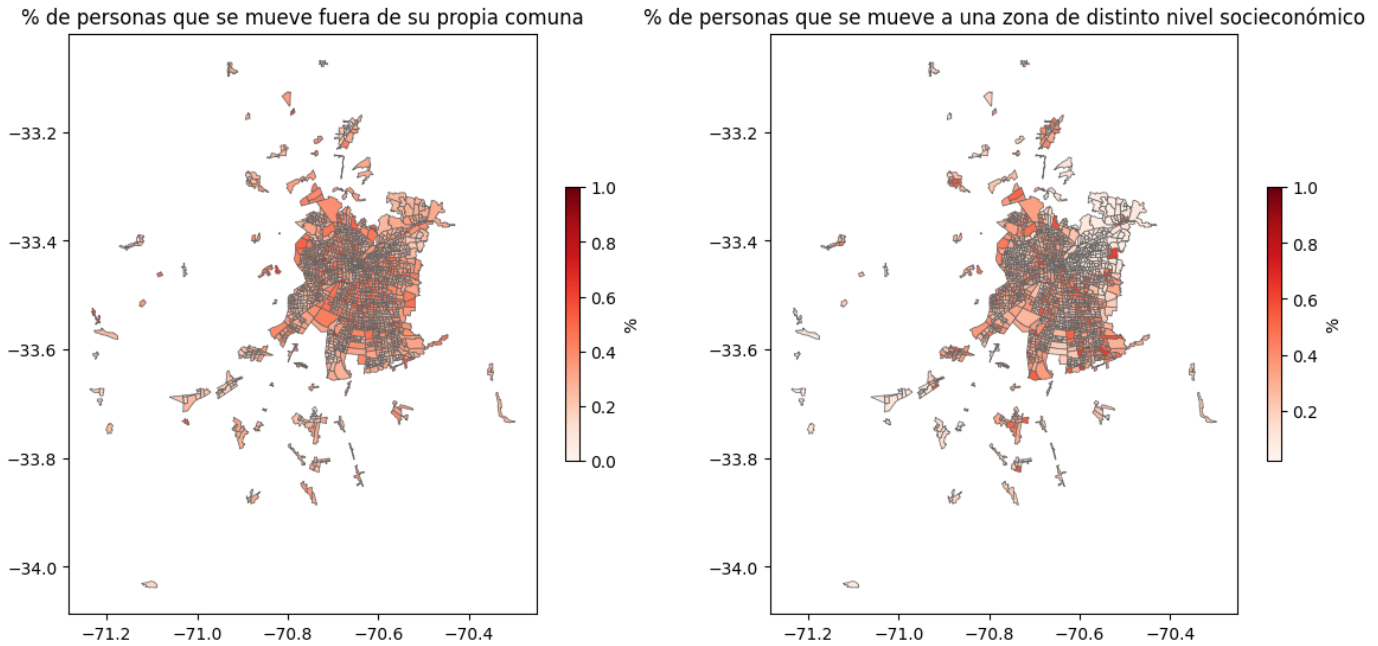


Figura A.16: Porcentaje de individuos que sale de una zona a otra zona de una comuna distinta a la suya, y porcentaje de individuos que sale de una zona de un determinado nivel de ingresos a otra zona de un nivel de ingresos distinto al propio.

A.2.4. Estadísticos descriptivos por grupo socioeconómico de algunas variables

A partir de la clasificación *b*) descrita al inicio de esta sección (clasificar una zona a partir del grupo que posee un mayor número de hogares en esa zona), obtenemos las siguientes tablas con estadísticos descriptivos por grupo socioeconómico.

Para mejorar la visualización de las tablas (tablas A.4, A.5, A.6, A.7, A.8), utilizamos la siguiente notación: *prom_ismt* (promedio índice ISMT), *Pueblo or.* (n° personas que pertenece a pueblo originario), *MCH* (mujeres con hijos), *AVB* (razón de áreas verdes buenas), *SPH* (superficie de parques por habitante) y *DC* (distancia al centro).

En cuanto las distintas variables presentadas, tenemos el siguiente análisis descriptivo:

1. Inmigrantes: La media de esta variable presenta un comportamiento creciente con el ingreso, siendo el que en promedio posee más el grupo *ABC1* y el que tiene menos el grupo *E*, donde además, cada grupo presenta una gran desviación estándar, existiendo zonas con cero inmigrantes.
2. Pueblo originario: Tenemos que en promedio la media es mayor en grupos de ingresos medios bajos (*C3* y *D*), y menor para sujetos de altos ingresos. Al igual que otras variables, existen una gran desviación entre zonas dentro de un mismo grupo, donde hay zonas con cero individuos pertenecientes a un pueblo originario.
3. Escolaridad: Esta variable es creciente con el ingreso. En promedio individuos de altos ingresos poseen 12 años de escolaridad, mientras que individuos del grupo *E* poseen un promedio cercano a 7. La desviación es menor a la de otras variables, donde a través de grupos, todos poseen una desviación menor a 1.09, siendo este valor el máximo y que lo posee el grupo *ABC1*.

4. Edad: Tenemos que a través de los grupos, la edad es relativamente similar, donde se posee un promedio de cercano a 35 años para todos los grupos, y una desviación cercana a 3 años.
5. Hijos: Tenemos que en promedio los grupos de ingresos medios bajos ($C3$ y D), son los que poseen una mayor cantidad de hijos. En cuanto a la desviación, no se observa un patrón a lo largo los ingresos, donde todos están cercanos al rango de 800-1000.
6. Jubilados: Tenemos que con excepción del grupo $C2$ que posee un promedio de 99 individuos jubilados, las demás zonas presentan un promedio cercano a los 300. En cuanto al grupo E , este posee en promedio 1 jubilado, no obstante, este grupo posee una baja representatividad (solo 2 zonas).
7. Desempleo: Tenemos que el desempleo es menor en grupos de ingresos altos y medios altos ($ABC1$ y $C2$), mientras que es mayor en sectores de ingresos medios bajos ($C3$ y D). A su vez, la desviación de zonas de ingresos medios bajos es mayor a la de ingresos altos. Estos últimos presentan una desviación promedio cercana a 38, mientras que sectores $C3$ una desviación de 48, y D una desviación de casi 60.
8. MCH: se comporta de manera similar al número de hijos, siendo grupos $C3$ y D los que poseen en promedio un mayor número de mujeres con hijos.
9. AVB: Esta variable presenta un comportamiento creciente con el nivel de ingresos, siendo el grupo $ABC1$ el cual posee en promedio cerca de 90 % de áreas verdes en buen estado, mientras que los grupos de ingresos bajos D solo un 60 %, y E un 50 %. A su vez, la desviación es menor en sectores de altos ingresos, por lo que sectores de altos ingresos tienden a poseer mejores áreas verdes que sectores de ingresos bajos.
10. SPH: Esta variable no presenta un patrón a lo largo de los ingresos, donde el grupo $C3$ es el que posee una mayor proporción promedio de superficie de áreas verdes a habitantes, y el que menos el grupo $C2$. Al mismo tiempo, el grupo $C3$ es el que posee la mayor desviación siendo muy superior a la de otros grupos, lo cual se traduce en que dichas zonas deben existir parques metropolitanos muy grandes, mientras que en otras zonas no deben existir áreas verdes (el mínimo es de cero en $C3$).
11. DC: En cuanto a esta variable, en promedio el grupo más cercano es el grupo $ACB1$, y el más lejano el grupo D , mientras que el grupo $C2$ posee un promedio similar al grupo $C3$. Por último, la desviación es similar entre grupos siendo cercana a 5 kilómetros, con excepción del grupo D que es de casi el doble ($\sim 9[Km]$).

Tabla A.4: Estadísticos descriptivos para grupo $ABC1$.

	prom_ismt	Personas	Inmigrantes	Pueblo or.	Escolaridad	edad	hijos	Jubilados	Desempleados	MCH	AVB	SPH	DC
count	466.00	466.00	466.00	466.00	466.00	466.00	466.00	466.00	466.00	466.00	450.00	448.00	466.00
mean	0.88	3439.97	395.14	175.14	12.62	36.85	2062.20	237.44	92.41	888.44	0.90	58.03	6.85
std	0.02	1293.69	323.60	115.58	1.09	3.93	1027.66	128.57	39.77	383.43	0.13	65.57	4.74
min	0.81	214.00	0.00	0.00	9.45	26.62	121.00	7.00	2.00	46.00	0.26	0.00	0.21
25 %	0.87	2600.00	194.00	94.25	11.87	33.95	1317.00	133.25	66.00	610.00	0.91	16.19	2.38
50 %	0.89	3306.00	310.00	157.50	12.82	37.35	1980.00	228.00	88.00	866.50	0.96	39.55	6.32
75 %	0.90	4099.00	477.50	232.75	13.38	39.94	2734.75	325.00	114.00	1163.25	0.98	79.86	10.22
max	0.92	11031.00	2438.00	1494.00	15.04	46.78	6811.00	684.00	292.00	2897.00	1.00	680.15	25.93

Tabla A.5: Estadísticos descriptivos para grupo *C2*.

	prom_ismt	Personas	Inmigrantes	Pueblo or.	Escolaridad	edad	hijos	Jubilados	Desempleados	MCH	AVB	SPH	DC
count	8.00	8.00	8.00	8.0	8.00	8.00	8.00	8.00	8.00	8.00	7.00	7.00	8.00
mean	0.85	3455.50	313.75	257.0	10.46	30.05	1837.62	99.00	94.25	878.75	0.65	8.96	9.71
std	0.01	1426.22	429.52	137.0	0.95	2.52	809.98	56.90	38.81	393.22	0.16	7.77	4.20
min	0.82	964.00	39.00	0.0	9.34	26.73	524.00	27.00	22.00	240.00	0.54	1.03	4.18
25 %	0.84	2961.50	70.25	233.5	9.93	28.15	1271.00	70.25	71.75	635.00	0.54	3.34	5.97
50 %	0.85	3593.50	95.00	264.0	10.10	29.87	1937.00	84.50	107.50	903.50	0.62	5.77	10.23
75 %	0.86	4315.50	327.75	314.5	10.83	31.43	2403.25	114.75	124.00	1147.75	0.67	13.25	12.93
max	0.86	5123.00	1205.00	478.0	12.04	34.52	2834.00	212.00	132.00	1358.00	0.98	22.70	15.23

Tabla A.6: Estadísticos descriptivos para grupo *C3*.

	prom_ismt	Personas	Inmigrantes	Pueblo or.	Escolaridad	edad	hijos	Jubilados	Desempleados	MCH	AVB	SPH	DC
count	471.00	471.00	471.00	471.00	471.00	471.00	471.00	471.00	471.00	471.00	439.00	439.00	471.00
mean	0.81	3950.33	206.33	374.23	10.12	35.57	2707.52	245.13	132.39	1157.68	0.62	87.27	9.37
std	0.02	1323.64	401.03	165.59	0.75	3.94	841.76	123.45	48.53	364.06	0.18	1360.56	5.60
min	0.76	10.00	0.00	0.00	7.88	25.72	3.00	0.00	0.00	2.00	0.04	0.00	0.18
25 %	0.80	3093.50	48.00	262.50	9.66	32.82	2167.00	151.00	97.00	913.00	0.51	4.78	5.84
50 %	0.81	3965.00	88.00	355.00	10.10	35.73	2744.00	233.00	134.00	1165.00	0.69	11.43	8.99
75 %	0.82	4746.00	171.00	463.00	10.68	38.71	3242.50	331.50	162.50	1403.00	0.71	26.37	11.71
max	0.86	9518.00	3437.00	1141.00	12.71	48.10	6446.00	661.00	378.00	2662.00	0.98	28522.58	53.29

Tabla A.7: Estadísticos descriptivos para grupo *D*.

	prom_ismt	Personas	Inmigrantes	Pueblo or.	Escolaridad	edad	hijos	Jubilados	Desempleados	MCH	AVB	SPH	DC
count	853.00	853.00	853.00	853.00	853.00	853.00	853.00	853.00	853.00	853.00	718.00	715.00	853.00
mean	0.76	3693.77	186.63	443.88	8.72	36.14	2917.85	291.38	142.49	1100.40	0.59	32.59	9.97
std	0.02	1350.17	290.27	221.57	0.81	3.14	1062.94	134.25	59.95	397.31	0.21	48.85	9.86
min	0.68	42.00	0.00	0.00	6.44	24.52	16.00	0.00	0.00	9.00	0.04	0.00	0.29
25 %	0.74	2836.00	51.00	289.00	8.13	34.21	2263.00	188.00	104.00	853.00	0.46	8.25	4.41
50 %	0.76	3747.00	101.00	427.00	8.61	36.57	2995.00	289.00	141.00	1130.00	0.62	16.25	7.30
75 %	0.77	4624.00	198.00	578.00	9.26	38.26	3632.00	387.00	184.00	1380.00	0.74	34.32	11.29
max	0.82	8904.00	3459.00	1275.00	12.06	44.48	5922.00	719.00	368.00	2206.00	0.98	387.98	64.79

Tabla A.8: Estadísticos descriptivos para grupo *E*.

	prom_ismt	Personas	Inmigrantes	Pueblo or.	Escolaridad	edad	hijos	Jubilados	Desempleados	MCH	AVB	SPH	DC
count	2.00	2.00	2.0	2.0	2.00	2.00	2.00	2.0	2.0	2.0	2.00	2.00	2.00
mean	0.62	26.50	0.0	0.0	6.98	35.82	15.50	1.0	0.0	5.0	0.48	0.39	6.10
std	0.01	7.78	0.0	0.0	0.43	5.91	3.54	0.0	0.0	0.0	0.09	0.06	1.65
min	0.61	21.00	0.0	0.0	6.68	31.64	13.00	1.0	0.0	5.0	0.41	0.35	4.93
25 %	0.61	23.75	0.0	0.0	6.83	33.73	14.25	1.0	0.0	5.0	0.44	0.37	5.51
50 %	0.62	26.50	0.0	0.0	6.98	35.82	15.50	1.0	0.0	5.0	0.48	0.39	6.10
75 %	0.62	29.25	0.0	0.0	7.13	37.91	16.75	1.0	0.0	5.0	0.51	0.41	6.68
max	0.62	32.00	0.0	0.0	7.29	40.00	18.00	1.0	0.0	5.0	0.54	0.43	7.27

Anexo B. Comparación de formulaciones de medidas de segregación experiencial

A partir de la entropía en trabajos previos (Thiel, 1985-Frankel & Volij, 2011-Boary Giannonez, 2023) se usa para definir el *índice de entropía* (o información mutua) para una zona j como

$$H_j = H(\bar{Q}) - \sum_{i \in j} H(Q_i) * W_i \quad (\text{B.1})$$

donde Q_i es la distribución de cada zona en j y W_i el peso correspondiente de i dentro de j . El definir el índice de segregación de esta manera (ec. B.1), por concavidad de H se tiene que esto garantiza que el índice sea siempre positivo, esto es, si recordamos la definición de concavidad, tenemos que tomando dos distribuciones p_1, p_2 y una constante $0 \leq \lambda \leq 1$ se tiene

$$H(\lambda p_1 + (1 - \lambda)p_2) \geq \lambda H(p_1) + (1 - \lambda)H(p_2) \quad (\text{B.2})$$

lo cual se puede generalizar a un número n de distribuciones

$$H\left(\sum_i^n \lambda_i p_i\right) \geq \sum_i^n \lambda_i H(p_i) \quad (\text{B.3})$$

lo cual es equivalente a la ecuación B.1. Este punto, nos sirve para analizar en la siguiente sección las distintas formas en que la ecuación B.1 podría definirse, esto es, si la distribución base con la se compara \bar{Q} es considerada de una manera distinta (distribución uniforme por ejemplo) a lo que plantea la ecuación B.3, o en su defecto evaluar el término de la sumatoria y el peso W_i se agregan tal cual esta en B.1, y se reemplaza por un ponderado dentro de la función de entropía H (esto es $H(\sum_{i \in j} Q_i * W_i)$ en B.1)

Dado lo anterior, usamos el índice de entropía para definir nuestras dos medidas de segregación experiencial: segregación inbound y outbound. Simplificando notación definimos primero para outbound:

$$EI_j^{out} = A_j^{out} - B_j^{out} \quad (\text{B.4})$$

donde A_j^{out} corresponde a la entropía de una distribución base o ideal con la se quiere comparar, y B_j^{out} corresponde a un ponderado que da cuenta de las personas con las que se encuentran sujetos de j . A su vez, definimos para inbound:

$$EI_j^{in} = A_j^{in} - B_j^{in} \quad (\text{B.5})$$

donde A_j^{in} corresponde a la entropía de una distribución base o ideal con la se quiere comparar, y B_j^{in} corresponde a un ponderado que da cuenta de las personas que se encuentran en j . Por último para segregación residencial definimos

$$RI_j = A_j^{res} - B_j^{res} \quad (\text{B.6})$$

donde A_j^{res} corresponde a la entropía de la distribución base considerada, y B_j^{res} a la entropía de la distribución residencial de j . En este caso B_j^{res} queda *fijo*, con $B_j^{res} = h_j^{res}$ y h_j^{res} definido de la siguiente manera

$$h_j^{res} = - \sum_{g \in \Sigma} Q_{j,g}^{res} \cdot \log_2(Q_{j,g}^{res}) \quad (\text{B.7})$$

donde $Q_{j,g}^{res}$ es la distribucion de ingresos en j .

Uno quisiera poder comparar los indices con distintas distribuciones base como el promedio regional por ejemplo, lo cual seria considerar $A_j^{in} = A_j^{out} = \bar{Q}$, o definir B_j^{out} y B_j^{in} de manera agregada (lado izquierdo en B.3) o desagregada (lado derecho en B.3), no obstante si las distribuciones que se consideran no cumplen con la relacion planteada en la ecuacion B.3, la positividad del indice no esta garantizada, pudiendo eventualmente ser negativo para algunos zonas.

A pesar de esto, consideremos dos formas en que podemos definir B^{out} y B^{in} . Primero para B^{in} tenemos

- Definimos B_j^{in} para una zona j de manera agregada como:

$$h_{t,j}^{in} = - \sum_{g \in \Sigma} Q_{t,j,g}^{in} \cdot \log_2(Q_{t,j,g}^{in}) \quad (\text{B.8})$$

que corresponde a la entropia sobre la distribucion promedio ponderada de individuos que hay en un lugar j un dia t .

- Definimos B_j^{in} para una zona j de manera desagregada como:

$$\tilde{h}_{t,j}^{in} = - \sum_{g \in \Sigma} \sum_{k \in J} W_{k,j,t}^{in} * Q_{g,k}^{res} \cdot \log_2(Q_{g,k}^{res}) \quad (\text{B.9})$$

que corresponde a la entropia promedio ponderada de las distribuciones de origen de cada grupo de individuos que visita j , donde $W_{k,j,t}^{in} = \frac{f_{kjt}}{\sum_{k \in J} f_{kjt}}$ es el numero de personas que va de k a j el dia t , nomalizada por el total de personas que va a la zona j el dia t .

A su vez, tenemos para B_j^{out} :

- Definimos B_j^{out} para una zona j de manera agregada como:

$$h_{t,j}^{out} = - \sum_{g \in \Sigma} Q_{t,j,g}^{out} \cdot \log_2(Q_{t,j,g}^{out}) \quad (\text{B.10})$$

que corresponde a la entropia de la distribucion ponderada de individuos con la que sujetos de j se encuentran un dia t .

- Definimos B_j^{out} para una zona j de manera desagregada como:

$$\tilde{h}_{t,j}^{out} = - \sum_{g \in \Sigma} \sum_{k \in J} W_{j,k,t}^{out} * Q_{t,g,k}^{in} \cdot \log_2(Q_{t,g,k}^{in}) \quad (\text{B.11})$$

$W_{j,k,t}^{out} = \frac{f_{jkt}}{\sum_{k \in J} f_{jkt}}$, que corresponde al numero de personas que van de j a k el dia t normalizado por el total de personas que sale de j

Tabla B.1: Casos considerados para comparar distintas formas en que planteamos el indice de entropia.

Variable	Comentario
$A_j^{out} = A_j^{in} = \bar{Q}$ $B_j^{out} = h_j^{out}$ $B_j^{in} = \tilde{h}_j^{in}$	Caso 1: Consideramos como distribucion base la distribucion regional, y los indices de entropia de manera agregada
$A_j^{out} = A_j^{in} = \bar{Q}$ $B_j^{out} = \tilde{h}_j^{out}$ $B_j^{in} = \tilde{h}_j^{in}$	Caso 2: Consideramos como distribucion base la distribucion regional, y los indices de entropia de manera desagregada
$A_j^{out} = A_j^{in} = Q_j$ $B_j^{out} = \tilde{h}_j^{out}$ $B_j^{in} = \tilde{h}_j^{in}$	Caso 3: Consideramos como distribucion base la distribucion propia de cada zona, y los indices de entropia de manera agregada
$A_j^{out} = A_j^{in} = Q_j$ $B_j^{out} = \tilde{h}_j^{out}$ $B_j^{in} = \tilde{h}_j^{in}$	Caso 4: Consideramos como distribucion base la distribucion propia de cada zona, y los indices de entropia de manera desagregada

Ahora, considerando todo lo anterior, evaluemos empiricamente tomar distintos valores y agregaciones sobre A_j^{in} , A_j^{out} , B_j^{out} y B_j^{in} , los que resumimos en la tabla B.1

A priori para cada uno de los casos propuestos en la tabla B.1 tenemos lo siguiente: primero notamos que para cada uno de los casos 1-4, por concavidad de h , y dado como definimos A, B en cada caso, la ecuacion B.3 no se satisface necesariamente, por lo que el indice de segregacion puede ser negativo en algunos casos evaluados.

Por otro lado, para los casos (1) y (2), se tiene que la distribucion base considerada ($A^{in,out}$) es constante, por lo que el ordenamiento (segregacion) viene dado por $B^{in,out}$. De esto, las diferencias en la segregacion medida EI agregada y la segregacion medida desagregada $\tilde{E}I$ viene dado por h y \tilde{h} , donde de esto mismo y por concavidad de h (ecuacion B.3) se tendra que EI será a lo menos tan segregado como $\tilde{E}I$.

En este mismo sentido, si queremos realizar un analisis de integracion y ver la diferencia entre la segregaciones $EI^{in,out}$ y RI , si la distribucion base A es constante e igual en todas las métricas (cosa que debe ser para que los indices sean comparables) se tiene que la diferencia se traduce a diferencias en la entropia B ($EI^{out,in} - RI = (A^{out,in} - B^{out,in}) - (A^{res} - B^{res}) = -B^{out,in} + B^{res}$).

B.1. Resultados de casos evaluados

A continuacion observamos los resultados de obtener los indices de entropia a partir de lo propuesto en la tabla B.1. De los resultados tenemos los siguientes puntos:

1. En cuanto la diferenciacion entre definir el indice de manera agregada o desagregada

tenemos: 1) se cumple el que por concavidad el agregado es siempre mayor al desagregado (ver. B.4).

2. El índice de segregacion experiencial inbound definido de manera desagregada, no es efectivo para capturar cambios en el tiempo de la segregacion (ver B.6). Esto es visible en la definicion de este indice (ver ec. B.9) donde si derivamos con respecto al tiempo

$$\frac{\partial}{\partial t} \tilde{h}_{t,j}^{in} = \frac{\partial}{\partial t} \left(- \sum_{g \in \Sigma} \sum_{k \in J} W_{k,j,t}^{in} * Q_{g,k}^{res} \cdot \log_2(Q_{g,k}^{res}) \right) = - \sum_{g \in \Sigma} \sum_{k \in J} Q_{g,k}^{res} \cdot \log_2(Q_{g,k}^{res}) * \frac{\partial W_{k,j,t}^{in}}{\partial t} \quad (\text{B.12})$$

se aprecia que este indice solo logra capturar variaciones en el flujo de personas de un lugar a otro, por lo que si en promedio el flujo es relativamente constante, el indice no deberia variar a traves del tiempo, lo cual es lo que se obtiene en figura la B.6.

3. Las entropias definidas de manera agregada o desagregada pese a no establecer el mismo orden sobre todas las zonas (segregacion) tienden a tener una distribucion parecida, siendo mayor (por concavidad) en el caso agregado, lo cual ocasiona que el promedio de la forma agregada se encuentre más cercano a la entropia regional (fig. B.3). Esto sugiere, que la forma agregada puede estar sub-dimensionando la segregacion real de un lugar (lo cual se puede deber a falta de detalle que se produce al agregar la distribucion), o en su defecto, que la version desagregada este sobre estimando la segregacion real de una zona (ver B.1,B.3).
4. Otras relaciones de la segregacion medida tanto de manera agregada como desagregada, son consistentes entre ambos casos: la relacion frente a los ingresos es la misma (ver por ejemplo fig. B.4, B.6,B.4), o su evolucion en el tiempo es similar (con la excepcion explicada previamente).

De lo anterior, dado que los resultados son relativamente similares, la formulación en el documento se hará de forma agregada, dado que esta logra capturar la variación en el tiempo.

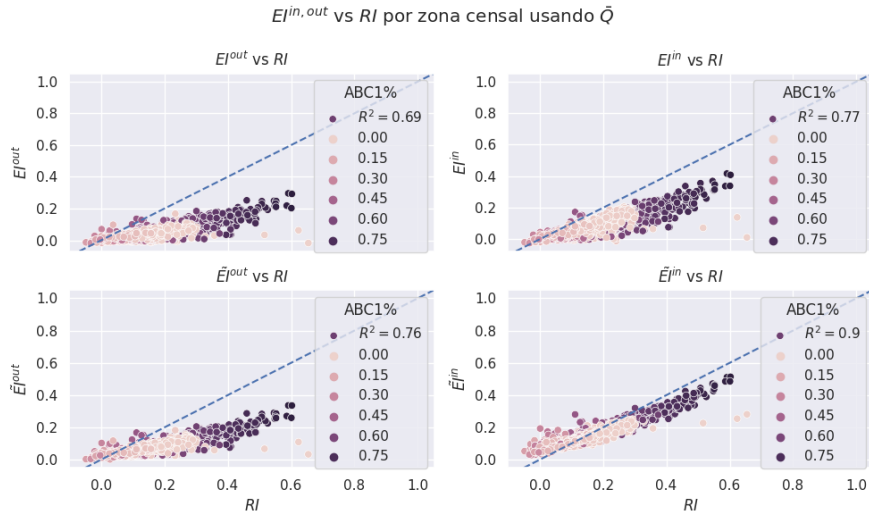


Figura B.2: Segregación experiencial en su forma agregada y desagregada frente a la segregacion residencial utilizando como entropía base A, la entropía de \bar{Q} .

$E^{out,in}$ y $\tilde{E}^{out,in}$ usando \bar{Q}

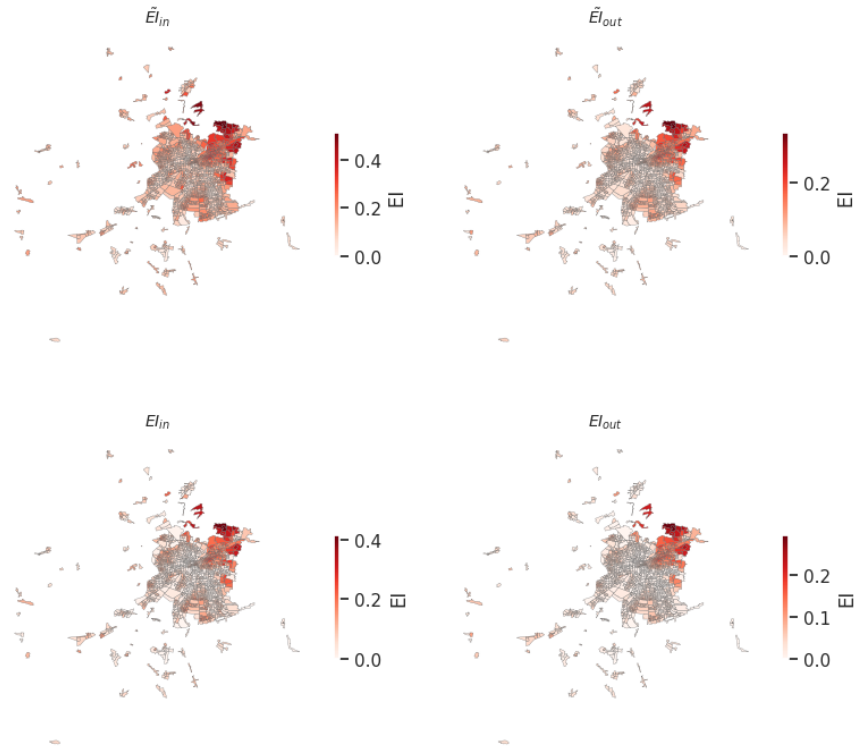


Figura B.1: Segregación experiencial en su forma agregada y desagregada frente a la segregación residencial utilizando como entropía base A , la entropía de \bar{Q} .

$E^{out,in}$ y $\tilde{E}^{out,in}$ usando \bar{Q}

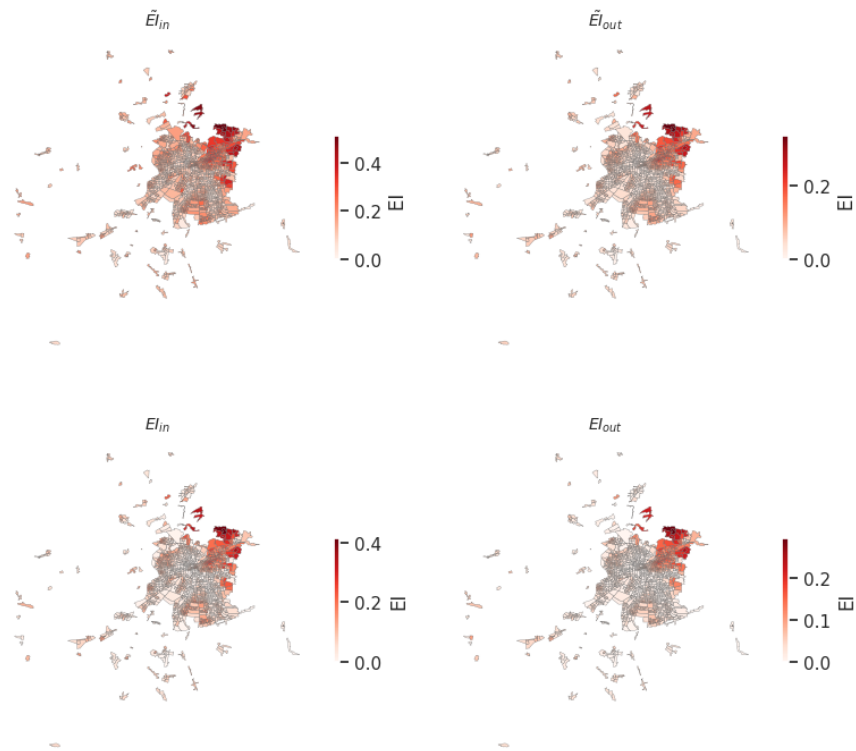


Figura B.3: Distribución de entropías.

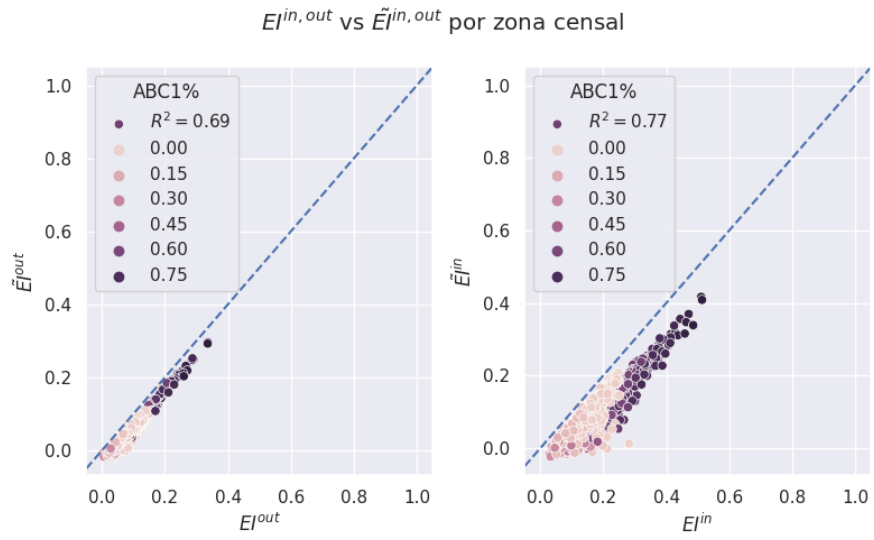


Figura B.4: Segregación experiencial outbound e inbound en sus forma desagregada versus agregada.

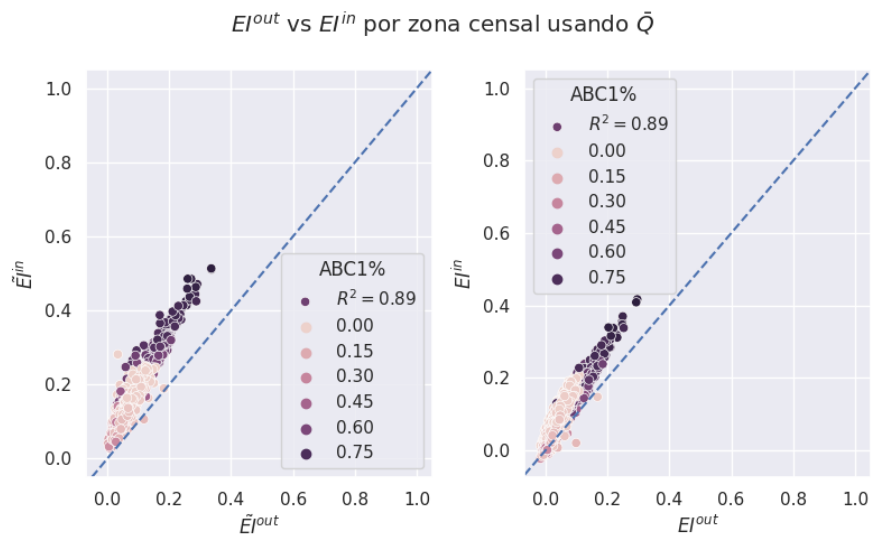


Figura B.5: Segregación experiencial inbound vs segregación outbound, agregada y desagregada.

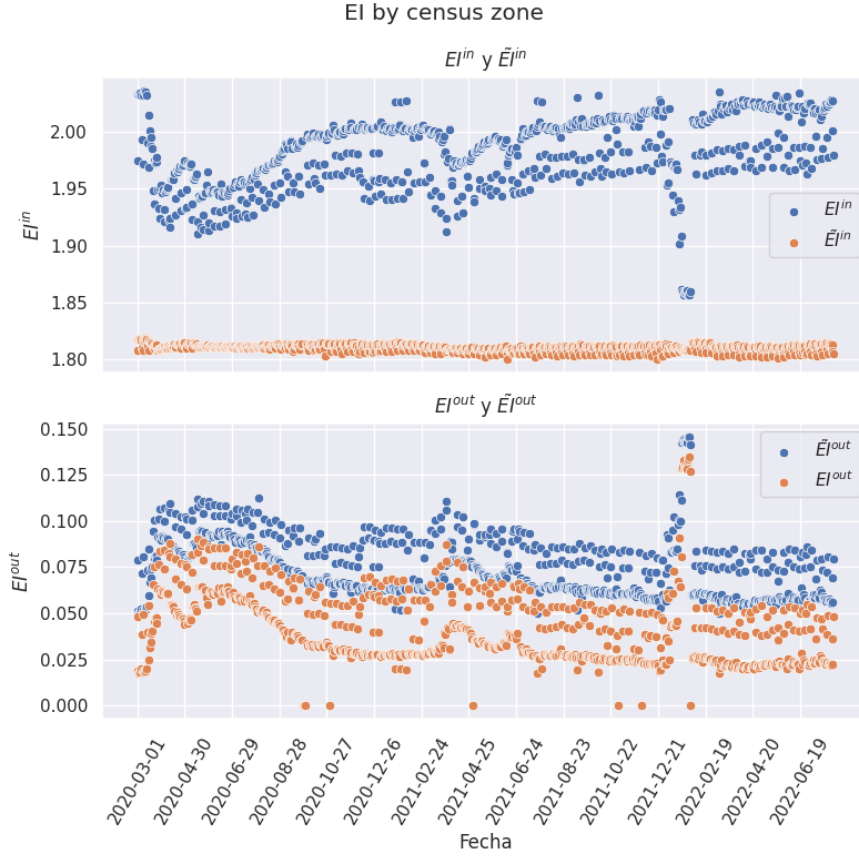


Figura B.6: Evolución temporal de los índices planteados de forma agregada y desagregada.

Anexo C. Sobre los datos

C.1. Datos

C.1.1. Representatividad de datos GPS

Para evaluar la representatividad que poseen los datos de GPS en relación a la población real, se estima el número de individuos en cada zona censal que poseen los datos de GPS por zona censal por medio de

$$\hat{N}_{c(i)} = \frac{\sum_j \sum_t f_{ijt}}{|T|} \quad (\text{C.1})$$

Donde $\hat{N}_{c(i)}$ corresponde a la cantidad de individuos estimados en zona censal c que posee un individuo i . Notar que $\hat{N}_{c(i)}$, corresponde a la suma de los lugares que visitaron individuos en $c(i)$, en promedio, durante los T días de análisis. De esta manera, sabiendo el número de individuos que salieron de i , tenemos el número de individuos que son representados por los datos GPS para la zona i . De esto último y de los datos censales, se estima el factor λ para cada zona, que corresponde al número de individuos que un individuo de los datos GPS debe representar, y se define como

$$\lambda_i = \frac{n^\circ Personasi}{N_{c(i)}} \quad (C.2)$$

donde $n^\circ Personasi$ corresponde al número de personas que habitan en i obtenido de datos censales.

Notar que si este factor λ es muy alto se tiene un problema de representación, dado que un solo individuo en los datos GPS estaría representando a un gran número de individuos de la vida real, lo cual haría que el modelo sobreestime o subestime características importantes para medir segregación. De esta manera, buscamos valores de λ lo más cercanos a 1 (el ideal). En el caso de este estudio, no se consideraron para el análisis todas aquellas cuyo λ fuera mayor a 30.

C.2. Selección de datos para caracterización de individuos y zonas censales

Para extraer una caracterización tanto a nivel de persona como de zona censal, se utilizan tres conjuntos de datos: datos censales a nivel de persona y manzana, y datos ISMT de la RM.

Los datos censales a nivel de persona para la RM, se componen por una matriz de 6848480 filas y 48 columnas. Cada fila representa a una persona encuestada, indicando como caracterización un código de persona y código de manzana censal respectivo. A su vez, las 48 columnas corresponden a preguntas que caracterizan tanto a la persona como el lugar donde vive. Para obtener una identificación por zona censal, utilizando el id de la comuna, del distrito censal, del área censal y zona, se construye el id a nivel de zona censal (11 números). Luego, de la matriz original, en conjunto con el id de zona censal creado, se seleccionan 7 columnas correspondientes a las preguntas²⁵:

- Pregunta 8: sexo (hombre o mujer)
- Pregunta 9: edad.
- Pregunta 12: lugar de nacimiento.
- Pregunta 16: se considera perteneciente a un pueblo indígena u originario.
- Pregunta 17: trabajó durante la semana pasada.
- Pregunta 19: total hijos/as nacidos vivos.
- Escolaridad: años de escolaridad.

En relación a los datos de manzana censal para la RM, estos se componen de 51070 filas, y 59 columnas. Las filas representan a cada manzana censal de la RM y cada columna el agregado en cada variable en la respectiva manzana censal. Al igual que con los datos censales a nivel de persona, se construye el id de zona censal, y luego, en este caso se seleccionan 9 columnas del total (59):

- Personas: número de personas en manzana censal.

²⁵ Ver Manual de usuario de la base de datos del censo de población y vivienda 2017, 2018.

- Hombres: número de hombres en manzana censal.
- Mujeres: número de mujeres en manzana censal.
- Edad 0-5: número de personas entre 0-5 años en manzana censal.
- Edad 6-14: número de personas entre 6-14 años en manzana censal.
- Edad 16-64: número de personas entre 16-64 años en manzana censal.
- Edad 65-: número de personas de 65 o más años en manzana censal.
- Inmigrantes: número de personas inmigrantes en la manzana censal.
- Pueblo: número de personas pertenecientes a pueblos originarios en manzana censal.

Finalmente los datos ISMT de la RM se componen de 1865 filas y 27 columnas. Cada fila corresponde a una zona censal, y las columnas tienen distintas caracterizaciones a nivel de zonas censales. De este conjunto de datos se seleccionan todas sus filas y 11 columnas:

- Comuna: código de comuna.
- nom_cmn: nombre de comuna.
- ID_MANZENT2: id de zona censal.
- uf2018: valor en unidades de fomento (UF) de m^2 en zona censal.
- prom_ismt: promedio de zona censal en índice ISTM.
- ABC1, C2, C3, D, E: número de hogares en cada grupo socioeconómico.
- NSE: total de hogares.

Tabla C.1: Estadísticas descriptivas de los datos disponibles a nivel de zona censal.

Variable	Media	Des. est	min	25 %	50 %	75 %	max
Zona censal							
Tamaño m	0.03	0.02	0.01	0.02	0.02	0.03	0.28
Tamaño Area m^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Características socioeconómicas							
Valor uf_{2018} m^2	24.94	17.08	0.0	14.90	20.33	29.87	101.52
promedio ISMT	0.80	0.05	0.60	0.75	0.79	0.84	0.92
ABC1	228.61	292.80	0.0	26.00	84.00	343.00	1594.00
C2	144.36	105.60	0.0	58.00	122.00	209.0	673.00
C3	297.14	158.48	0.0	176.00	280.00	397.0	1094.0
D	311.75	202.93	1.0	135.00	308.00	454.00	961.00
E	70.89	64.56	0.00	17.00	53.00	108.00	356.00
Características demográficas							
n° personas	3662.62	1362.04	10.0	2766.00	3644.00	4551.00	11031.00
n° hombres	1775.97	696.84	0.0	1328.00	1750.00	2212.00	10080.00
n° mujeres	1884.13	694.27	0.00	1428.00	1884.00	2343.00	5443.00
n° personas edad entre 0-5 años	257.32	135.96	0.00	165.00	239.00	327.00	1175.00
n° personas edad entre 6-14 años	390.27	229.33	0.00	229.00	350.00	516.00	1622.00
n° personas edad entre 15-64 años	2552.93	976.34	0.0	1919.00	2517.00	3193.00	10441.00
n° personas edad entre 65 o más años	378.37	201.48	0.00	218.00	367.00	515.00	1000.00
n° inmigrantes	241.35	338.49	0.00	61.00	132.00	284.00	3459.00
n° personas pueblo indígena	352.14	215.35	0.00	196.00	319.00	480.00	1494.00
Años de Escolaridad promedio zona	10.06	1.82	6.44	8.60	9.70	11.12	15.04
Edad promedio zona	36.12	3.66	24.52	33.68	36.47	38.68	48.10
n° Hijos	2618.85	1070.80	3.00	1889.00	2650.00	3335.00	6811.00
n° Jubilados	262.05	133.18	0.0	158.00	251.00	354.00	719.00
n° Desempleados	125.70	56.88	0.00	85.00	120.00	163.00	378.00
n° Mujeres con Hijos	1051.36	404.30	2.00	786.00	1066.00	1332.00	2897.00
n°mujeres/n°hombres	0.51	0.02	0.08	0.50	0.51	0.52	0.57
Razón de áreas verdes buenas	0.68	0.22	0.03	0.53	0.71	0.91	0.99
n° viviendas	1260.91	461.97	6.00	971.75	1251.00	1520.25	4301.00
Áreas Verdes							
Poblacion total	6861287.33	6909539.91	1550.00	2282436.00	4242214.00	8841469.50	39015360.00
Población parque $P_{POBP}ARQ$	0.013	0.011	0.00	0.00	0.01	0.01	0.07
CAPCOM $P}AR$	0.03	0.06	0.00	0.04	0.09	0.08	1.09
superficie de parque /n° habitantes	54.51	705.46	0.00	8.07	18.09	44.07	28522.57
Mall							
Número	19.16	10.92	1.00	9.75	19.50	28.25	37.00
Número estacionamientos	2148.66	1860.63	0.00	415.00	1500.00	3650.00	5000.00
Número de tiendas	152.21	109.68	13.00	63.00	117.00	243.00	411.00
Número tiendas anclas	3.45	2.30	0.00	1.50	3.00	5.00	8.00
n° Cine	0.70	0.46	0.00	0.00	1.00	1.00	1.00