



**Trabajo 2 geomarketing: Microsimulación espacial a zona censal de  
adultos mayores que reciben pensión solidaria para el Gran Santiago**

**Pedro Pablo Silva Antilef**

Profesor Cátedra: Ricardo Crespo

Santiago – Chile  
2020

# TABLA DE CONTENIDO

<b>1</b>	<b>Introducción</b>	<b>1</b>
1.1	Antecedentes generales . . . . .	1
1.2	Área de estudio . . . . .	1
1.3	Objetivos . . . . .	2
1.3.1	Objetivo general . . . . .	2
1.3.2	Objetivos específicos . . . . .	2
1.4	Metodología y herramientas utilizadas . . . . .	3
1.4.1	Metodología . . . . .	3
1.4.2	Herramientas utilizadas . . . . .	3
<b>2</b>	<b>Marco Teórico</b>	<b>4</b>
2.1	Descripción de Datasets . . . . .	4
2.1.1	Encuesta de Caracterización Socioeconómica Nacional año 2017 . . . . .	4
2.1.2	Censo de Población y Vivienda año 2017 . . . . .	4
2.2	Microsimulación . . . . .	4
2.2.1	Microsimulación Espacial . . . . .	4
2.2.2	Weighting algorithm: Iterative Proportional Fitting . . . . .	5
<b>3</b>	<b>Desarrollo</b>	<b>6</b>
3.1	Preprocesamiento . . . . .	6
3.1.1	Encuesta Casen . . . . .	6
3.1.2	Censo de Población y Vivienda . . . . .	7
3.2	Microsimulación . . . . .	8
<b>4</b>	<b>Análisis de resultados</b>	<b>9</b>
<b>5</b>	<b>Conclusiones</b>	<b>11</b>
	<b>Referencias bibliográficas</b>	<b>12</b>
	<b>Anexos</b>	<b>13</b>

## ÍNDICE DE TABLAS

Tabla 3.1 Discretización de variables para encuesta CASEN. Fuente: Elaboración propia a partir de datos de encuesta CASEN . . . . .	7
Tabla 3.2 Discretización y descripción de variables para Censo de población y vivienda. Fuente: Elaboración propia a partir de datos de encuesta Censo 2017 . . . . .	8

## ÍNDICE DE ILUSTRACIONES

Figura 1.1	Mapa Gran Santiago . . . . .	2
Figura 4.1	Microsimulacion . . . . .	9

# **CAPÍTULO 1. INTRODUCCIÓN**

## **1.1 ANTECEDENTES GENERALES**

La evolución de las ciudades ha presentado diferencias socio-espaciales en su estructura desde sus principios y primeros ejemplos. Las primeras ciudades de Asia no solo presentaban calles de uso intensivo de algún tipo de actividad como comercial, artesanal y de funciones administrativas, sino también tuvieron barrios "altos y bajos" en términos de estratos socioeconómicos. El mismo fenómeno se presentó en ciudades de Europa medieval, japonesas o chinas, también segregadas por motivos religiosos o étnicos (Borsdorf, 2002).

Por eso no es de extrañar que en Latinoamérica exista un fenómeno parecido donde ciudades como Santiago de Chile, presentan una estructura urbana fragmentada, donde la clase social define la ubicación geográfica de los grupos socioeconómicos (Borsdorf, 2002).

Existieron varias épocas de desarrollo urbano en las ciudades Latinoamericanas donde destaca el proceso o "fase más reciente del desarrollo urbano en América Latina (1970 hasta hoy): la ciudad fragmentada", así denominado por (Borsdorf, 2003).

Uno de los hitos que marcó el desarrollo residencial fue la promulgación de la Ley de Venta por pisos de 1937, que permitió las edificaciones en altura y normó el régimen de copropiedad inmobiliaria. Bajo este amparo el desarrollo residencial aumentó en Santiago de Chile, un claro ejemplo de esto es el sector de Fernando Castillo Velasco vieron un crecimiento acelerado de desarrollo residencial con una particularidad: se trataba de familias de intelectuales y artistas que copartían tendencias políticas y culturales, aumentando el desarrollo residencial de este sector de Santiago y la fragmentación socioeconómica de los habitantes de la capital. El mismo fenómeno se repite para varias ciudades de Chile (Hidalgo, 2005).

## **1.2 ÁREA DE ESTUDIO**

El área de estudio corresponde a la conurbación "Gran Santiago", el cual corresponde a las comunas de la Provincia de Santiago, sumadas a las comunas de San Bernardo y Puente Alto, todas pertenecientes a la Región Metropolitana (RM). Esta metrópolis es la capital de la RM y del extenso país de Chile y tiene una población de 6.119.984 según el Censo de población y vivienda del año 2017, siendo la ciudad más populosa del país, lo que la convierte en el centro económico más importante del país.

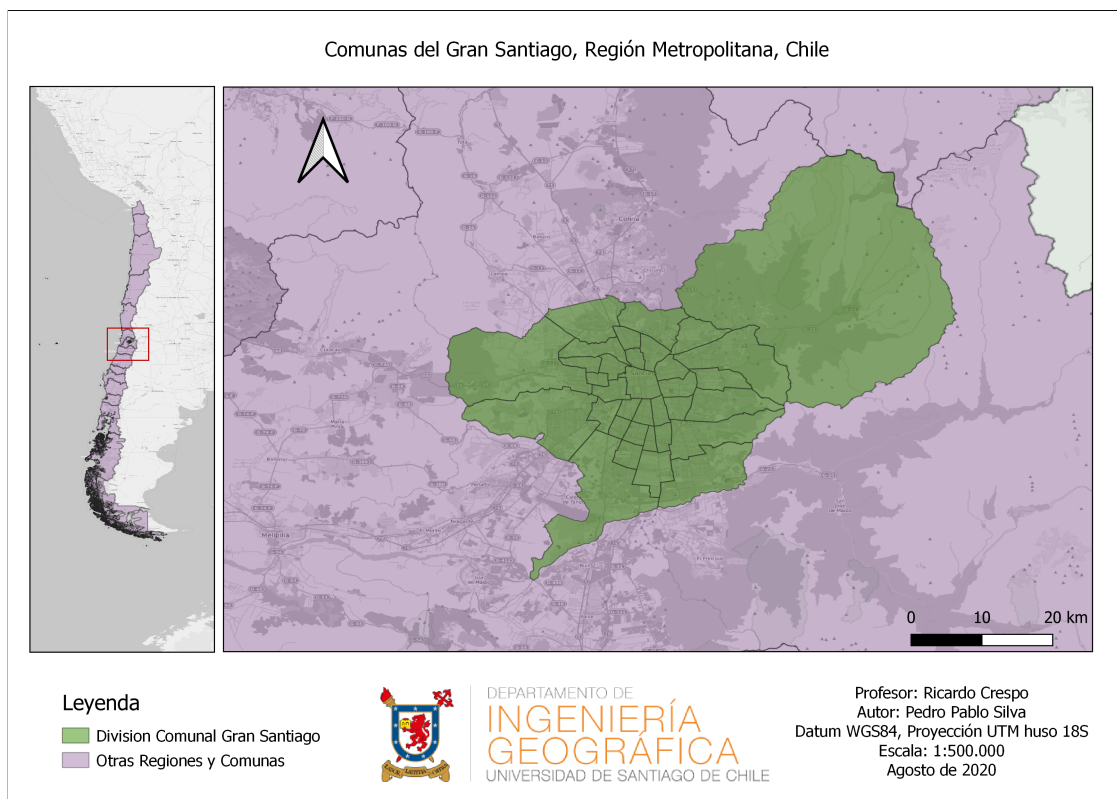


Figura 1.1: Cartografía Gran Santiago.  
 Fuente: Elaboración propia con datos del INE, 2020.

## 1.3 OBJETIVOS

### 1.3.1 Objetivo general

Realizar una microsimulación espacial a nivel de zona censal de adultos mayores que reciben pensión solidaria (total o complementaria) para el Gran Santiago utilizando información pública: Encuesta CASEN 2017 y Censo de Población y Vivienda 2017.

### 1.3.2 Objetivos específicos

1. Explorar encuesta CASEN 2017 para encontrar las variables objetivo a microsimular.
2. Preprocesar los datos para ser ingresados como input a la microsimulación.

3. Microsimular la variable objetivo y desplegar los resultados.
4. Analizar el resultado de la microsimulación, identificando patrones sustentados con referencias.

## **1.4 METODOLOGÍA Y HERRAMIENTAS UTILIZADAS**

### **1.4.1 Metodología**

En primera instancia se analizaron las variables disponibles en la encuesta CASEN para el estudiar los beneficiarios de pensión solidaria en personas mayor a 60 años. Luego de seleccionar las variables que se utilizarán, se realiza el preprocesamiento de los datos con el fin de obtener la estructura necesaria (y estricta) que deben tener los datos de input en el modelo de microsimulación. También se concatenan los códigos de censales para obtener el geocódigo único que permite realizar la unión entre los resultados de la microsimulación y su representación espacial explícita. Después de realizar la microsimulación, los resultados se representan en una cartografía con la simbología adecuada para representar el fenómeno estudiado. Finalmente se realiza un análisis de los resultados, en términos espaciales según información obtenida de indicadores y literatura relacionada.

### **1.4.2 Herramientas utilizadas**

Los datos de la encuesta CASEN 2017 y del Censo de Población y Vivienda 2017 fueron cargados en una base de datos PostgreSQL, desde donde fueron recuperadas las variables necesarias en el software estadístico RStudio (Basado en R). En este último programa se realizó el preprocesamiento de los datos, la microsimulación en sí a través de la librería "rakeR" y su posterior despliegue en una cartografía.

## **CAPÍTULO 2. MARCO TEÓRICO**

### **2.1 DESCRIPCIÓN DE DATASETS**

#### **2.1.1 Encuesta de Caracterización Socioeconómica Nacional año 2017**

La Encuesta de Caracterización Socioeconómica Nacional (CASEN) del año 2017 es una iniciativa encargada por el Ministerio de Desarrollo Social y desarrollada por el Centro de Encuestas y Estudios Longitudinales de la Pontificia Universidad Católica de Chile y asesorada por diferentes entidades como el Instituto Nacional de Estadística y la Comisión Económica Para América Latina y el Caribe (Social, 2018). Tiene como fin caracterizar los hogares chilenos en términos socioeconómicos a nivel comunal.

#### **2.1.2 Censo de Población y Vivienda año 2017**

El Censo de Población y Vivienda año 2017 (Censo 2017) es un proceso liderado por el Instituto Nacional de Estadística, el cual movilizó a más de 546.000 participantes, donde se contabilizó un total de 17.574.003 de personas y un total de 6.499.355 viviendas en la totalidad del territorio nacional (INE, 2018). El Censo es un importante esfuerzo para caracterizar demográficamente la población en función de la edad, escolaridad, tipo de vivienda y hogar, su dinámica de migración, etc. Los resultados se pueden encontrar a nivel persona, hogar, vivienda, zona censal, distrito, área, comuna, región y nacional, siendo zona censal, el mínimo nivel representable en una cartografía.

### **2.2 MICROSIMULACIÓN**

#### **2.2.1 Microsimulación Espacial**

La microsimulación espacial es una muy útil herramienta para el análisis territorial de la población, al desagregar (espacialmente) información preexistente, generando resultados fácilmente interpretables que permitan identificar patrones y tendencias en la población. (Ballas, 2005). Esto vuelve a la microsimulación, una metodología útil para generar información relevante en la toma de decisiones.



### 2.2.2 Weighting algorithm: Iterative Proportional Fitting

El procedimiento de ajuste iterativo o IPF (por sus siglas en inglés) fue introducido en 1950 por Deming & Stephan (1994) para estimar la probabilidad de celda en tablas de contingencia sujeto a ciertas restricciones marginales. Su convergencia y sus propiedades estadísticas han sido investigadas por diferentes autores y sometido a diferentes métodos de evaluación (Ruschendorf, 1995). En el caso de la microsimulación, el IPF es usado para distribuir individuos a zonas y su representación matemática es la siguiente (Lovelace, 2018).

Para la regresión

$$w(i, z, t + 1) = w(i, z, t) \cdot \frac{const_t(z, c, ind(i, c))}{\sum_{j=1}^{n_{ind}} w(j, z, t) \cdot I(ind(j, c) = ind(i, c))} \quad (2.1)$$

Donde:

- *inds*: Matriz de dos dimensiones que contiene variables categóricas de cada individuo.
- *const<sub>t</sub>*: Matriz que contiene el conteo restrictivo de datos.
- *w*: Función de peso o ponderación.
- *c*: Variable o característica.
- *z*: Zona desagregada.
- *i, j*: Individuos.
- *t*: Paso o iteración.
- *I(x)*: Función indicador, donde el valor 1 es TRUE y 0 en otro caso.

## CAPÍTULO 3. DESARROLLO

### 3.1 PREPROCESAMIENTO

El procedimiento de ajuste proporcional iterativo con el que funciona la librería rakeR de R, es exigente en cuanto a la estructura de los datos de entrada, siendo esta la gran dificultad de implementación de este algoritmo.

#### 3.1.1 Encuesta Casen

El preprocesamiento de la encuesta CASEN parte con la imputación de valores no observados para la variable escolaridad, los cuales fueron obtenidos mediante una regresión lineal y predicción mediante el método de mínimos cuadrados. Luego se procede a discretizar las variables objetivo que el algoritmo comparará con las variables del Censo, estas variables son Edad, Escolaridad, Sexo y "Ayuda". Esta última corresponde a la variable a microsimular.

La variable edad fue separada en personas menores a 60 años y personas de 60 años o más, ya que 60 años es la edad mínima para jubilar en Chile en el caso del sexo femenino. En cuanto a la escolaridad se dividió en 4 categorías, personas sin estudios, personas entre 1 y 8 años, tiempo en el que se debería completar la educación primaria o básica en Chile, personas con 8 o 12 años de educación que corresponden al rango de años en que se debería realizar la educación media o secundaria en Chile, finalmente personas con más de 12 años de educación que corresponden a personas con algún grado de estudios universitarios. La variable sexo ya estaba representada de forma binaria y tomaba valor 1 si el encuestado era un hombre y 2 en caso de que el encuestado fuera mujer, por lo que se cambió el valor a un string, donde sex\_f corresponde a personas del sexo femenino y sex\_m a personas del sexo masculino.

La variable a microsimular corresponde a y26\_1a y y26\_1a o "Mes pasado Pensión Básica Solidaria de Vejez" y Mes pasado Jubilación de Vejez con Aporte Previsional Solidario" respectivamente, donde serán seleccionadas las personas que contestaron de manera positiva alguna de estas dos preguntas y agrupadas en la nueva variable "ayuda", la cual será utilizada como target para la microsimulación. El resumen de las nuevas variables se puede observar en la tabla 3.1.

Finalmente se dividió el dataframe en 34 sub dataframes correspondientes a la información de cada comuna.

Variable	Valor	Descripción
Edad	edad_menor_60	Personas que tienen menos de 60 años al momento de la encuesta
	edad_mayor_60	Personas que tienen 60 años o más al momento de la encuesta
Escolaridad	esc_0	Personas con 0 años de escolaridad
	esc_1_8	Persona con más de 0 años de escolaridad y 8 o menos años de escolaridad
	esc_8_12	Persona con más de 8 años de escolaridad y 12 o menos años de escolaridad
	esc_mas_12	Personas con más de 12 años de escolaridad
Sexo	sex_f	Personas de sexo femenino
	sex_m	Personas de sexo masculino
Ayuda	2	Personas que reciben pensión solidaria parcial o total

Tabla 3.1: Discretización de variables para encuesta CASEN. Fuente: Elaboración propia a partir de datos de encuesta CASEN

### 3.1.2 Censo de Población y Vivienda

El preprocesamiento de la base de datos del Censo de Población y Vivienda corresponde a realizar un conteo agrupado por comuna de las personas con las mismas características que para la encuesta CASEN: personas con edad menor a 60 años y personas con 60 años o más (variable P09 en el diccionario censal), personas con sexo femenino o sexo masculino (p08 en diccionario censal). Para la variable de escolaridad, la cual es de variable categórica y corresponde a la variable P15 en el diccionario censal del INE y que corresponde al "Nivel del curso más alto aprobado" donde se consideraron las personas tenían como curso aprobado más alto Sala cuna o jardín infantil, Prekínder, Kinder dentro de la categoría esc\_0 o personas con 0 años de escolaridad, a las personas que aprobaron Especial o diferencial, Educación básica o primaria o preparatoria (sistema antiguo) en el rango de escolaridad de 1 a 8 años de estudio, a las personas que completaron sus estudios científicos humanistas, técnico profesional o humanidades (sistema antiguo), Técnica comercial, industrial/normalista (sistema antiguo) como personas que estudiaron entre 8 y 12 años y por último a las personas que estudiaron técnico superior (1-3 años), Profesional (4 años o más) y estudios de postgrados dentro de las personas con 12 años de escolaridad o más. El resumen de los atributos se puede encontrar en la tabla 3.2. Finalmente se divide el dataframe del censo en 34 dataframes, correspondientes a cada comuna del GS.

Variable	Valor	Descripción
Edad	edad_menor_60	Personas que tienen menos de 60 años al momento de la encuesta
	edad_mayor_60	Personas que tienen 60 años o más al momento de la encuesta
Nivel del curso más alto aprobado	esc_0	Personas con sala cuna o jardín infantil, prekínder o kínder aprobado
	esc_1_8	Persona con educación especial o diferencial, educación básica o primaria o preparatoria (sistema antiguo) aprobado
	esc_8_12	Personas con estudios científicos humanistas, técnico profesional o humanidades (sistema antiguo), Técnica comercial, industrial/normalista (sistema antiguo) aprobado
	esc_mas_12	Personas con técnico superior (1-3 años), Profesional (4 años o más) o estudios de postgrados finalizados
Sexo	sexo_f	Personas de sexo femenino
	sexo_m	Personas de sexo masculino

Tabla 3.2: Discretización y descripción de variables para Censo de población y vivienda. Fuente: Elaboración propia a partir de datos de encuesta Censo 2017

### 3.2 MICROSIMULACIÓN

Para la microsimulación (y como se explicó en la sección 1.4 y 2.2) se utilizará el software estadístico R, específicamente la librería rakeR, la cual utiliza el algoritmo de ponderación ajuste proporcional iterativo. Por parte del Censo, se utilizaron las variables edad, sexo y años de escolaridad para hacer el match con estas mismas variables en la encuesta Casen, las cuales ponderan la correlación entre estas variables y la variable objetivo "ayuda". Se mantiene la columna de identificación del censo, la cual corresponde al código único territorial (CUT) para cada zona censal.

Finalmente con el CUT, se procede a mapear la información para realizar una interpretación espacialmente explícita de los resultados obtenidos, para esto se utiliza una capa de información territorial recuperada del INE. El análisis de esta representación se realizará en la sección 4.

Se probaron varios métodos para la obtención de quiebres para la representación de la información, como método de jenks, geométrico, equals, pretty, entre otros que ofrece la librería Cartography de R, siendo los cuantiles la mejor representación más adecuada para su representación debido a que lograba explicar las diferencias espaciales entre las comunas.

## CAPÍTULO 4. ANÁLISIS DE RESULTADOS

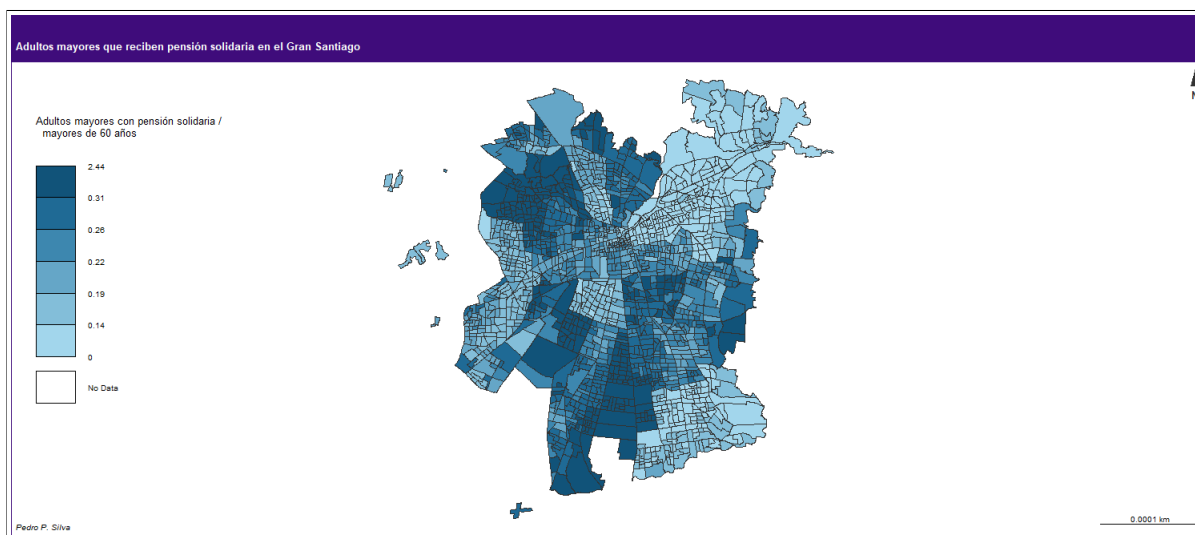


Figura 4.1: Microsimulación adultos mayores con pensión solidaria, Gran Santiago.  
Fuente: Elaboración propia, 2020.

La representación utilizada como resultado de la microsimulación muestra una clusterización de adultos mayores con pensión solidaria de manera parcial o total, siendo el sector sur y el sector nor-poniente de la capital aquellos que concentran este tipo de personas. El sector oriente, específicamente las comunas de Providencia, Las Condes, Vitacura y Lo Barnechea presentan un bajo nivel ayudas a las pensiones por parte del estad. Esto es congruente con distintos estudios que muestran que estas comunas presentan niveles de calidad de vida superiores al resto del GG, de hecho, estás comunas son las 4 comunas con mayor índice de calidad de vida urbana de Chile (ICVU), desarrollado por el Institutos de Estudios Urbanos y Territoriales de la Universidad Católica de Chile junto con la Cámara Chilena de la Construcción (CChC, 2019). Puente Alto, San Miguel y la Reina (N°5 en el ranking ICVU) también presentan bajos niveles de entrega de ayuda a adultos mayores, pero en menor medida que las comunas anteriormente descritas.

Los clusters de entrega de beneficios de pensión de vejez se encuentran en las periferias de la capital. El sector sur, en las comunas de La Pintana, La Granja, San Ramón, Lo Espejo, Renca, Cerro Navia, Pudahuel Norte, el sector sur de San Bernardo, sector norte de huechuraba (específicamente en La Pincoya), algunos sectores de San Joaquín, entre otros. Los cuales presentan una gran cantidad de personas que reciben los beneficios de pensión solidaria. Además estos territorios se encuentran en el rango inferior del ICVU, donde las condiciones laborales, el ambiente de negocios, condiciones socio-culturales y la vivienda y entorno son las principales variables que influyen en que estas comunas se encuentren en el umbral de este ranking (CChC, 2019).

En términos generales los resultados obtenidos son parecidos a los expuestos por (Crespo,

2020a), donde la distribución por ingresos para el GS son muy parecidos a los obtenidos en este trabajo, donde el sector sur y nor-poniente perciben una menor cantidad de ingresos y el sector nor-oriente de la capital ingresos más altos. En temas de salud también se puede observar una correlación entre los ingresos y el padecimiento de ciertas enfermedades y acceso a centros de salud (Crespo, 2020b), fenómeno que presenta patrones parecidos a los encontrados en esta microsimulación.

Puente alto se presenta como una zona interesante, ya que pese a que presenta bajos niveles de ingresos (Crespo, 2020a), también presenta una gran variabilidad respecto a las condiciones socioeconómicas de sus habitantes, por lo que resulta llamativo la homogeneidad de los resultados en esa comuna acompañada de la baja proporción de ayuda a adultos mayores. Lo mismo sucede para la comuna de Pedro Aguirre Cerda.

También otros sectores resultan interesantes desde el punto de vista de la heterogeneidad de los resultados dentro de las límites comunales, zonas donde conviven (vecinos en términos de zona censal) adultos mayores que reciben beneficios del estado junto con otros que no.

Por otra parte y considerando la situación actual de pandemia por Covid-19, es importante destacar a las comunas de La Florida, La Pintana y La Granja como aquellas que tienen altos niveles de hacinamiento (crítico) promedio (INE, 2018), donde las últimas dos comunas nombradas además poseen una gran cantidad de adultos mayores socioeconómicamente vulnerables según este trabajo y también otros. (Crespo, 2020a), (Crespo, 2020b), (CChC, 2019),

## **CAPÍTULO 5. CONCLUSIONES**

En general la microsimulación se presenta como una herramienta potente para generar información a partir de datos extraídos de información pública que permite su representación espacial y un análisis territorial sobre la fragmentación de las personas en términos socioeconómicos.

En términos generales, la metodología propuesta muestra buenos resultados, pero existen algunas pequeñas incongruencias que podrían ser explicados por algún error experimental o instrumental (rakeR Package), lo cual debería ser analizado más adelante.

Se recomienda calcular algún indicador o estadístico que permita medir el rendimiento del algoritmo, como el ajuste de bondad, o una tabla de contingencia e índice kappa a partir de la información georreferenciada existente de la encuesta CASEN correspondiente a personas encuestadas.

Se plantea como un desafío la búsqueda de nuevas formas de microsimulación, ya sea por otras librerías disponibles en R o en otro lenguaje de programación, así como también otro algoritmo de machine learning u optimización que mejore el rendimiento de la microsimulación.

## REFERENCIAS BIBLIOGRÁFICAS

- Ballas, D. (2005). Simbritain: A spatial microsimulation approach to population dynamics. Wiley InterScience DOI: 10.1002.
- Borsdorf, A. (2002). Barrios cerrados en santiago de chile, quito y lima: Tendencias de la segregación socio-espacial en capitales andinas. Latin American cities Project.
- Borsdorf, A. (2003). Como modelar el desarrollo y la dinámica de la ciudad latinoamericana. URE Santiago.
- CChC (2019). Presentación icvu 2019. Recuperado el día 03 de Agosto desde <https://www.cchc.cl/>.
- Crespo, R. (2020a). On the spatially explicit gini coefficient the case study of chile a high income developing country. Letters in Spatial and Resource Sciences.
- Crespo, R. (2020b). A spatially explicit analysis of chronic diseases in small areas: a case study of diabetes in santiago, chile. International Journal of Health Geographics.
- Hidalgo, R. (2005). La exclusión residencial y el desarrollo de la ciudad moderna en america latina: De la polarización a la fragmentación. el caso de santiago de chile. Geographicalia journal.
- INE (2018). Sintesis de resultados censo de población y vivienda 2017. Recuperado el día 03 de Agosto desde <https://www.censo2017.cl>.
- Lovelace (2018). Spatial microsimulation with r. Recuperado el día 03 de Agosto desde <https://spatial-microsim-book.robinlovelace.net/>.
- Ruschendorf, L. (1995). Convergence of the iterative proportional fitting procedure. The annals of Statistics.
- Social, M. D. (2018). Sintesis de resultados casen, situación pobreza. Recuperado el día 03 de Agosto desde [http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Resultados\\_pobreza\\_Casen\\_2017.pdf](http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Resultados_pobreza_Casen_2017.pdf).



## Anexo 1: Código fuente de modelación en R.

```
library(RPostgreSQL)
library(dplyr)
library(ggplot2)
library(GGally)
library(plyr)
library(rakeR)
library(data.table)
library(geojsonio)
library(rgdal)
library(seqHMM)
library(sf)
library(sp)
library(writexl)
#library(leaflet)
#library(raster)
library(dplyr)
library(spData)
#-----conectar con base de datos y extraer información necesaria-----#

#-----Extracción de encuesta CASEN-----#

#Conectar con dbpsql
#clave de acceso
pw = {
  "admin"
}
drv = dbDriver("PostgreSQL")

#Crear conexión a base de datos
con = dbConnect(drv, dbname = "postgres",
                host = "localhost", port = 5432,
                user = "postgres", password = "admin")

#Solicitud SQL data "general"
q = "SELECT id_vivienda, region, provincia, comuna, edad, esc, e6a, sexo, oficio4, pco1, y26_1a, y26_1b"
data_casen = dbSendQuery(con, q)
data_casen = dbFetch(data_casen)
casen_gs = data_casen
#> ID=c(1:35357)
#casen_gs$ID=ID
#names(casen_gs)

#-----Extracción de Censo 2017-----#

#Conectar con dbpsql
#clave de acceso
pw = {
  "admin"
}
drv = dbDriver("PostgreSQL")
```

```

#Crear conexión a base de datos
con = dbConnect(drv, dbname = "postgres",
                host = "localhost", port = 5432,
                user = "postgres", password = "admin")

#Solicitud SQL data "general"
q = "SELECT region, provincia, comuna, dc, area, zc_loc, P08, P09, P15 FROM public.censo WHERE provincia"
data_censo = dbSendQuery(con, q)
data_censo = dbFetch(data_censo)
censo_gs = data_censo

#-----Fin de extraccion de datos-----#

#-----Preprocesamiento de datos-----#

#-----Encuesta Casen-----#

#Agregar datos faltantes para esc con regresion lineal x mínimos cuadrados.
index_esc_na = which(is.na(as.numeric(casen_gs$esc)==T))
casen_2 = casen_gs
casen_2$esc = as.numeric(casen_gs$esc)
casen_2$e6a = as.numeric(casen_gs$e6a)
edu_lm = lm(esc~e6a,casen_2[-index_esc_na,])
edu_pred = predict.lm(edu_lm, casen_2[index_esc_na,])
casen_2$esc[index_esc_na]= round(edu_pred)

#Clasificación por rango de edad
casen_2$edad = as.numeric(casen_2$edad)
index_menos_60 = which(casen_2$edad<60)
index_mas_60 = which(casen_2$edad>=60)
casen_2$edad_f = NA
casen_2$edad_f[index_menos_60] = "edad_menor_60"
casen_2$edad_f[index_mas_60] = "edad_mayor_60"

#Clasificación por rango de escolaridad
casen_2$esc = as.numeric(casen_2$esc)
index_0 = which(casen_2$esc==0)
index_1_8 = which(casen_2$esc>0 & casen_2$esc<=8)
index_8_12 = which(casen_2$esc>8 & casen_2$esc<=12)
index_mas_12 = which(casen_2$esc>12)
casen_2$esc_f = NA
casen_2$esc_f[index_0] = "esc_0"
casen_2$esc_f[index_1_8] = "esc_1_8"
casen_2$esc_f[index_8_12] = "esc_8_12"

```

```

casen_2$esc_f[index_mas_12] = "esc_mas_12"

#Clasificacion por sexo
index_mujer = which(casen_2$sexo==2)
casen_2$sex_f = NA
casen_2$sex_f[index_mujer]="sex_f"
casen_2$sex_f[-index_mujer]="sex_m"

#Clasificacion si recibe ayuda de pensión solidaria
index_pension_sol = which(casen_2$y26_1a==1 | casen_2$y26_1b==1)
casen_2$ayuda = NA
casen_2$ayuda[index_pension_sol] = 1
casen_2$ayuda[-index_pension_sol] = 2

casen_f = data.frame("comuna" = casen_2$comuna,
                     "esc_f" = as.character(casen_2$esc_f),
                     "edad_f" = as.character(casen_2$edad_f),
                     "sexo_f" = as.character(casen_2$sex_f),
                     "ayuda" = as.character(casen_2$ayuda))

casen_f_comunas = dplyr(casen_f,.(comuna))

#-----Censo 2017-----#

#for (j in 1:6119984){
#  if (nchar(censo_3$dc[j])==1){
#    censo_3$dc[j] = paste0("0",censo_3$dc[j])
#  }
#}
#
#for (k in 1:6119984){
#  if (nchar(censo_3$zc_loc[k])==1){
#    censo_3$zc_loc[k] = paste0("00",censo_3$zc_loc[k])
#  }else if (nchar(censo_3$zc_loc[k]==2)){
#    censo_3$zc_loc[k] = paste0("0",censo_3$zc_loc[k])
#  }
#}
#}

#Se construye GEOCODIGO
censo_2 = censo_gs
censo_2$dc = as.character(censo_2$dc)
censo_2$dc[censo_2$dc == "1"] = "01"
censo_2$dc[censo_2$dc == "2"] <- "02"
censo_2$dc[censo_2$dc == "3"] <- "03"
censo_2$dc[censo_2$dc == "4"] <- "04"
censo_2$dc[censo_2$dc == "5"] <- "05"
censo_2$dc[censo_2$dc == "6"] <- "06"
censo_2$dc[censo_2$dc == "7"] <- "07"
censo_2$dc[censo_2$dc == "8"] <- "08"
censo_2$dc[censo_2$dc == "9"] <- "09"

censo_2$zc_loc = as.numeric(censo_2$zc_loc)

```

```

index_zl_2 = which((nchar(censo_2$zc_loc))==2)
index_zl_1 = which((nchar(censo_2$zc_loc))==1)
censo_2$zc_loc[index_zl_2] = paste0("0",censo_2$zc_loc[index_zl_2])
censo_2$zc_loc[index_zl_1] = paste0("00",censo_2$zc_loc[index_zl_1])
censo_2$geocode = paste(censo_2$comuna,censo_2$dc,censo_2$area,censo_2$zc_loc)

#Se cambian los nombres de columnas del censo para facilitar lectura
names(censo_2)[names(censo_2) == "p08"] <- "sexo"
names(censo_2)[names(censo_2) == "p09"] <- "edad"
names(censo_2)[names(censo_2) == "p15"] <- "esc"

#se hace el conteo de personas por edad y zona local
censo_2$edad = as.numeric(censo_2$edad)
edad_menos_60 = aggregate (edad ~ geocode + comuna, data = censo_2, FUN = function (a) {length(which(a<60))})
edad_mas_60 = aggregate (edad ~ geocode + comuna, data = censo_2, FUN = function (a) {length(which(a>=60))})
names(edad_menos_60)[names(edad_menos_60) == "edad"] <- "edad_menor_60"
names(edad_mas_60)[names(edad_mas_60) == "edad"] <- "edad_mayor_60"

#se hace el conteo de personas por sexo zona local
sexo_mas = aggregate(sexo ~ geocode + comuna, data = censo_2, FUN = function (a) {length(which(a==1))})
sexo_fem = aggregate(sexo ~ geocode + comuna, data = censo_2, FUN = function (a) {length(which(a==2))})
names(sexo_mas)[names(sexo_mas) == "sexo"] <- "sex_m"
names(sexo_fem)[names(sexo_fem) == "sexo"] <- "sex_f"

#se hace el conteo de personas por años de escolaridad y zona local
censo_2$esc = as.numeric(censo_2$esc)
esc_0 = aggregate (esc ~ geocode + comuna, data = censo_2, FUN = function (a) {length(which(a>=0 & a<=3))})
esc_1_8 = aggregate (esc ~ geocode + comuna, data = censo_2, FUN = function (a) {length(which(a>=4 & a<=8))})
esc_8_12 = aggregate(esc ~ geocode + comuna, data = censo_2, FUN = function (a) {length(which(a>=9 & a<=12))})
esc_mas_12 = aggregate (esc ~ geocode + comuna, data = censo_2, FUN = function (a) {length(which(a>12))})
names(esc_0)[names(esc_0) == "esc"] <- "esc_0"
names(esc_1_8)[names(esc_1_8) == "esc"] <- "esc_1_8"
names(esc_8_12)[names(esc_8_12) == "esc"] <- "esc_8_12"
names(esc_mas_12)[names(esc_mas_12) == "esc"] <- "esc_mas_12"

tabla_consolidada = Reduce(function(x,y) merge(x = x, y = y, by = "geocode"),
list(edad_menos_60, edad_mas_60, sexo_mas, sexo_fem, esc_0, esc_1_8, esc_8_12, esc_mas_12))
tabla_consolidada = tabla_consolidada[,c(1,3,5,7,9,11,13,15,17,16)]
names(tabla_consolidada)[names(tabla_consolidada) == "comuna.y"] <- "comuna"

for (i in 2:9){
  tabla_consolidada[i] = as.numeric(tabla_consolidada[,i])
}

#se consolidan los dataframes (1 dataframe x comuna)
censo_f_comunas = dplyr::group_by(tabla_consolidada,.(comuna))

```

```

#----- Microsimulacion -----#

vars = c("edad_f", "esc_f", "sexo_f")

sim_list = list()
for (i in 1:34){
  casen_f_comunas[[i]]$esc_f = as.character(casen_f_comunas[[i]]$esc_f)
  casen_f_comunas[[i]]$edad_f = as.character(casen_f_comunas[[i]]$edad_f)
  casen_f_comunas[[i]]$sexo_f = as.character(casen_f_comunas[[i]]$sexo_f)
  casen_f_comunas[[i]]$zone = as.character(seq.int(nrow(casen_f_comunas[[i]])))
  casen_f_comunas[[i]] = casen_f_comunas[[i]][,c(1,6,3,2,4,5)]
  censo_f_comunas[[i]] = censo_f_comunas[[i]][,c(1,3,2,6,7,8,9,5,4,10)]
  pesos = weight(cons = censo_f_comunas[[i]][,-10], inds = casen_f_comunas[[i]][,-1], vars=vars)
  pesos_int = integerise(pesos, inds = casen_f_comunas[[i]][,-1])
  names(pesos_int)[1]="geocode"
  sim_list[[i]]=pesos_int
}

sim_df = rbindlist(sim_list)
names(sim_df)[1]="GEOCODIGO"
#-----Representación-----#

#Se genera la columna de adultos mayores y se hace join al spatialdatagrame
gs_layer <- rgdal::readOGR("C:/Users/ppsa_/Documents/Drope/u/Nivel 10/Geomarketing/T2/zonas_cens_gs/zona
poblacion_mayor = edad_mas_60[,c(1,3)]
poblacion_mayor$geocode <- gsub(' ', '', poblacion_mayor$geocode)
names(poblacion_mayor)[1]="GEOCODIGO"
gs_layer@data = join(gs_layer@data,poblacion_mayor, by="GEOCODIGO")

#Se genera la columna ayuda y se hace join al spatialdatagrame
ayuda_count = aggregate(ayuda ~ GEOCODIGO, data = sim_df, FUN = function(a) {length(which(a==1))})
ayuda_count$GEOCODIGO <- gsub(' ', '', ayuda_count$GEOCODIGO)
gs_layer@data = join(gs_layer@data,ayuda_count, by="GEOCODIGO")

gs_layer@data$indicador = gs_layer@data$ayuda / gs_layer@data$edad_mayor_60

library(cartography)

#Se generan los quiebres quantiles, se plotea y agregan preferencias del layout
bks1 <- getBreaks(v = gs_layer$indicador, nclass = 6, method = "quantile")
choroLayer(spdf = gs_layer, df = gs_layer@data, var = "indicador", breaks = bks1, legend.pos = n)
layoutLayer(title="Adultos mayores que reciben pensión solidaria en el Gran Santiago", author = "Pedro P
north =TRUE, col = "#3F0C7B")

```

```

legendChoro(pos = "topleft",
            title.txt = "Adultos mayores con pensión solidaria / \n  mayores de 60 años",
            breaks = bks1,
            col = carto.pal("blue.pal", 6),
            nodata = TRUE, nodata.txt = "No Data")

#se exporta dataframe de capa gs_layer
writeOGR(obj = gs_layer, dsn = "gs_layer.geojson", layer="gs_layer", driver="GeoJSON")
write_xlsx(ayuda_count, "tabla_microsim.xlsx")

```