

Artículo de investigación

Las viviendas en el Aglomerado del Gran San Juan según ingresos de los hogares. Aplicaciones desde el análisis simbólico de datos

José Ernesto Torres^{1*} y Jorgelina Carrizo^{2**}

¹ Universidad Nacional de San Juan. Departamento de Geografía. Instituto de Geografía Aplicada

² Universidad Nacional de San Juan. Departamento de Matemática

* E-mail: jtorres@ffha.unsj.edu.ar

** E-mail: jorgelinav_carrizo@yahoo.com.ar

Recibido: 13/07/2020; Aceptado: 15/08/2020; Publicado: 15/12/2020

Resumen

El presente trabajo tiene como objetivo plantear un panorama de las viviendas en el Aglomerado del Gran San Juan, en función de los niveles de ingresos de los hogares. La fuente de datos es la Encuesta Permanente de Hogares (EPH) correspondiente al primer trimestre del año 2019. Sobre la base hogares se seleccionan variables que permitan describir las viviendas. La metodología utilizada se basa en el Análisis Simbólico de Datos (SDA), herramienta de gran utilidad frente a la complejidad de la EPH, en cuanto a su elevado número de variables y de casos. El resumen de datos se realiza en términos de objetos simbólicos (OS), representados por los deciles para la variable Ingreso Total Familiar. Los softwares utilizados son SODAS y R. Se realiza una descripción de los OS construidos y se aplican métodos de clustering para agruparlos. Finalmente, se realiza una Regresión Simbólica de Datos para variables de intervalo. Los principales resultados obtenidos muestran que los OS presentan ciertas características similares, a pesar de las diferencias en los ingresos. Los métodos de agrupamiento muestran claramente las diferencias entre los OS. La regresión simbólica indica una asociación entre el ingreso total familiar y las deficiencias de las viviendas.

Palabras clave: viviendas; hogares; ingresos; análisis simbólico de datos

Dwellings in Great San Juan according to household income. Applications from symbolic data analysis

Abstract

The aim of this paper is to present a panorama of the dwellings in Great San Juan, based on the levels of household income. The data source is the Permanent Household Survey (EPH) corresponding to the first quarter of the year 2019. Variables that allow the description of the dwellings are selected, based on household. The methodology used is based on the Symbolic Data Analysis (SDA), the very useful tool for dealing with the complexity of the EPH, in terms of its high number of variables and cases. The summary of data is done in terms of symbolic objects (OS), represented by deciles for the Total Family Income variable. The programs used are SODAS and R. A description of the OS is made and clustering methods are used to group them. Finally, a Symbolic Data Regression is performed for interval variables. The main results obtained show that the OS share certain similar characteristics, despite differences in revenue. Grouping methods show the differences between the OS clearly. The symbolic regression indicates an association between total family income and housing deficiencies.

Keywords: dwellings; household; income; symbolic data analysis

1. Introducción

Según los datos de la Encuesta Permanente de Hogares (EPH), para el primer trimestre de 2019, el 99,7 % de los hogares del Aglomerado del Gran San Juan (AGSJ) habitaba en casas o departamentos. Este dato, tomado de manera aislada, puede conducir a pensar en que los habitantes de dicho aglomerado urbano presentan una muy buena calidad de vida. Sin embargo, la observación cotidiana, demuestra que la realidad está muy distante de esa presunción.

Las características físicas y las condiciones sanitarias de las viviendas, entre otras dimensiones, aparecen despegadas o separadas del tipo de vivienda. Se ha optado por la generalización de la palabra “casa” sin contemplar las particularidades de cada una de ellas. Al respecto, en las últimas décadas en la provincia de San Juan se han desarrollado diversos planes de vivienda, puestos en marcha por diferentes gobiernos. A ello se suma el Plan de Erradicación de Villas de Emergencia aplicado en la década del 2000, que cumplió con el objetivo de erradicar este tipo de viviendas del departamento Capital. Según este plan las familias erradicadas recibieron nuevas viviendas en departamentos aledaños a Capital y pertenecientes al Gran San Juan.

En virtud de ello se plantea que el número de viviendas dignas ha aumentado, disminuyendo la cantidad de hogares pobres, por lo menos en esta dimensión. Ahora bien ¿el hecho de que la mayor parte de los hogares habiten en casa o departamento es correlativo con la disminución de la pobreza? En este punto se hace necesario considerar otras variables que permitan una aproximación más cercana a la realidad. En función de ello se plantea como objetivo general conocer la situación de los hogares del AGSJ, en relación a las características de sus viviendas, tomando como variable de control su nivel de ingresos. Como objetivos específicos se propone analizar los objetos simbólicos contruidos a partir de los deciles según el ingreso total familiar (ITF); distinguir grupos de objetos simbólicos mutuamente excluyentes y establecer asociaciones entre variables aplicando técnicas de regresión simbólica.

El estudio se hace siguiendo la metodología del Análisis Simbólico de Datos (SDA), introducido por Diday (1987, 1988, 1991). Se trata de una herramienta estadística fundamental para Data Mining. Frente a las características de la EPH, que contiene un gran número de variables con un número mayor aún de unidades de análisis, por ende un gran cúmulo de información, se impone un proceso de extracción de conocimiento que permita encontrar la información más importante o más significativa para la toma de decisiones. Al respecto Ruiz Shulcloper et al. (1997) plantea que:

En la medida en que los datos se hacen más complejos, por ejemplo dejan de ser exclusivamente numéricos para presentarse mezclados con datos de naturaleza cualitativa, con subjetividad, imprecisión y otros elementos de esta índole, mayor es la dificultad de extraer información útil de los mismos” (p. 13)

De acuerdo a Mallea et al. (2018) “El análisis de datos simbólicos proporciona un marco que permite representar datos con variabilidad, usando nuevos tipos de variables” (p. 81). Los datos simbólicos son “... conjuntos finitos de valores, intervalos, o de forma más general distribuciones.” (Mallea et al., 2018, p. 81) que pueden ser ordenados en forma matricial. La misma autora define a los objetos simbólicos (OS) o conceptos de la siguiente manera:

Los objetos simbólicos representan conceptos, entendidos como la intención y extensión del mismo. La intención de un concepto representa las propiedades que lo definen y que lo hacen distinto de los demás conceptos. La extensión de un concepto se compone de los individuos que se definen por el concepto o que cumplen las propiedades que definen el concepto (p. 82)

En función de lo expuesto, es importante considerar, que el presente trabajo se enmarca en dos proyectos que se desarrollan en la Facultad de Filosofía, Humanidades y Artes de la Universidad Nacional de San Juan. Ellos son Observatorio de la población de San Juan: su estado en vísperas del Censo Nacional de Población, Hogares y Viviendas 2020 y Herramientas estadísticas para Data Mining y modelos dinámicos.

2. Materiales y métodos

La metodología sigue los lineamientos del Análisis Simbólico de Datos. La fuente de información es la EPH correspondiente al primer trimestre del año 2019. El área de estudio es el AGSJ. En primer lugar se seleccionan las variables que describan el estado de las viviendas y el número de miembros del hogar, según sean menores de 10 años y mayores de 10 años. La tabla 1 y la tabla 2 muestran las variables de intervalo y las variables modales respectivamente.

Los OS considerados están representados por la variable modal “Número de decil del ingreso total del hogar del total EPH”. Se trata de 12 objetos simbólicos, ya que se agregan aquellos casos en que los miembros del hogar no perciben ingresos y aquellos casos de no respuesta. La obtención de los objetos simbólicos (OS) se realiza mediante el software SODAS (Symbolic Official Data Analysis System).

Tabla 1. Variables de intervalo seleccionadas

Código	Variable
II1	Cuántos ambientes/ habitaciones tiene este hogar para uso exclusivo
II2	De esos, cuántos usan habitualmente para dormir?
IX_TOT	Cantidad de miembros del Hogar
IX_MEN10	Cantidad de miembros del Hogar menores de 10 años
IX_MAYEQ10	Cantidad de miembros del Hogar de 10 y más años
ITF	Monto de Ingreso Total Familiar

Fuente: elaboración propia sobre la base de la EPH (INDEC)

La tabla simbólica se construye a través de una query SQL de Tipo 1, trabajando con el módulo DB2SO del paquete SODAS. De la base original de la EPH, se filtran los datos correspondientes al AGSJ. Sobre esa base se seleccionan las variables, quedando una matriz de 597 filas (casos) por 20 columnas (variables y concepto).

Habiendo creado la tabla simbólica, se realiza una cadena de métodos, integrada por DSTAT con el fin de realizar la estadística descriptiva de las variables y VIEW con el fin de visualizar los OS que aparecen en la tabla simbólica, mediante gráficos Zoom Star. Se propone, además, la clasificación de los OS, a partir de la aplicación de los métodos DIV y SCLUST. El primero es un método de agrupamiento divisivo y jerárquico, mientras que el segundo se puede usar para dividir un conjunto de n datos simbólicos p -dimensionales en un número m de grupos homogéneos, de modo que el método de agrupamiento resultante, queda a criterio del investigador, quien puede ajustar una serie de parámetros con el fin de lograr una clasificación óptima.

Tabla 2. Variables modales seleccionadas

CÓDIGO	Variable	Categorías
IV1	Tipo de vivienda	1. Casa; 2. Departamento; 3. Pieza de inquilinato; 4. Pieza en hotel/pensión; 5. Local no construido para habitación
IV3	Los pisos interiores son principalmente de...	1. Mosaico/baldosa/madera/cerámica/alfombra; 2. Cemento/ladrillo fijo; 3. Ladrillo suelto/tierra
IV4	La cubierta exterior del techo es de....	1. Membrana/cubierta asfáltica; 2. Baldosa/losa sin cubierta; 3. Pizarra/teja; 4. Chapa de metal sin cubierta; 5. Chapa de fibrocemento/plástico; 6. Chapa de cartón; 7. Caña/tabla/paja con barro/paja sola; 9. N/S. Depto en propiedad horizontal
IV6	Tiene agua.....	1. Por cañería dentro de la vivienda; 2. Fuera de la vivienda pero dentro del terreno; 3. Fuera del terreno
IV7	El agua es de.....	1. Red pública (agua corriente); 2. Perforación con bomba a motor; 3. Perforación con bomba manual
IV8	¿Tiene baño/letrina?	1. Si; 2. No
IV9	El baño o letrina está....	1. Dentro de la vivienda; 2. Fuera de la vivienda pero dentro del terreno; 3. Fuera del terreno
IV10	El baño tiene....	1. Inodoro con botón/ mochila/cadena y arrastre de agua; 2. Inodoro sin botón/cadena y con arrastre de agua (a balde); 3. Letrina (sin arrastre de agua)
IV11	El desagüe del baño es....	1. A red pública (cloaca); 2. A cámara séptica y pozo ciego; 3. Sólo a pozo ciego; 4. A hoyo/excavación en la tierra
II4_1	¿Tiene cuarto de cocina?	1. Si; 2. No
II4_2	¿Tiene lavadero?	1. Si; 2. No
II4_3	¿Tiene garage?	1. Si; 2. No
II7	Régimen de tenencia	1. Propietario de la vivienda y el terreno; 2. Propietario de la vivienda solamente; 3. Inquilino/arrendatario de la vivienda; 4. Ocupante por pago de impuestos/expensas; 5. Ocupante en relación de dependencia; 6. Ocupante gratuito (con permiso); 7. Ocupante de hecho (sin permiso); 8. ¿Está en sucesión?
II8	Combustible utilizado para cocinar	1. Gas de red; 2. Gas de tubo/garrafa; 3. Kerosene/leña/carbón
DECIFR	Nº de decil del ingreso total del hogar del total EPH	0 = Sin ingresos; 1 = decil 1; 2 = decil 2; 3 = decil 3; 4 = decil 4; 5 = decil 5; 6 = decil 6; 7 = decil 7; 8 = decil 8; 9 = decil 9; 10 = decil 10; 12 = no respuesta de ingresos

Fuente: elaboración propia sobre la base de la EPH (INDEC)

Seguidamente, a través de software R se realiza una regresión lineal simbólica simple, considerando la tabla simbólica de datos, que tiene como conceptos a los deciles según el nivel de ingreso y como variables de intervalo, el monto de ingreso total familiar (ITF) y el número de casas con cubierta exterior del techo deficiente (CTD). Se toma entonces, para la aplicación de los métodos de regresión simbólica, al CTD como variable respuesta y al ITF como variable explicativa.

3. El área de estudio

En relación al área de estudio, el AGSJ, cabe destacar que se trata del centro que presenta la mayor concentración poblacional de la provincia de San Juan con un Índice de Concentración Urbana de 69,9 %, según los datos aportados por el Censo Nacional de Población, Hogares y Viviendas 2010. De acuerdo a las estimaciones realizadas por el Instituto Nacional de Estadísticas y Censos, se ha

convertido en un aglomerado de más de 500000 habitantes, alcanzando en abril de 2020 los 535339 habitantes. Según puede apreciarse en la figura 1 comprende los departamentos centrales de la provincia de San Juan, extendiéndose por Capital, Chimbab y Santa Lucía y parte de Rivadavia, Rawson y Pocito.

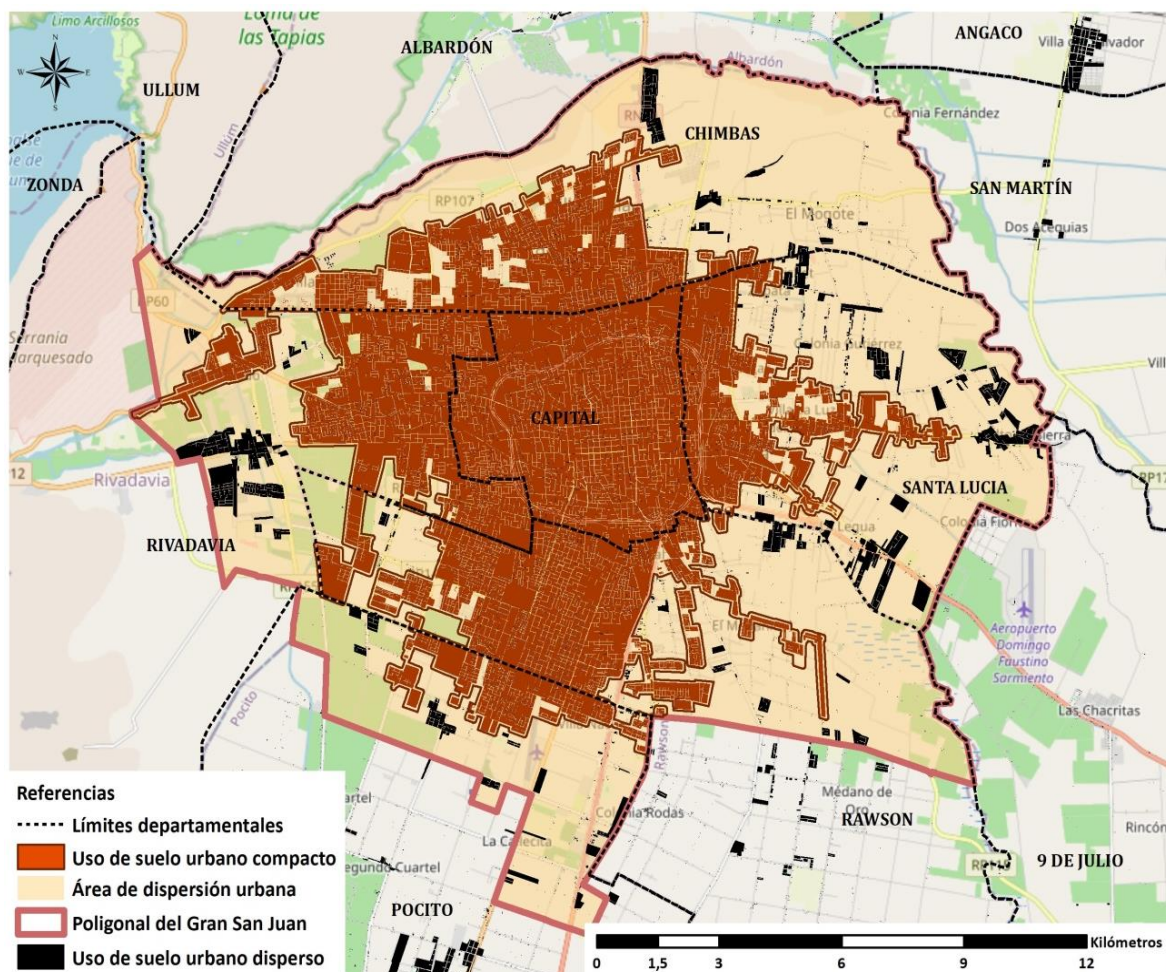


Figura 1. El área del Gran San Juan.

Fuente. Programa “Geografía de la Provincia de San Juan a través de Estudios Departamentales. Propuestas de Ordenamiento Territorial” Instituto de Geografía Aplicada. FFHA, UNSJ. Elaboración propia en base a relevamientos de campo e imágenes satelitales. Límite poligonal del AGSJ INDEC (2010). Límites departamentales CEFOCA (2016).

4. Los objetos simbólicos

4.1. Variable utilizada para la construcción de los objetos simbólicos

En primer lugar se realiza una presentación de la variable utilizada para la construcción de los objetos simbólicos. Se trata de “Número de decil del ingreso total del hogar del Total EPH”. En el Cuadro 3 se muestra la frecuencia para cada caso y los valores mínimo y máximo para cada decil. Se observa que aquellos hogares sin ingresos representan un porcentaje de 0,34 %, equivalente sólo a 2 casos. Por su parte la no respuesta ante la pregunta acerca de los ingresos sólo está representada por un caso.

Tabla 3. Deciles del ingreso total del hogar

Deciles	Frecuencia	Porcentaje	Valor Mínimo	Valor Máximo
0	2	0,34		
1	56	9,38	1830	11100
2	69	11,56	11600	16300
3	67	11,22	16400	20900
4	46	7,71	21000	25600
5	59	9,88	25630	29990
6	58	9,72	31000	38000
7	61	10,22	38400	46300
8	60	10,05	46700	58000
9	59	9,88	58315	77130
10	59	9,88	79200	290000
12	1	0,17		
Total	597	100		

Fuente: elaboración propia sobre la base de la EPH (INDEC)

4.2. Análisis de los objetos simbólicos

El análisis de los objetos simbólicos se realiza solamente para los deciles 1, 2 y 3, que representan las situaciones más desfavorables y los deciles 8, 9 y 10 que representan las más favorables.

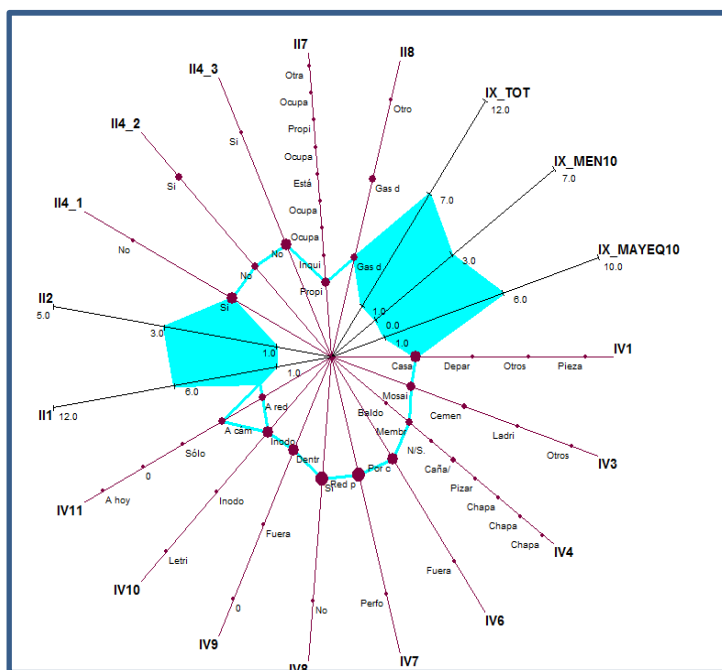


Figura 2. Gráfico Zoom Star que representa el Objeto Simbólico OS = 1 (Decil 1)

Fuente: elaboración propia sobre la base de la EPH (INDEC)

Los hogares agrupados en el OS = 1 (figura 2) habitan, en su mayoría, en casas con pisos de mosaico/baldosa, madera, cerámica/alfombra y con cubierta exterior del techo elaborado con membrana/cubierta asfáltica. Reciben agua por cañería dentro de la vivienda, procedente de red pública. Tienen baño dentro de la vivienda con inodoro con botón, mochila/cadena. El desagüe del baño es a red pública (cloaca) y a cámara séptica y pozo ciego. Disponen de entre uno y ocho ambientes, de los cuales, entre uno y dos son para dormir. Tienen cuarto de cocina y no tienen

lavadero ni garaje. En su mayoría son propietarios de la vivienda y el terreno y utilizan como combustible para cocinar el gas natural. Los hogares están conformados por entre uno y siete miembros, de los cuales menos de tres tienen diez años y de uno a seis tienen más de diez años. Los datos de la tabla 4 permiten apreciar la participación de cada categoría, en porcentaje.

Tabla 4. Participación de las diferentes categorías de las variables en el OS =1

IV1 = Casa (0.87), Depar (0.10), Otros (0.02), Pieza (0.02)
And IV3 = Mosai (0.68), Cemen (0.30), Ladri (0.02)
And IV4 = Baldo (0.17), Membr (0.50), N/S. (0.02), Caña/ (0.25), Pizar (0.02), Chapa (0.02), Chapa (0.03)
And IV6 = Por c (0.98), Fuera (0.02)
And IV7 = Red p (1.00)
And IV8 = Si (1.00)
And IV9 = Dentr (0.82), Fuera (0.18)
And IV10 = Inodo (0.93), Inodo (0.07)
And IV11 = A red (0.43), A cám (0.43), Sólo (0.13)
And I11 = [1.00 : 6.00]
And I12 = [1.00 : 3.00]
And I14_1 = Si (0.95), No (0.05)
And I14_2 = No (0.53), Si (0.47)
And I14_3 = No (0.82), Si (0.18)
And I17 = Propi (0.62), Inqui (0.10), Ocupa (0.07), Ocupa (0.07), Está (0.10), Ocupa (0.02), Propi (0.02), Otra (0.02)
And I18 = Gas d (0.57), Gas d (0.43)
And IX_TOT = [1.00 : 7.00]
And IX_MEN10 = [0.00 : 3.00]
And IX_MAYEQ10 = [1.00 : 6.00]

Fuente: elaboración propia sobre la base de la EPH (INDEC)

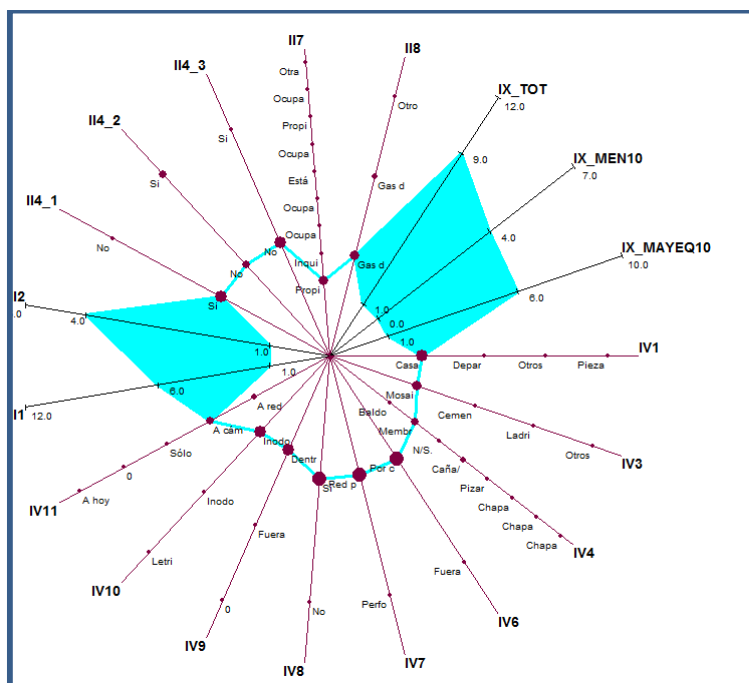


Figura 3. Gráfico Zoom Star que representa el Objeto Simbólico OS = 2 (Decil 2)

Fuente: elaboración propia sobre la base de la EPH (INDEC)

Los hogares agrupados en el OS = 2 (figura 3) habitan, en su mayoría, en casas con pisos de mosaico/baldosa, madera, cerámica/alfombra y con cubierta exterior del techo elaborado con membrana/cubierta asfáltica. Reciben agua por cañería dentro de la vivienda, procedente de red pública. Tienen baño dentro de la vivienda con inodoro con botón, mochila/cadena. El desagüe del baño es a cámara séptica y pozo ciego. Disponen de entre uno y ocho ambientes, de los cuales, entre uno y tres son para dormir. Tienen cuarto de cocina y no tienen lavadero ni garaje. En su mayoría son propietarios de la vivienda y el terreno y utilizan como combustible para cocinar el gas natural. Los hogares están conformados por entre uno y nueve miembros, de los cuales menos de cuatro

tienen diez años y de uno a seis tienen más de diez años. La tabla 5 muestra la participación de cada categoría, en porcentaje.

Tabla 5. Participación de las diferentes categorías de las variables en el OS =2

IV1 = Casa (0.87), Depar (0.13)
And IV3 = Mosai (0.78), Cemen (0.16), Ladri (0.03), Otros (0.03)
And IV4 = Baldo (0.15), Membr (0.49), N/S. (0.04), Caña/ (0.26), Pizar (0.03), Chapa (0.03)
And IV6 = Por c (1.00)
And IV7 = Red p (1.00)
And IV8 = Si (1.00)
And IV9 = Dentr (0.90), Fuera (0.10)
And IV10 = Inodo (0.94), Inodo (0.06)
And IV11 = A red (0.38), A cám (0.49), Sólo (0.13)
And II1 = [1.00 : 6.00]
And II2 = [1.00 : 4.00]
And II4_1 = Si (0.87), No (0.13)
And II4_2 = No (0.59), Si (0.41)
And II4_3 = No (0.96), Si (0.04)
And II7 = Propi (0.65), Inqui (0.12), Ocupa (0.06), Ocupa (0.07), Está (0.06), Ocupa (0.01), Propi (0.03)
And II8 = Gas d (0.62), Gas d (0.38)
And IX_TOT = [1.00 : 9.00]
And IX_MEN10 = [0.00 : 4.00]
And IX_MAYEQ10 = [1.00 : 6.00]

Fuente: elaboración propia sobre la base de la EPH (INDEC)

Los hogares agrupados en el OS = 3 (figura 4) habitan, en su mayoría, en casas con pisos de mosaico/baldosa, madera, cerámica/alfombra y con cubierta exterior del techo elaborado con membrana/cubierta asfáltica. Reciben agua por cañería dentro de la vivienda, procedente de red pública. Tienen baño dentro de la vivienda con inodoro con botón, mochila/cadena. El desagüe del baño es a red pública. Disponen de entre uno y nueve ambientes, de los cuales, entre uno y tres son para dormir. Tienen cuarto de cocina y no tienen lavadero ni garaje. En su mayoría son propietarios de la vivienda y el terreno y utilizan como combustible para cocinar el gas natural. Los hogares están conformados por entre uno y nueve miembros, de los cuales menos de cinco tienen diez años y de uno a seis tienen más de diez años. Los datos de la tabla 6 indican la participación de cada categoría, en porcentaje.

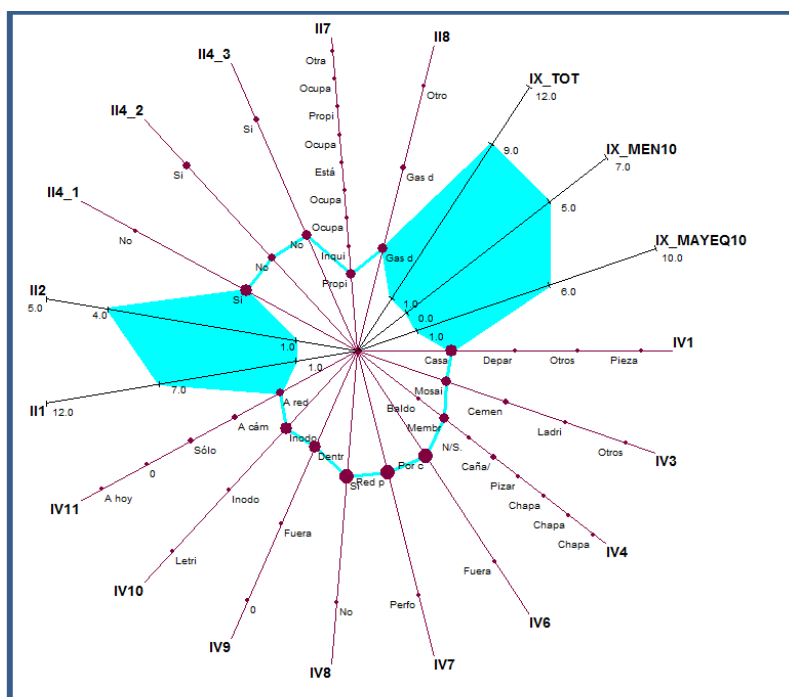


Figura 4. Gráfico Zoom Star que representa el Objeto Simbólico OS = 3 (Decil 3)
Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

Tabla 6. Participación de las diferentes categorías de las variables en el OS =3

IV1 = Casa (0.88), Depar (0.12)
And IV3 = Mosai (0.68), Cemen (0.25), Ladri (0.04), Otros (0.03)
And IV4 = Baldo (0.14), Membr (0.61), N/S. (0.01), Caña/ (0.23)
And IV6 = Por c (1.00)
And IV7 = Red p (1.00)
And IV8 = Si (1.00)
And IV9 = Dentr (0.96), Fuera (0.04)
And IV10 = Inodo (0.96), Inodo (0.04)
And IV11 = A red (0.46), A cám (0.29), Sólo (0.25)
And II1 = [1.00 : 7.00]
And II2 = [1.00 : 4.00]
And II4_1 = Si (0.91), No (0.09)
And II4_2 = No (0.54), Si (0.46)
And II4_3 = No (0.75), Si (0.25)
And II7 = Propi (0.64), Inqui (0.17), Ocupa (0.01), Ocupa (0.07), Está (0.06), Propi (0.01), Ocupa (0.03)
And II8 = Gas d (0.67), Gas d (0.33)
And IX_TOT = [1.00 : 9.00]
And IX_MEN10 = [0.00 : 5.00]
And IX_MAYEQ10 = [1.00 : 6.00]

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

Los hogares agrupados en el OS = 8 (figura 5) habitan, en su mayoría, en casas con pisos de mosaico/baldosa, madera, cerámica/alfombra y con cubierta exterior del techo elaborado con membrana/cubierta asfáltica. Reciben agua por cañería dentro de la vivienda, proveniente de red pública. Tienen baño dentro de la vivienda con inodoro con botón, mochila/cadena. El desagüe del baño es a cámara séptica con pozo ciego. Disponen de entre uno y siete ambientes, de los cuales, entre uno y cuatro son para dormir. Tienen cuarto de cocina y tienen lavadero, pero no tienen garaje. En su mayoría son propietarios de la vivienda y el terreno y utilizan como combustible para cocinar el gas natural. Los hogares están conformados por entre uno y nueve miembros, de los cuales entre cero y cuatro tienen menos de diez años y de uno a seis tienen más de diez años. Los datos de la tabla 7 permiten apreciar la participación de cada categoría en porcentaje.

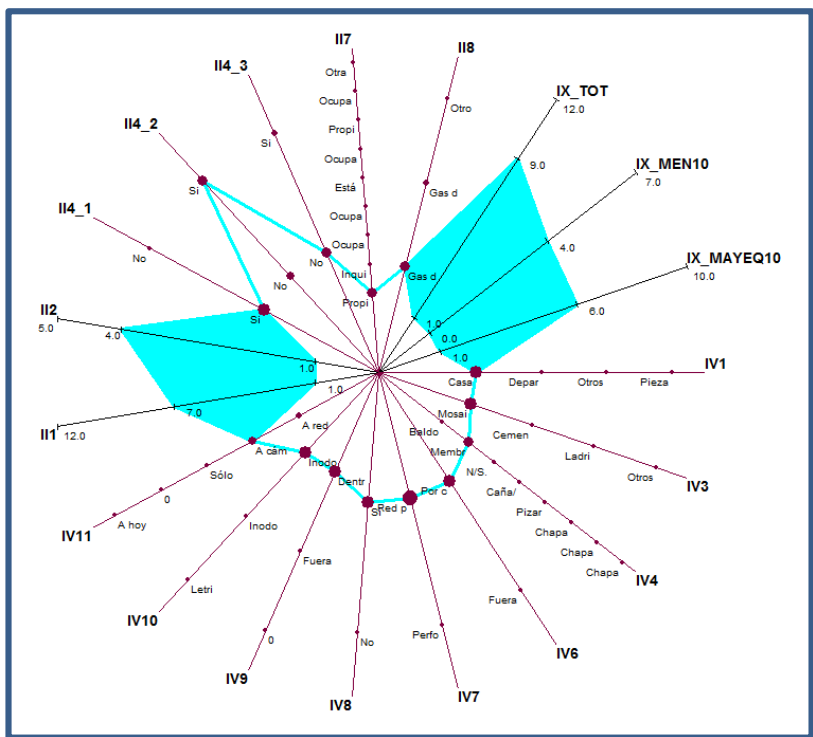


Figura 4. Gráfico Zoom Star que representa el Objeto Simbólico OS = 8 (Decil 8)

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

Tabla 7. Participación de las diferentes categorías de las variables en el OS =8

IV1 = Casa (0.87), Depar (0.13)
And IV3 = Mosai (0.89), Cemen (0.09), Ladri (0.02)
And IV4 = Baldo (0.18), Membra (0.67), Caña/ (0.07), Pizar (0.07), Chapa (0.02)
And IV6 = Por c (0.98), Fuera (0.02)
And IV7 = Red p (1.00)
And IV8 = Si (0.98), No (0.02)
And IV9 = Dentr (0.96), Fuera (0.04)
And IV10 = Inodo (0.93), Inodo (0.04), Letri (0.02)
And IV11 = A red (0.36), A cám (0.53), Sólo (0.09), A hoy (0.02)
And I11 = [1.00 : 7.00]
And I12 = [1.00 : 4.00]
And I14_1 = Si (0.93), No (0.07)
And I14_2 = No (0.40), Si (0.60)
And I14_3 = No (0.76), Si (0.24)
And I17 = Propi (0.64), Inqui (0.18), Ocupa (0.04), Ocupa (0.09), Está (0.04)
And I18 = Gas d (0.76), Gas d (0.22), Otro (0.02)
And IX_TOT = [1.00 : 9.00]
And IX_MEN10 = [0.00 : 4.00]
And IX_MAYEQ10 = [1.00 : 6.00]

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

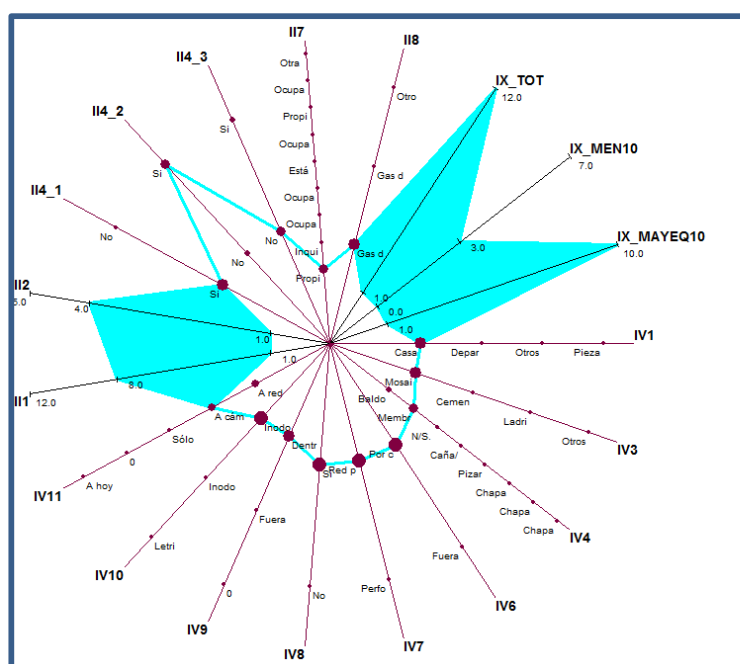


Figura 6. Gráfico Zoom Star que representa el Objeto Simbólico OS = 9 (Decil 9)

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

Los hogares agrupados en el OS = 9 (Figura 6) habitan, en su mayoría, en casas con pisos de mosaico/baldosa, madera, cerámica/alfombra y con cubierta exterior del techo elaborado con membrana/cubierta asfáltica. Reciben agua por cañería dentro de la vivienda, procedente de red pública. Tienen baño dentro de la vivienda con inodoro con botón, mochila/cadena. El desagüe del baño es a cámara séptica con pozo ciego. Disponen de entre uno y diez ambientes, de los cuales, entre uno y tres son para dormir. Tienen cuarto de cocina, lavadero y no tienen garaje. En su mayoría son propietarios de la vivienda y el terreno y utilizan como combustible para cocinar el gas natural. Los hogares están conformados por entre uno y doce miembros, de los cuales menos de tres tienen diez años y de uno a diez tienen más de diez años. Los datos de la tabla 8 muestran la participación de cada categoría en porcentaje.

Los hogares agrupados en el OS = 10 (Figura 7) habitan, en su mayoría, en casas con pisos de mosaico/baldosa, madera, cerámica/alfombra y con cubierta exterior del techo elaborado con

membrana/cubierta asfáltica. Reciben agua por cañería dentro de la vivienda, procedente de red pública. Tienen baño dentro de la vivienda con inodoro con botón, mochila/cadena. El desagüe del baño es a red pública y a cámara séptica con pozo ciego. Disponen de entre dos y nueve ambientes, de los cuales, entre uno y cuatro son para dormir. Tienen cuarto de cocina, lavadero y garaje. En su mayoría son propietarios de la vivienda y el terreno y utilizan como combustible para cocinar el gas natural. Los hogares están conformados por entre uno y nueve miembros, de los cuales menos de tres tienen diez años y de uno a nueve tienen más de diez años. Los datos de la tabla 9 permiten apreciar la participación de cada categoría en porcentaje.

Tabla 8. Participación de las diferentes categorías de las variables en el OS =9

IV1 = Casa (0.90), Depar (0.10)
And IV3 = Mosai (0.88), Cemen (0.10), Otros (0.02)
And IV4 = Baldo (0.20), Membr (0.63), N/S. (0.04), Caña/ (0.08), Pizar (0.04)
And IV6 = Por c (1.00)
And IV7 = Red p (1.00)
And IV8 = Si (1.00)
And IV9 = Dentr (0.94), Fuera (0.06)
And IV10 = Inodo (1.00)
And IV11 = A red (0.45), A cám (0.51), Sólo (0.04)
And II1 = [1.00 : 8.00]
And II2 = [1.00 : 4.00]
And II4_1 = Si (0.96), No (0.04)
And II4_2 = No (0.31), Si (0.69)
And II4_3 = No (0.63), Si (0.37)
And II7 = Propi (0.76), Inqui (0.08), Ocupa (0.04), Ocupa (0.08), Está (0.02), Propi (0.02)
And II8 = Gas d (0.88), Gas d (0.12)
And IX_TOT = [1.00 : 12.00]
And IX_MEN10 = [0.00 : 3.00]
And IX_MAYEQ10 = [1.00 : 10.00]

Tabla 9. Participación de las diferentes categorías de las variables en el OS = 10

IV1 = Casa (0.86), Depar (0.14)
And IV3 = Mosai (0.93), Cemen (0.07)
And IV4 = Baldo (0.21), Membr (0.61), N/S. (0.11), Caña/ (0.04), Pizar (0.04)
And IV6 = Por c (1.00)
And IV7 = Red p (1.00)
And IV8 = Si (1.00)
And IV9 = Dentr (1.00)
And IV10 = Inodo (1.00)
And IV11 = A red (0.50), A cám (0.50)
And II1 = [2.00 : 9.00]
And II2 = [1.00 : 4.00]
And II4_1 = Si (0.96), No (0.04)
And II4_2 = No (0.25), Si (0.75)
And II4_3 = No (0.46), Si (0.54)
And II7 = Propi (0.75), Inqui (0.07), Ocupa (0.07), Ocupa (0.07), Propi (0.04)
And II8 = Gas d (0.89), Gas d (0.07), Otro (0.04)
And IX_TOT = [1.00 : 9.00]
And IX_MEN10 = [0.00 : 3.00]
And IX_MAYEQ10 = [1.00 : 9.00]

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

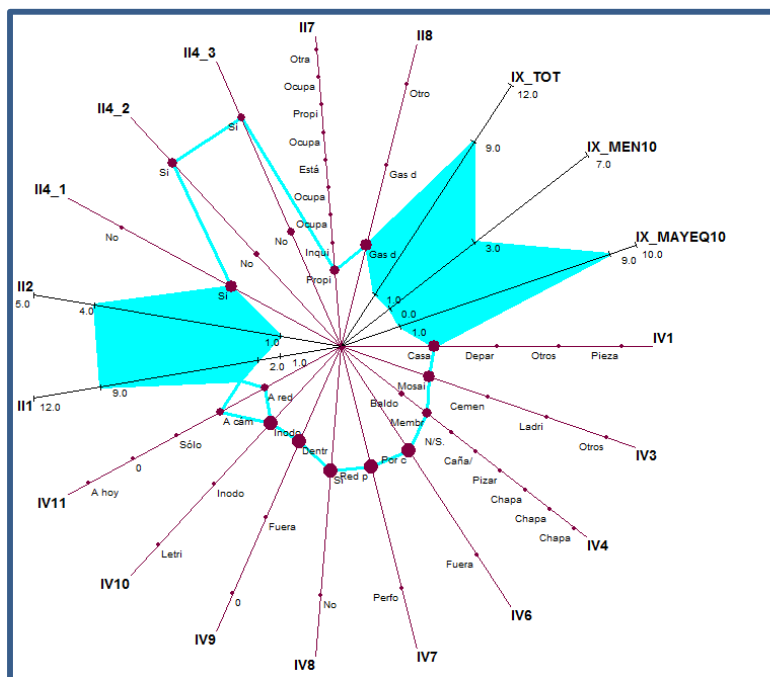


Figura 7. Gráfico Zoom Star que representa el Objeto Simbólico OS = 10 (Decil 10)
Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

5. Clasificación de los objetos simbólicos

5.1. Método DIV

Con el propósito de clasificar los objetos simbólicos se utiliza el método DIV, perteneciente al módulo Clustering de SODAS. Se trata de un método de agrupamiento jerárquico que comienza con todos los objetos en un grupo y continúa por divisiones sucesivas de cada uno. En cada paso, un grupo se divide en dos de acuerdo con una pregunta binaria. Dicha pregunta induce la mejor partición en dos grupos de acuerdo con una extensión del criterio de inercia. En este caso la matriz de entrada es la matriz de datos simbólicos con las variables modales seleccionadas. Se trabaja con cinco clusters y la disimilitud entre los objetos es no normalizada.

La tabla 10 muestra los objetos simbólicos por cluster. Se observa que el cluster 2 contiene el grupo de hogares sin ingresos, el cluster 3 aquellos hogares que no respondieron la pregunta y el cluster 4 los que poseen mayor nivel de ingresos y se encuentran en el decil 10. El cluster 1 está integrado por los deciles de menores ingresos con excepción del decil 4. El cluster 5 contiene a este último decil y los mayores a 6.

Tabla 10. Partición en cinco cluster

Cluster 1 (n=4) : 05 02 03 01
Cluster 2 (n=1) : 00
Cluster 3 (n=1) : 12
Cluster 4 (n=1) : 10
Cluster 5 (n=5) : 06 09 04 07 08
Explicated inercia : 82.296907

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

La tabla 11 y la figura 8 muestran la conformación de los clusters. Se observa que la variable material de los pisos interiores (IV3) es la primera que divide el árbol. Los clusters 1, 3, 4 y 5 son los que contienen viviendas con piso de mosaico/baldosa/madera/cerámica/alfombra, mientras que el cluster 2 tiene piso de cemento/ladrillo fijo o ladrillo suelto/tierra. La segunda división la realiza la variable tenencia de garaje (II4_3). Los clusters 1 y 5 contienen viviendas sin garaje, mientras que los cluster 3 y 4 si poseen. La última división la produce la variable tenencia de lavadero (II4_2). Los hogares que pertenecen al cluster 3 no tienen lavadero, mientras que los del cluster 4 si lo tienen.

Tabla 11. Descripción de los cluster

Cluster 1 : IF 4- [II4_2 = No] AND 2- [II4_3 = No] AND 1- [IV3 = Mosaico/baldosa/madera/cerámica/alfombra]
Cluster 2 : IF 1- [IV3 = Otros OR Ladrillo suelto/tierra OR Cemento/ladrillo fijo]
Cluster 3 : IF 3- [II4_2 = No] AND 2- [II4_3 = Si] AND 1- [IV3 = Mosaico/baldosa/madera/cerámica/alfombra]
Cluster 4 : IF 3- [II4_2 = Si] AND 2- [II4_3 = Si] AND 1- [IV3 = Mosaico/baldosa/madera/cerámica/alfombra]
Cluster 5 : IF 4- [II4_2 = Si] AND 2- [II4_3 = No] AND 1- [IV3 = Mosaico/baldosa/madera/cerámica/alfombra]

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

En definitiva los 4 clusters obtenidos pueden caracterizarse de la siguiente manera:

- Cluster 1: integrado por aquellos hogares que no tienen lavadero ni garaje y cuyos pisos son de mosaico/baldosa/madera/cerámica/alfombra.

- Cluster 2: con hogares que tienen piso de ladrillo suelto o cemento/ladrillo fijo u otros.
- Cluster 3: los hogares que componen este cluster habitan en viviendas que no tienen lavadero, pero si tienen garaje. Los pisos son de mosaico/baldosa/madera/cerámica/alfombra.
- Cluster 4: integrado por viviendas que tienen lavadero y garaje y con pisos de mosaico/baldosa/madera/cerámica/alfombra.
- Cluster 5: conformado por hogares que habitan en viviendas que tienen lavadero y no tienen garaje. Los pisos son de mosaico/baldosa/madera/cerámica/alfombra.

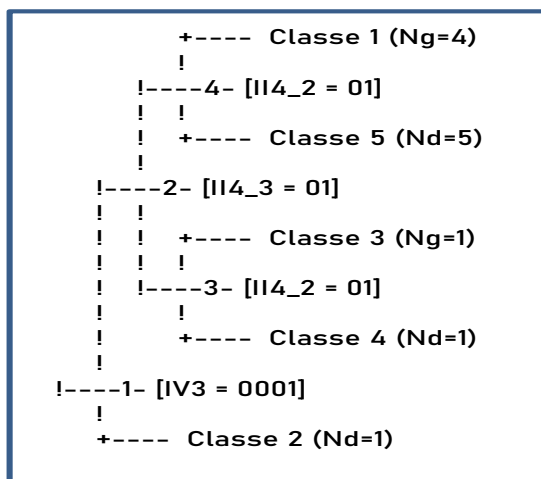


Figura 8. Árbol de cluster

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

5.2. Método SCLUST

En segundo lugar se emplea el método clustering dinámico (SCLUST) perteneciente al software SODAS. El objetivo es el mismo que en el caso anterior: clasificar los objetos simbólicos y comparar los resultados obtenidos por ambos métodos. Para ello se trabaja con cinco cluster y con las 14 variables modales. Se emplea la distancia modal de Carvalho “Tipo L1” y la opción prototipos aleatorios. La salida del módulo nos brinda la descripción de clases, a través de sus prototipos, la relación con las variables y la extensión de las clases, es decir sus elementos. El porcentaje de criterio explicado es de 91,86.

Tabla 12. Descripción de las variables intervinientes en el cluster

Nombre	Contribución con la partición	Contribución con el total de inercia	Calidad
IV1	87,84	2,85	0,96
IV3	92,58	14,64	1,01
IV4	86,12	10,48	0,94
IV6	37,99	0,33	0,41
IV7	9,09	0,02	0,1
IV8	100	0,41	1,09
IV9	80,27	2,14	0,87
IV10	57,74	1,29	0,63
IV11	97,61	21,95	1,06
II4_1	100	1,88	1,09
II4_2	100	7,22	1,09
II4_3	100	14,2	1,09

II7	88,3	18,23	0,96
II8	90,56	4,35	0,99

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

En el tabla 12 se observa la descripción de las variables. La calidad indica el poder de discriminación de la variable para la clasificación. En tal sentido las variables que mejor discriminan son tenencia de cuarto de cocina (II4_1), tenencia de lavadero (II4_2), tenencia de garaje (II4_3), tenencia de baño/letrina (IV8), desagüe del baño (IV_11) y material de los pisos (IV3). Las cuatro primeras variables mencionadas son las que más contribuyen a la partición, aunque de ellas, la tenencia de garaje es la que más contribuye a la inercia. Por su parte, la variable desagüe del baño (IV_11) es la que más contribuye a la inercia con un alto valor en la contribución a la partición y alto valor de calidad.

Tabla 13. Contribución de cada variable a los cluster

Variables	Categorías	Conjunto	Cluster				
			1	2	3	4	5
IV1	Casa	0,9	0,88	1	0,87	1	0,91
	Departamento	0,1	0,12	0	0,13	0	0,09
	Otros	0	0	0	0	0	0
	Pieza en hotel/pensión	0	0	0	0	0	0
IV3	Mosaico/baldosa/madera/cerámico/alfombra	0,75	0,9	0	0,76	1	0,84
	Cemento/ladrillo fijo	0,23	0,09	1	0,21	0	0,14
	Ladrillo suelto/tierra	0,01	0	0	0,02	0	0,02
	Otros	0,01	0,01	0	0,01	0	0
IV4	Baldosa/losa sin cubierta	0,14	0,21	0	0,15	0	0,17
	Membrana/cubierta asfáltica	0,66	0,62	1	0,57	1	0,62
	N/S. Departamento	0,04	0,07	0	0,04	0	0,01
	Caña/tabla/paja con barro/paja sola	0,14	0,06	0	0,2	0	0,15
	Pizarra/teja	0,02	0,04	0	0,01	0	0,04
	Chapa de cartón	0	0	0	0	0	0
	Chapa de metal sin cubierta	0,01	0	0	0,02	0	0,01
	Chapa de fibrocemento/plástico	0	0	0	0,01	0	0
IV6	Por cañería dentro de la vivienda	0,99	1	1	0,99	1	0,98
	Fuera de la vivienda pero dentro del terreno	0,01	0	0	0,01	0	0,02
IV7	Red pública (agua corriente)	1	1	1	1	1	1
	Perforación con bomba a motor	0	0	0	0	0	0
IV8	Si	1	1	1	1	1	0,98
	No	0	0	0	0	0	0,02
IV9	Dentro de la vivienda	0,94	0,97	1	0,92	1	0,93
	Fuera de la vivienda pero dentro del terreno	0,06	0,03	0	0,08	0	0,07
	Fuera del terreno	0	0	0	0	0	0,01
IV10	Inodoro con botón/ mochila/cadena y arrastre de agua	0,98	1	1	0,97	1	0,95
	Inodoro sin botón/cadena y con arrastre de agua (a balde)	0,02	0	0	0,03	0	0,03
	Letrina (sin arrastre de agua)	0	0	0	0	0	0,02
IV11	A red pública (cloaca)	0,44	0,47	0	0,44	1	0,36
	A cámara séptica y pozo ciego	0,37	0,51	0	0,4	0	0,5
	Sólo a pozo ciego	0,19	0,02	1	0,16	0	0,12
	0	0	0	0	0	0	0,01
	A hoyo/excavación en al tierra	0	0	0	0	0	0,01
II4_1	Si	0,93	0,96	1	0,91	1	0,92
	No	0,07	0,04	0	0,09	0	0,08
II4_2	Si	0,51	0,72	0,5	0,51	0	0,55
	No	0,49	0,28	0,5	0,49	1	0,45
II4_3	No	0,69	0,55	1	0,78	0	0,76
	Si	0,31	0,45	0	0,22	1	0,24
II7	Propietario de la vivienda y el terreno	0,64	0,75	0	0,65	1	0,62
	Inquilino/arrendatario de la vivienda	0,11	0,08	0	0,15	0	0,15
	Ocupante gratuito (con permiso)	0,09	0,06	0,5	0,05	0	0,05
	Ocupante por pago de impuestos/expensas	0,05	0,08	0	0,04	0	0,1
	Está en sucesión	0,09	0,01	0,5	0,06	0	0,07
	Ocupante de hecho (sin permiso)	0	0	0	0,01	0	0
	Propietario de la vivienda solamente	0,01	0,03	0	0,01	0	0,01
	Ocupante en relación de dependencia	0,01	0	0	0,01	0	0
	Otra situación	0	0	0	0	0	0,01
II8	Gas de red	0,71	0,89	0,5	0,63	1	0,74
	Gas de tubo/garrafa	0,28	0,1	0,5	0,36	0	0,25
	Kerosene/ leña/ carbón	0,01	0,02	0	0	0	0,02

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

La Tabla 13 muestra la contribución de cada variable a los prototipos (cluster). Sobre su base y en función de las figuras 9 a 13 se describen los clusters en función de los prototipos obtenidos. Ellos son:

- Cluster 1 (figura 9). El prototipo del primer cluster se caracteriza por hogares que viven en casas (88 %), que tienen pisos de mosaico (99 %) y cubierta exterior del techo de membrana/cubierta asfáltica (62 %). El 100 % tiene agua por cañería dentro de la vivienda procedente de red pública. Todos tienen baño con inodoro con botón, de los cuales el 97 % está dentro de la vivienda. Descargan a cámara séptica y pozo ciego (51 %). El 96 % tiene cocina, el 72 % lavadero y el 55 % no tiene garaje. El 75 % es propietario de la vivienda y el terreno. El 89 % usan gas de red (natural) para cocinar.

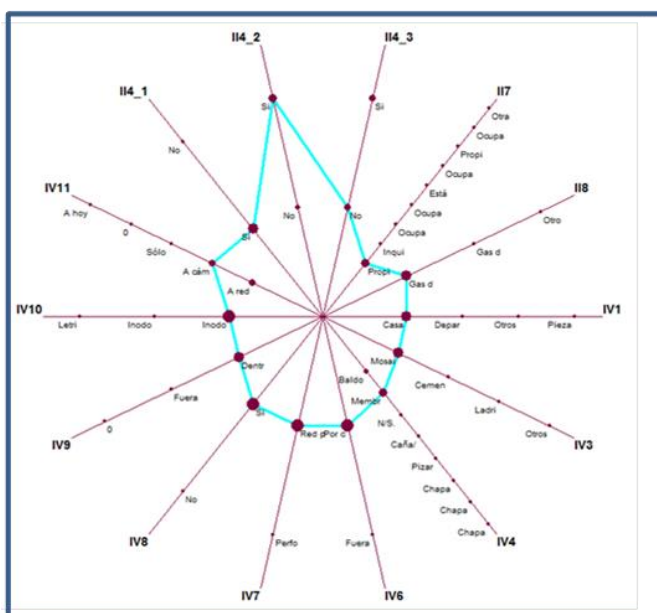


Figura 9. Cluster 1 (prototipo 1/5)

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

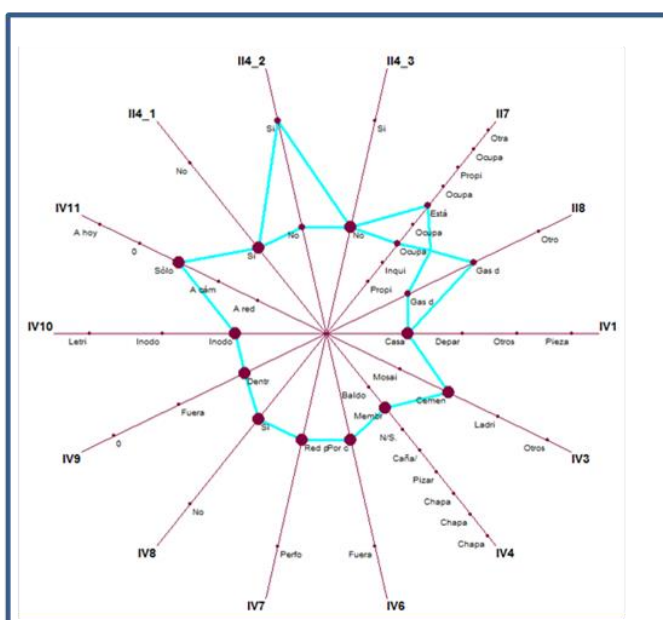


Figura 10. Cluster 2 (prototipo 2/5)

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

- Cluster 2 (figura 10). El prototipo del segundo cluster se caracteriza por hogares, que en un 100% viven en casas, tienen pisos de cemento/ladrillo fijo, tienen cubierta exterior del techo con membrana/cubierta asfáltica, tienen agua por cañería dentro de la vivienda procedente de red pública, tienen baño con inodoro con botón y con desagüe sólo a pozo ciego. El 100 % tiene cuarto de cocina, mientras que el 50 % tiene lavadero y el 100 % no tiene garaje. En cuanto al régimen de tenencia el 50 % es ocupante gratuito (con permiso) y el otro 50 % está en sucesión. El combustible utilizado para cocinar se distribuye en un 50 % en gas natural y en un 50 % en gas de tubo/garrafa.
- Cluster 3 (figura 11). El prototipo del tercer cluster presenta hogares que viven en casas (87 %), que tienen pisos de mosaico (76 %) y cubierta exterior del techo de membrana/cubierta asfáltica (57 %). El 99 % tiene agua por cañería dentro de la vivienda y procede de red pública en un 100%. Todos tienen baño, de los cuales el 92 % está dentro de la vivienda y el 97 % con inodoro con botón. Descargan a red pública (44 %). El 91 % tiene cocina, el 51 % lavadero y el 78 % no tiene garaje. El 65 % es propietario de la vivienda y el terreno. El 63 % usan gas de red (natural) para cocinar.

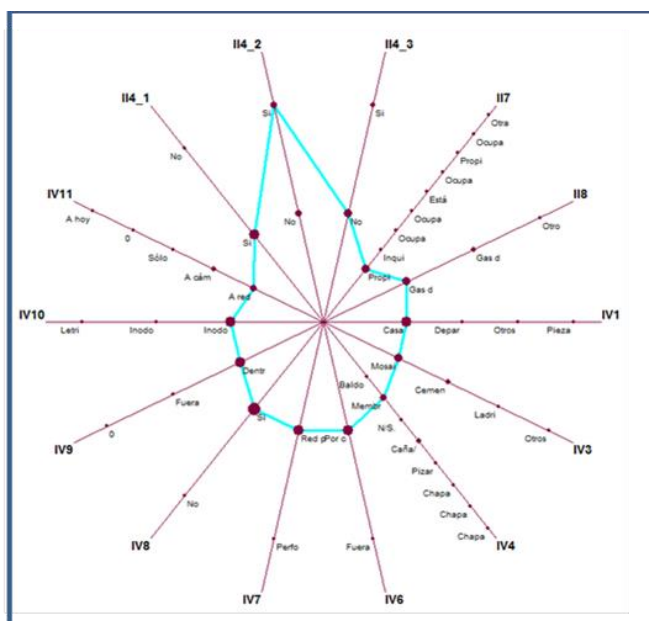


Figura 11. Cluster 3 (prototipo 3/5)

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

- Cluster 4 (figura 12). El prototipo del cuarto cluster contiene hogares, que en un 100% viven en casas, tienen pisos de mosaico, tienen cubierta exterior del techo con membrana/cubierta asfáltica, tienen agua por cañería dentro de la vivienda procedente de red pública, tienen baño con inodoro con botón y con desagüe a red pública (cloaca). El 100 % tiene cuarto de cocina y garaje y el 100 % no tiene lavadero. En cuanto al régimen de tenencia el 100 % es propietario de la vivienda y el terreno. Todos utilizan como combustible para cocinar el gas natural.
- Cluster 5 (figura 13). El prototipo del quinto cluster se caracteriza por hogares que viven en casas (91 %), que tienen pisos de mosaico (84 %) y cubierta exterior del techo de membrana/cubierta asfáltica (62 %). El 98 % tiene agua por cañería dentro de la vivienda y procede de red pública en un 100%. El 98 % tiene baño, de los cuales el 93 % está dentro de la vivienda y el 95 % tiene inodoro con botón. Descargan a cámara séptica y pozo ciego (50 %). El 92 % tiene cocina, el 55 % lavadero y el 76 % no tiene garaje. El 64 % es propietario de la vivienda y el terreno. El 74 % usan gas de red (natural) para cocinar.

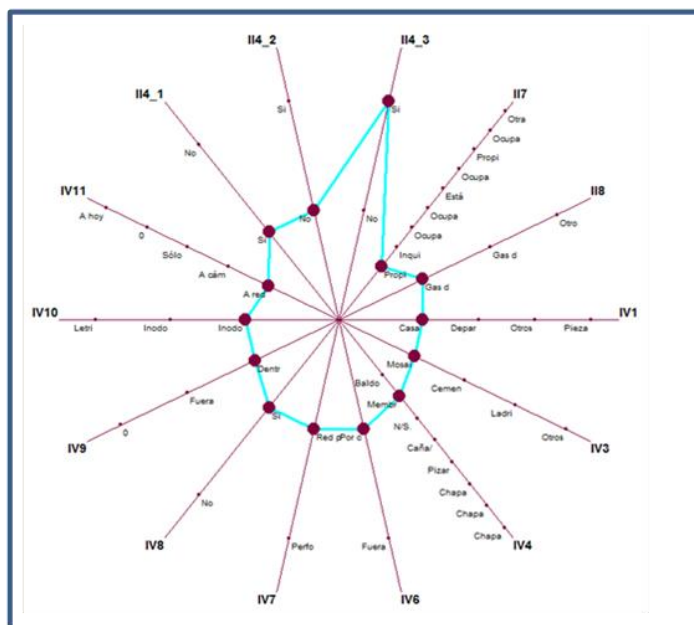


Figura 12. Cluster 4 (prototipo 4/5)

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

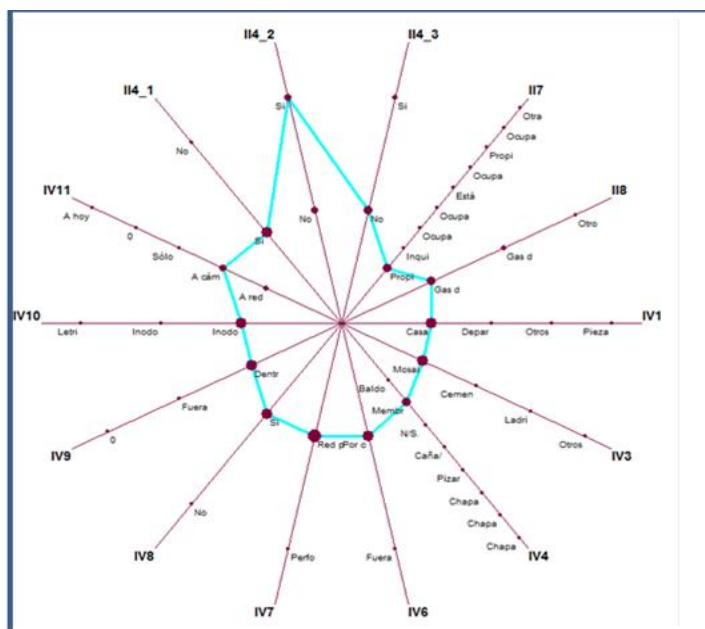


Figura 13. Cluster 5 (prototipo 5/5)

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

4.4. Regresión simbólica para variables de intervalo

La regresión simbólica se aplica a la tabla simbólica de datos construida, que toma como concepto a los deciles según el nivel de ingreso y como variables de intervalo el monto de ingreso total familiar (ITF) y el número de casas con techo deficientes (CTD). La primera es tomada como variable explicativa y la segunda como variable respuesta. Antes de realizar la regresión simbólica se comprueba la asociación entre ambas variables con una regresión simple clásica. Se observa en la figura 13 que hay una relación inversa con un coeficiente de correlación de -0,87. Esta información habilita la realización de la regresión simbólica.

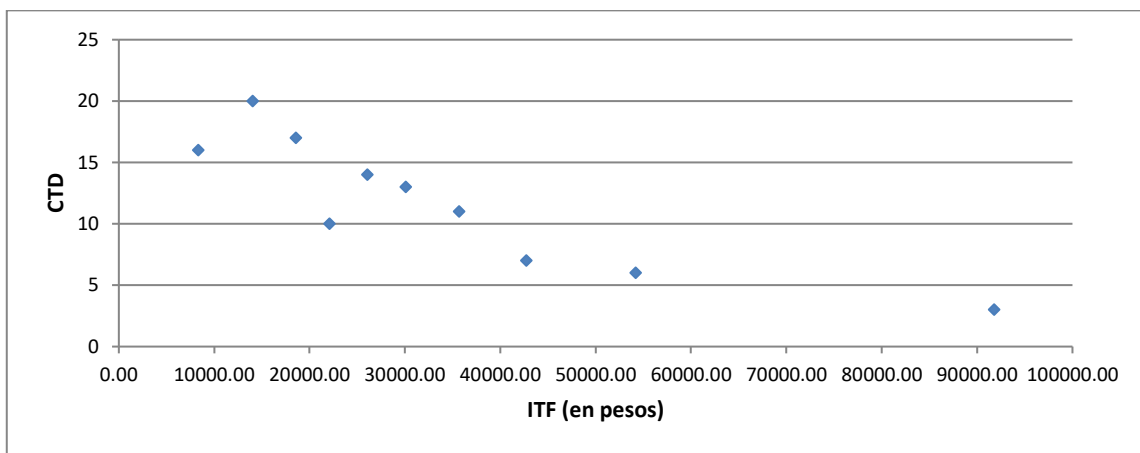


Figura 13. Regresión simple entre las variables Ingreso Total Familiar (ITF) y Número de casas con techos deficientes (CTD)

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

Tabla 14. Valores máximo y mínimo de las variables CTD (Y) e ITF (X1)

```
# A Symbolic Data Table : 10 x 2
      Y      X1
1 [0,13] [1830,11100]
2 [0,18] [11600,16300]
3 [0,17] [16400,20900]
4 [0,10] [21000,25600]
5 [0,14] [25630,31000]
6 [0,13] [31000,38000]
7 [0,9] [38400,46300]
8 [0,5] [46700,58000]
9 [0,5] [58315,77130]
10 [0,4] [79200,290000]
```

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

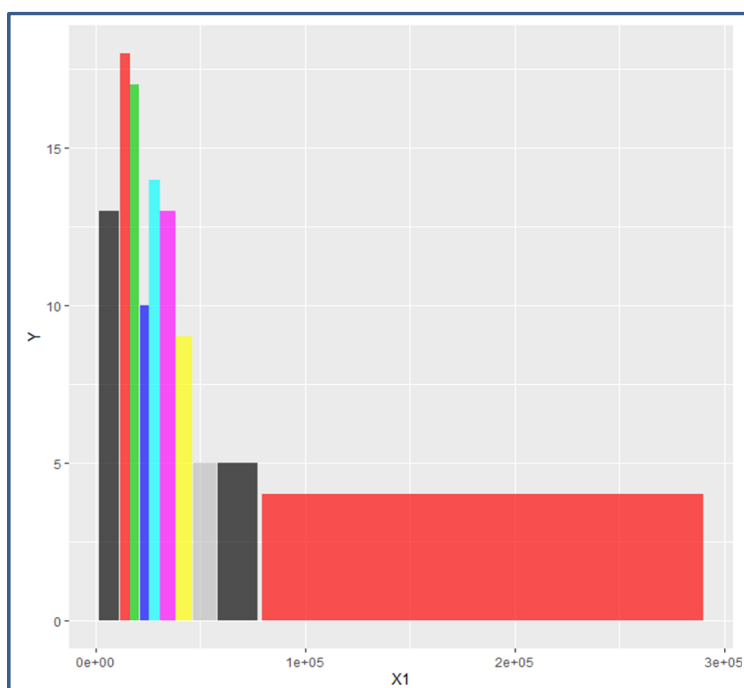


Figura 14. Regresión simbólica entre las variables ITF Y CTD

Fuente. Elaboración propia sobre la base de datos de la EPH (INDEC)

La tabla 14 muestra los valores máximo y mínimo de las dos variables consideradas, para cada uno de los deciles, mientras que la figura 14 muestra la salida gráfica de la regresión simbólica realizada. Se observa que existe una relación lineal inversa entre el ITF y las CTD. Se observa una mayor concentración de viviendas con cubierta del techo inconvenientes en los primeros deciles. Además, a medida que aumentan se advierte una mayor amplitud entre los valores máximo y mínimo.

6. Conclusiones

La EPH es un instrumento que tiene la finalidad de proporcionar información sobre las características sociales y económicas de la población. En cada una de sus bases, ya sea hogar o individual, contiene una gran número de variables que permiten conocer las características mencionadas. La obtención de los datos se realiza a través de un diseño muestral que se aplica a los aglomerados urbanos de la Argentina. En el caso del AGSJ, la muestra alcanzó en el primer trimestre de 2019 un total de 527 casos.

Dada su magnitud, se hace necesario buscar procedimientos estadísticos que permitan la reducción de los datos y que posibiliten recabar la información más importante, oculta tras ese gran volumen. El SDA es un buen recurso para conseguir tal objetivo. En este caso sus métodos fueron aplicados para conocer las características de las viviendas habitadas por los hogares del AGSJ.

Los OS muestran características similares en cuanto a ciertas variables, como que el tipo de vivienda que prevalece en ellos es la casa. De igual modo, en los seis OS analizados, las viviendas tienen agua por cañería dentro de las viviendas procedente de red pública. Algo similar ocurre con la tenencia de baño dentro de la vivienda y el uso de inodoro con botón o cadena. En el caso del baño se observan diferencias en el desagüe. Si bien hay un predominio del desagüe a cámara séptica y pozo ciego y a red pública, persisten situaciones de desagüe sólo a pozo ciego. La mayor parte de los hogares tienen cocina y en menor proporción lavadero y garaje. También prevalece el régimen de tenencia propietario de la vivienda y el terreno y el uso de gas de red como combustible para cocinar. En cuanto a las variables de intervalo, el número de ambientes aumenta con los deciles al igual que el número de miembros del hogar.

Al aplicar el método de agrupamiento DIV se obtiene cinco clusters. En líneas generales el cluster 1 contiene los OS más bajos en cuanto a los deciles, mientras que el cluster 5 contiene a los más altos. El OS = 10 es separado en otro cluster (4) y es el que presenta las mejores condiciones de la vivienda. Por su parte, el cluster 2 presenta las situaciones más desfavorables. La variable principal que produce la primera división es el piso de mosaico/baldosa/madera/cerámica/alfombra. Le siguen las variables tenencia de garaje y la tenencia de lavadero.

Por su parte, el método SCLUST muestra los 5 cluster. En este caso pueden apreciarse mejor las diferencias entre los OS. El principal contraste se produce entre los cluster 2 y 4. Éste último presenta el porcentaje máximo en aquellas variables favorables como la tenencia de cocina y garaje, tenencia de baño con desagüe a red pública y piso de mosaico. Por el contrario no registra casos en variables desfavorables como desagüe a cámara séptica y pozo ciego, desagüe sólo a pozo y piso de cemento/ladrillo fijo. Sólo llama la atención en este caso la ausencia de lavadero.

Finalmente la regresión simbólica muestra que existe una asociación entre el ITF y las CTD. Es decir a mayor ingreso del hogar se produce una disminución del número de casas que tienen materiales deficientes como cobertura del techo. A manera de cierre puede afirmarse que los ingresos no explican por sí solos las características de las viviendas. Al comparar los OS se observan más similitudes que diferencias.

Referencias bibliográficas

- Diday, E. (1987). Introduction à l'approche symbolique en analyse des données. *Premieres Journees Symbolique - Numerique*. CEREM ADE, Universite Paris Dauphine, 21-56.
- Diday, E. (1988). The symbolic approach in clustering and related methods of data analysis: the basic choices. In: H.H. Bock (Ed.). *Classification and related methods of data analysis*. Aachmen, Amsterdam, 673-684.
- Diday E. (1991). Des objets de l'Analyse des Données aó ceux de l'Analyse des Connaissances in *Induction symbolique et numerique*. Y. Kodratoff and E. Diday edit. CEPADUES EDITIONS, Toulouse, France.
- Instituto Nacional de Estadísticas y Censos. República Argentina. (25 de abril de 2020). *Aglomerados de la Argentina de 500.000 habitantes y más*. https://www.indec.gob.ar/ftp/cuadros/poblacion/aglomerados_argentina_500_mil.pdf
- Lima Neto, E., De Carvalho, F. (2008). Centre and Range method for fitting a linear regression model to symbolic interval data. *Computational Statistics and Data Analysis*. Vol. 52, 1500-1515. <https://www.sciencedirect.com/science/article/abs/pii/S0167947307001934>.
- Lima Neto, E., De Carvalho, F. (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis*. Vol 54, 333-347. <https://www.sciencedirect.com/science/article/abs/pii/S0167947309003065>.
- Lima Neto, E., Cordeiro, G., De Carvalho, F. (2011). Bivariate symbolic regression models for interval-valued variables. *Journal of Statistical Computation and Simulation* . Vol. 8, 1727-1744. <https://www.tandfonline.com/doi/abs/10.1080/00949655.2010.500470>.
- Mallea, A., Herrera, M., Lund, M. (2018). *Análisis simbólico de datos: una potente herramienta para Big Data*. Repositorio Institucional de la UNLP. <http://sedici.unlp.edu.ar/handle/10915/69945>
- Ruiz Shulcloper, J (1997). Bases conceptuales para una teoría de objetos Simbólicos. *Revista Computación y sistemas*. Vol 1, N° 1, 13-20. <https://www.repositoriodigital.ipn.mx/bitstream/123456789/14894/1/art2.%20Vol1%20N%201.pdf>



Esta obra se encuentra bajo Licencia Creative Commons Reconocimiento-NoComercial 4.0. Internacional. Reconocimiento - Permite copiar, distribuir, exhibir y representar la obra y hacer obras derivadas siempre y cuando reconozca y cite al autor original. No Comercial – Esta obra no puede ser utilizada con fines comerciales, a menos que se obtenga el permiso.