



Universidad de Chile
Facultad de Ciencias Sociales
Departamento de Sociología

ESTILOS DE VIDA A PARTIR DEL CONSUMO EN CHILE

Tesis para optar al título de Sociólogo

IGNACIO AGLONI JORQUERA

Profesor Guía: Rodrigo Asún

Santiago de Chile
Julio, 2019

Índice general

Resumen.....	4
1. Problematización: estilos de vida a partir del consumo en Chile.....	5
2. Diseño de investigación.....	11
2.1 Pregunta de investigación.....	11
2.2 Objetivos.....	11
2.2.1 Objetivo general:.....	11
2.2.2 Objetivos específicos:.....	11
2.3 Hipótesis.....	12
3. Antecedentes teóricos y empíricos.....	13
3.1 Sobre las necesidades.....	13
3.2 Sobre el consumo.....	16
3.3 Sobre los estilos de vida.....	19
4. Diseño metodológico.....	23
4.1 Descripción general de la VIII Encuesta de Presupuestos Familiares.....	23
4.1.1 Estructura de los datos.....	24
4.2 Técnicas utilizadas.....	25
4.2.1 Técnicas de <i>clustering</i> y clasificación.....	25
4.2.2 <i>Clustering</i> a través de K-Medias.....	28
4.2.3 Técnicas para determinar el valor óptimo de K.....	29
4.2.4 Análisis de Componentes Principales.....	31
4.3 Plan de análisis.....	33
5. Resultados.....	35
5.1 Sobre la construcción de una clasificación.....	35
5.2 Construcción de variables activas en la clasificación.....	37
5.2.1 Nivel de desagregación del gasto.....	37
5.2.2 Definición de gasto para la clasificación.....	42
5.2.3 Reducción de dimensionalidad con ACP.....	46
5.3 Construcción, identificación y descripción de los estilos de vida de los hogares chilenos.....	50
5.3.1 Variables suplementarias para la caracterización de los clusters.....	51
5.3.2 Descripción general de los clustersobtenidos con K-medias.....	56
5.3.3 Denominación de los clusters.....	59
5.3.4 Los estilos de vida de los hogares chilenos.....	64

6.	Conclusiones	84
6.1	Reflexiones metodológicas	84
6.2	Reflexiones sobre los estilos de vida.....	85
6.2.1	Sobre la flexibilidad de los estilos de vida.....	85
6.2.2	Estilos de vida y estratificación socioeconómica	87
6.3	Sobre el carácter construido de la clasificación	88
6.4	Potenciales líneas de investigación	89
7.	Bibliografía	91
8.	Anexos.....	96
	Anexo 1: división y grupo de la Clasificación del Consumo Individual por Finalidades	96
	Anexo 2: detección y depuración de <i>outliers</i> o casos atípicos	99
	Detección de <i>outliers</i>	99
	Depuración de <i>outliers</i>	100
	Anexo 3: Descripción de variables suplementarias.....	103

Resumen

Los esquemas de estratificación tradicional, basados en la posición de los individuos en el mercado del trabajo, han visto mermada su capacidad para explicar la diferenciación social y los nuevos problemas que emergen a consecuencia de las transformaciones sociales y culturales que ha traído consigo la profundización de las sociedades neoliberales. Ante esto se hace necesario intentar nuevas taxonomías que atiendan a los espacios que han ido cobrando relevancia para las personas, como es el del consumo. Utilizando datos de la VIII Encuesta de Presupuestos Familiares 2016-2017, el propósito de esta tesis fue generar una clasificación de los hogares chilenos en distintos “estilos de vida”, basados en sus gastos de consumo, identificando los rasgos distintivos de cada estilo de vida y las características de los hogares que los conforman. Para lograrlo se diseñó una metodología que incluye técnicas de reducción de información mediante Análisis de Componentes Principales, y *clustering* a través de K-medias, lo que decantó en una paleta de 8 clusters de estilos de vida. El presente trabajo tiene un énfasis en la estrategia metodológica, esperandotambién servir de hoja de ruta para quienes se vean enfrentados a un desafío similar en relación a sus objetivos y el tipo de datos con los que cuentan.

Palabras clave: consumo, gasto, estilos de vida, *clustering*, análisis de componentes principales, K-medias.

1. Problematización: estilos de vida a partir del consumo en Chile

Desde sus inicios, la sociología ha hecho esfuerzos por establecer criterios que permitan clasificar adecuadamente a los individuos y grupos, desde el punto de vista de la diferenciación social (Aguilar, 2009). Esto ha implicado la búsqueda constante de medidas empíricas que permitan delimitar estos constructos. El tránsito por distintos enfoques teóricos y énfasis sobre cuál o cuáles son los rasgos fundamentales para diferenciar a los individuos y grupos, ha decantado en la exploración de diversas metodologías y variables explicativas.

Para las teorías de inspiración marxista, la diferenciación fundamental está dada por la clase, la cual se funda en la relación que tienen los individuos o grupos con los medios de producción. Ahora, operacionalizar la relación con los medios de producción en el sentido de determinar la pertenencia a una clase, resulta complejo desde los métodos tradicionales de producción de datos sociales, como son las encuestas, censos o registros administrativos.

Por otra parte, Max Weber identificó diferentes clasificaciones sociales, todas basadas en formas de distribución del poder en la sociedad: las clases, los estamentos y los partidos. La situación de clase para Weber estaba definida por la “posición ocupada en el mercado”, por lo que la pertenencia a una clase desde las teorías de inspiración weberiana tenían una aproximación más literal en el plano empírico: la posición en el mercado laboral, que se puede entender como la ocupación, la que opera como un *proxy* de las oportunidades que tendrían los individuos en relación a su destino personal y sobre su acceso a bienes y servicios. La posición en el mercado laboral también fue una variable de clasificación importante para la tradición funcionalista. Esta reflejaría adecuadamente la clase, al traducirse en recompensas *ad hoc* a la importancia que la sociedad le otorga a la ocupación desempeñada (Aguilar, 2009).

Sin embargo, como indican Azocar y Mayol (2011) y ya habían adelantado otros autores como Moulian (2002), ante el debilitamiento de vínculos asociativos como la política o el trabajo, otros fenómenos cobran relevancia a la hora de explicar la integración y la diferenciación en la sociedad chilena, asomando el consumo como un espacio de relaciones cargado de potencia explicativa. En relación a esta transformación, Araujo y Martuccelli (2012) sostienen que en Chile la instalación del proyecto neoliberal a comienzos de los años 70 produjo una verdadera

“revolución cultural”, que tuvo su núcleo central en el tránsito, más que desde el predominio del Estado al del mercado, desde el predominio de la producción hacia el del consumo.

Considerando la propuesta weberiana, pareciera que ante este escenario sería más relevante utilizar no las clases sino los estamentos como foco de interés para explicar las trayectorias vitales de los individuos. Weber define los estamentos como comunidades de carácter amorfo que, a diferencia de la clase, condicionada por motivos puramente económicos, se constituye alrededor de la idea de una estimación social específica, positiva o negativa, del “honor”, adscrito a alguna cualidad común a muchas personas. Además, el honor correspondiente al estamento, no debe necesariamente relacionarse con alguna situación de clase en particular, “poseedores y desposeídos pueden pertenecer al mismo estamento y esto ocurre con frecuencia” (Weber, 2002: 688). Entonces, donde se expresa el contenido del honor correspondiente a un estamento es en los “modos de vida”. Finalmente, el autor señala que “las «clases» se organizan según las relaciones de producción y de adquisición de bienes; (y) los «estamentos», según los principios de su consumo de bienes en las diversas formas específicas de su «manera de vivir»” (Weber, 2002: 692).

A finales de los 90’ en Chile se generó una serie de trabajos que de alguna manera inauguran la preocupación por el consumo, a causa de las consecuencias de la modernización neoliberal y la ampliación del acceso al crédito. Uno de ellos es el Informe de Desarrollo Humano 1998, titulado “Las paradojas de la modernización”. En esta investigación se plantea que el consumo se ha ido expandiendo en Chile desde los 80’, y con mucha mayor velocidad desde los 90’, pero no como resultado de un periodo de bonanza y mayor liquidez de los hogares, sino por la expansión del crédito, que ha provocado un severo endeudamiento en los hogares de clase media (PNUD, 1998). En este sentido, el PNUD problematiza un modelo de desarrollo donde la certidumbre que demuestran los chilenos al adquirir bienes vía crédito no se condice con la sensación de inseguridad laboral o la percepción que tienen de falta de oportunidades.

El diagnóstico del PNUD señalaba diversas aristas en el proceso de modernización que denunciaban falencias en el modelo económico de libre mercado y la estructura laboral, principalmente. Si bien se trataba de una crítica severa, era una crítica sistémica, que apuntaba a la corrección del modelo de desarrollo. El “Chile actual: Anatomía de un mito” de Moulian

(2002), por otra parte, presenta una crítica más dura, recriminando al modelo y apuntando a su superación.

Moulian (2002) señala que se ha transformado el rol del Estado frente a la sociedad. El Estado ha perdido su centralidad en la redistribución de los ingresos, siendo sustituido de alguna manera por la ficción de la igualdad a través de la masificación del crédito. El autor propone que se ha migrado desde una matriz populista a una matriz productivista – consumista. Se habla de matriz productivista por el sometimiento institucionalizado del trabajo al capital, justificado por la mística de la eficiencia y la competitividad; y consumista por los dispositivos económicos y simbólicos que compensan el énfasis productivista.

El crédito para Moulian es una de las formas de disimular la desigualdad social, ligándolo con la idea de la esperanza del acceso al mercado, generando la ilusión de igualdad y libertad. Esta idea se expresa en la aparición del mall, nuevo lugar de la integración social, como lo fue la iglesia o la plaza pública en el pasado (Moulian, 2002).

Otro de los trabajos destacados de finales de siglo en Chile es “La irrupción de las masas y el malestar de las élites” de Eugenio Tironi (1999). En esta investigación el autor vislumbra un cambio en los patrones culturales de la sociedad chilena, a causa de la desestructuración provocada por el proceso de modernización. Para el autor, Chile se convierte en la década de los 90’ en un nuevo tipo de sociedad, donde el carácter del sujeto se trastorna. El salto hacia una sociedad de consumo se produce no cuando las empresas deciden tratar a las personas como consumidores, sino que cuando estas comienzan a comportarse como tales; cuando las personas se comportan antes como individuos que como comunidades, antes como defensores de derechos e intereses privados que como promotores de fines colectivos, antes como defensores de la transparencia del mercado que como demandantes de beneficios del Estado (Tironi, 1999).

Emerge entonces el crédito y su revés, el endeudamiento, como un costo del progreso, al igual que la destrucción del medio ambiente, la congestión vehicular, el individualismo, etc., y como consecuencia del crédito aparecen las masas que armadas con sus tarjetas intentan fundar su identidad a través del consumo (Tironi, 1999). El revés de este movimiento es el malestar de las élites y sobre todo las clases medias acomodadas, que ven amenazada su posición por masas consumistas que intentan imitar su estilo de vida y alcanzar su estatus a través de un discurso de éxito económico mediante el consumo.

La centralidad otorgada al fenómeno del consumo a través de los trabajos del PNUD (1998), Moulian (2002) y Tironi (1999), se ve materializada durante la década siguiente en ideas más concretas para clasificar a los grupos sociales en Chile. Ya hacia el año 2009, Omar Aguilar proponía actualizar los métodos de estratificación, considerando la diferenciación material y simbólica, mientras que en el mismo documento compilatorio Carlos Catalán presentaba su propuesta acerca del consumo y la segmentación social. Se trata de la publicación titulada “El arte de clasificar a los chilenos” del 2009, en la cual, si bien aún priman los modelos de clasificación que buscan delimitar a la sociedad en grupos de acuerdo a un enfoque de clases sociales basado en la situación ocupacional, ya se comienza a vislumbrar una apertura de los marcos teóricos para observar e incluir el fenómeno del consumo.

En la actualidad, los esquemas clasificatorios basados en la idea de clase social se han ido viendo reemplazados por clasificaciones que buscan aproximarse más a los estilos de vida de las personas y hogares, para la tarea de explicar o predecir ideas políticas, preferencias de consumo, entre otros aspectos contingentes en la sociedad actual. Prueba de esto es que la Asociación de Investigadores de Mercado (AIM) publicó a principios del 2019 la nueva estratificación de su marco muestral, que transita desde la tradicional clasificación en 5 grupos (ABC1, C2, C3, D y E), a una clasificación de 7 grupos (AB, C1a, C1b, C2, C3, D y E), declarando que: “cambiamos desde un GSE más bien anclado en la idea de clase social, donde el acceso a ciertos puestos de trabajo y niveles educacionales definían el nivel de clasificación, a uno que incorpora el ingreso disponible en el hogar como una variable clave en los hábitos y conductas de las personas. Esta nueva clasificación recoge en mejor medida la heterogeneidad de las configuraciones familiares y las dinámicas de ciclo de vida en las trayectorias individuales” (GFK, 2019: 2). El ingreso disponible se utiliza para estimar las posibilidades de consumo.

Entonces, si bien podría sonar agresivo hablar de la preeminencia de la sociedad de consumo en Chile, mientras existe aún una proporción de la población que se encuentra muy restringida en cuanto a su acceso al mercado, es innegable el lugar preponderante que tiene la relación con los objetos y prácticas como ámbito de reproducción e integración social (Baudrillard, 1991; Lipovetsky, 2003; Baumann, 2011). La segmentación que el consumo ofrece no se reduce a las diferencias de acceso, entran en juego también preferencias, deseos y motivaciones, *“no existe una sola segmentación de la sociedad por el consumo. Lo que hace el consumo es entregar una*

diversidad de posibles segmentaciones” (Catalán, 2009: 42, el énfasis es del autor). Esto se relaciona con la lógica weberiana de las agrupaciones estamentales, donde no solo la situación de los grupos en el mercado explica la expresión de ciertos estilos de vida, también emergen otras variables, que ayudarían a explicar por qué la intención expresiva de los *habitus* mediante el consumo y el uso del tiempo difiere en grupos que comparten un nivel de ingreso o categoría en el mercado del trabajo.

Por otra parte, el contexto de elevada desigualdad en Chile complejiza la comprensión de las motivaciones detrás de la *performance* de consumo de los individuos. Entre el 2015 y 2017, el índice de Gini, utilizado como indicador de desigualdad en ingresos, sufrió en Chile un aumento que lo llevó a superar el 0.5 (MDS, 2018a)¹, a pesar de en 2015 ya estar entre los más altos, con el cuarto lugar a nivel mundial y el segundo en la OCDE, solo después de México (OCDE, 2018). Por otra parte, López, Figueroa y Gutiérrez(2013), en publicación de la Facultad de Economía y Negocios de la Universidad de Chile, utilizando datos del Servicio de Impuestos Internos, señalan que el 0,01% de la población de Chile acumula el 10,1% del ingreso total del país. Ronald Inglehart propone que los países que hayan logrado satisfacer las necesidades de seguridad económica y personal para proporciones mayores de su población, serán los que presenten mayor grado de valores posmaterialistas, lo que se traduce en una mayor preocupación por el bienestar subjetivo, el empoderamiento en el ámbito laboral, la preocupación por el medio ambiente y los temas estéticos e intelectuales, entre otros, mientras que cuanto menos garantizadas están esas necesidades (y para menor número de personas), será mayor el grado de valores materialistas que se podrá encontrar en su sistema de valores predominante (Inglehart, 1991). Sin embargo, en el contexto de alta desigualdad que se presenta en Chile, ciertamente hay una convivencia de distintas estructuras de sentido, que se traducen en discursos simbólicos, a través del consumo de objetos y prácticas, muy variados.

Esta conclusión ciertamente complejiza la comprensión de un escenario social colmado de transacciones materiales, que no son explicadas suficientemente por diferencias en el acceso. También son troqueladas por diferencias fundamentales en las estructuras valorativas y los *habitus*, que modelan las intenciones expresivas de los estilos de vida. En este sentido, lo que interesa en este trabajo es conocer cómo se expresa esta multiplicidad de aspectos materiales y

¹ El índice de Gini tiene un rango de 0 a 1, donde el cero representa la igualdad perfecta.

culturales a través del gasto, como vestigio objetivo y medible de la performance de los hogares en el espacio social, y cómo se traduce en una particular paleta de estilos de vida.

Keister *et al.* (2016) señalan que debido a que la cultura estadounidense fomenta y recompensa el consumo de bienes y servicios a partir de la primera infancia, el gasto refleja simultáneamente la posición estructural, las preferencias y la identidad. Por lo que, como tales, los perfiles de gasto, o la forma en que un hogar asigna los ingresos disponibles a través de un gran número de opciones de gasto, son excelentes medidas de resumen de los diversos incentivos y presiones que enfrenta un hogar, incluidas las preferencias, así como las limitaciones y oportunidades estructurales. En este sentido, la cultura chilena no pareciera discrepar mucho de la norteamericana, por lo que los perfiles de gasto son una forma ideal de identificar estilos de vida, porque abarcan un amplio conjunto de comportamientos socialmente relevantes y proporcionan indicadores claros de la posición del hogar en el espacio social mientras se mantienen empíricamente manejables.

Los gastos son marcadores de la estructura social y económica porque reflejan diferencias en los ingresos, la educación, la estructura familiar, el lugar de residencia, la etnia, la nacionalidad, etc. Dichos factores afectan los gastos de diferentes maneras pero, lo que es más importante, interactúan de manera matizada para determinar el gasto. Por ejemplo, el ingreso necesariamente restringe el gasto total, pero también dicta la asignación del gasto en varias categorías: si el alquiler, la comida, los servicios básicos u otros gastos necesarios utilizan una gran proporción del ingreso, a una familia le puede quedar poco dinero para el gasto discrecional en lujos, como vacaciones en el extranjero, salidas periódicas a restaurantes, etc.

Utilizando los datos de la VIII Encuesta de Presupuestos Familiares (en adelante VIII EPF) como fuente de información para el análisis del consumo de los hogares, esta tesis parte de la premisa de que el gasto, como expresión material del consumo, es un reflejo de preferencias, identidades y de una posición en la estructura social, que tiene el potencial de funcionar como una ventana para observar los estilos de vida (Bourdieu y Passeron, 1977; Keister et al., 2016).

2. Diseño de investigación

A continuación se expone la pregunta que guiará este trabajo, así como sus objetivos e hipótesis.

2.1 Pregunta de investigación

¿Cuáles son los estilos de vida de los hogares chilenos configurados a partir de sus gastos de consumo?

2.2 Objetivos

2.2.1 Objetivo general:

1. Clasificar a los hogares chilenos en distintos estilos de vida a partir de sus gastos de consumo, identificando los rasgos distintivos de cada estilo de vida y las características de los hogares que conforman cada uno de ellos.

2.2.2 Objetivos específicos:

- 1.1. Desarrollar una propuesta metodológica estructurada que resulte adecuada para la clasificación de hogares de acuerdo a sus gastos de consumo, identificando limitaciones y bondades de las técnicas implementadas para las características del set de datos utilizado.
- 1.2. Caracterizar e identificar a los distintos estilos de vida de los hogares chilenos, a partir del análisis de las variables de gastode consumo incluidas en la construcción del modelo de clasificación (variables activas).
- 1.3. Caracterizar a los distintos estilos de vida de los hogares chilenos, a partir del análisis de variables sociodemográficas y socioeconómicas no incluidas en el modelo (variables suplementarias).

2.3 Hipótesis

A partir de los gastos en consumo final de los hogares chilenos es posible identificar distintos estilos de vida, los que se relacionan con aspectos materiales: un cierto nivel de ingresos, o la presencia de *shocks* externos como una enfermedad u otro imprevisto; pero también con gustos y preferencias, que se verán influenciados por variados sistemas de valores y *habitus*. Esto tendrá como consecuencia que los grupos resultantes de la clasificación no serán un mero espejo de la estratificación socioeconómica, sino que reflejarán una dinámica más compleja en la construcción de los estilos de vida de los hogares que ciertamente desborda y enriquece lo explicado por los ingresos.

3. Antecedentes teóricos y empíricos

Para abordar el fenómeno de la configuración de los estilos de vida es pertinente transitar por los conceptos de necesidad y consumo. No es posible asimilar la discusión acerca de los estilos de vida sin incorporar el consumo, que está a la base de los discursos simbólicos en la sociedad actual, así como no es posible entender el consumo sin desmontarlo del concepto de necesidad.

3.1 Sobre las necesidades

Las principales teorías propuestas sobre el concepto de necesidad desde las ciencias sociales pueden ser distribuidas sobre un eje continuo que va desde el universalismo al culturalismo/historicismo, y utilizando la distinción entre necesidad como carencia, como estado; y la necesidad como objeto o práctica externos al sujeto (satisfactores).

El polo del universalismo corresponde al extremo del eje que representa a las necesidades como objetivas y compartidas por todas las personas independiente de su arraigo histórico o cultural; las necesidades serían, entonces, “reales”. Entre las propuestas más cercanas a una visión universalista de las necesidades se encuentra la Teoría de las Necesidades de Maslow, por un lado, y por otro los postulados de Halbwachs y Chombart de Lauwe, al alero del positivismo francés de la primera mitad del siglo XX.

Maslow propone una exhaustiva categorización y jerarquización de las necesidades (como motivaciones) representada en lo que se conoce como la Pirámide de Maslow (Alguacil, 2000). Esta está compuesta por 5 niveles que van ascendiendo en subjetividad desde la base hasta la cima. En ese mismo orden encontramos: i) Necesidades fisiológicas; ii) Necesidades de salud y seguridad; iii) Necesidades de pertenencia y amor; iv) Necesidades de estima; y finalmente, v) Necesidades de autorrealización o metanecesidades. Estas últimas se encuentran en un bloque sustantivamente diferente y fueron escasamente desarrolladas por Maslow, a diferencia de las demás (Alguacil, 2000).

Si bien Maslow elabora una gradiente desde lo objetivo a lo más subjetivo, en la cima de su pirámide, no por subjetiva deja de ser categórica, jerárquica y universal su teoría. Las críticas a la visión maslowiana provienen desde distintos flancos. Respecto a la performance de los

individuos en su esquema, Doyal y Gough rechazan que estos actúen motivados por impulsos que los guíen en una lógica jerárquica a través del continuo ascendente desde las necesidades más básicas hasta las de autorrealización (Alguacil, 2000). No existe una determinación universal para satisfacer, por ejemplo, las necesidades de seguridad antes que las de pertenencia y amor. Otra crítica fundamental a la propuesta de Maslow proviene de la perspectiva historicista y dialéctica que defiende Agnes Heller: las necesidades se desarrollan por la continua superación de estadios históricos, refundándose nuevos sistemas de necesidades con cada revolución de los modos de vida (Heller, 1978). Las necesidades no serían ontológicas.

También se asientan sobre una perspectiva universalista de las necesidades los postulados de Maurice Halwachs y Pierre-Henry Chombart de Lauwe, quienes, cobijados en el positivismo francés de la primera mitad del siglo XX, erigen una teoría del consumo desde la concepción de la necesidad como algo dado, como variable independiente que hay que descubrir y conocer.

Chombart de Lauwe es el continuador de la teoría de Halwachs, complementando la visión sobre las necesidades de este con una conceptualización de las “aspiraciones” (Alonso, 2009). El autor propone algunas distinciones que serán recurrentes en teorías posteriores sobre las necesidades. En primer lugar distingue entre lo que llama la necesidad-objeto y la necesidad-estado. Ambas son necesidades y por tanto son distintas de las aspiraciones. Las necesidades-objeto son, básicamente, lo que la mayoría de las teorías actuales llaman satisfactores, objetos o prácticas que están fuera del sujeto y que le sirven como herramientas para la satisfacción de una necesidad que es interior y objetiva. La necesidad vista como carencia – en el sentido sartreano – es lo que Chombart de Lauwe entiende como necesidad-estado, como un estado interior que refleja la falta de un elemento para su satisfacción.

Ahora, “mientras que las necesidades están originadas por restricciones físicas, inconscientes o sociales, las aspiraciones son «deseos vueltos hacia un fin, una finalidad u objeto»; además, y esta es su principal característica, «están orientados por imágenes, signos, símbolos»” (Alonso, 2009: 54). Las aspiraciones se comportan dialécticamente: las necesidades-aspiraciones se convierten paulatinamente, a medida que se va logrando su satisfacción generalizada, en necesidades-obligaciones (necesidades en el sentido estricto), teniendo en esto el Estado un rol fundamental, de satisfacer las segundas y siempre tomar nota de la evolución de las primeras, para así evitar que “la población reaccione” (Alonso, 2009).

En el medio del continuo entre universalismo y culturalismo/historicismo se encuentran las propuestas de los eco-humanistas: Galtung (1977), Max-Neef (1998) y Doyal y Gough (1994), quienes tienen en común la preocupación por la satisfacción de las necesidades en un contexto de depredación del medio ambiente, apuntando a imperativos morales propios del humanismo. Estos autores se sitúan en el medio de la polaridad entre universalismo y culturalismo/historicismo, planteando que, una vez advertida la fundamental diferencia entre necesidades y satisfactores, o entre necesidades y aspiraciones, se puede establecer un set de necesidades que sean universales, siendo los satisfactores y los bienes económicos determinados cultural e históricamente.

Finalmente, los postulados de Ágnes Heller y Jean Baudrillard, pueden ser situados en el polo culturalista/historicista de la concepción de las necesidades. Ambos autores provienen de la corriente de pensamiento marxista y completan una conceptualización sobre las necesidades que es ambigua en Marx. Para Heller, las necesidades no se pueden categorizar en falsas o verdaderas, reales o irreales, pues son determinadas históricamente. Todas las necesidades sentidas y formuladas por las personas conscientemente deben ser consideradas como reales (Izquierdo, 1997).

Según señala Alguacil (2000), se encuentran 4 categorías de clasificación de las necesidades en el trabajo de Ágnes Heller: i) Necesidades existenciales, que son necesidades básicas tanto materiales como no materiales; ii) Necesidades alienadas, de carácter cuantitativo (como la necesidad de poder, ambición, acumulación de riqueza); iii) Necesidades no alienadas, de carácter cualitativo (como la necesidad de amor, amistad, de desarrollo personal); y las Necesidades radicales, que son generadas dentro del sistema capitalista, y que como condición de satisfacción implican la superación de dicho modo de producción. Las necesidades son reales en tanto que son vividas de manera consciente, pero no son universales, dado que pueden ser superadas históricamente dando paso a nuevos sistemas de necesidades.

En el mismo sentido Baudrillard plantea que “no hay necesidades sino porque el sistema las necesita” (Baudrillard, 1991: 80). Radicaliza la postura de que las necesidades no están dadas ontológicamente, declarando incluso que si el sistema no necesitara que los trabajadores consuman alimentos o se reprodujeran, estos no lo harían. La crítica de Baudrillard apunta a las teorías que asumen las necesidades como naturales y ocultan su componente ideológico,

ocultando a su vez relaciones de dominación. Es por eso que una Teoría de las necesidades no tiene sentido; no puede haber más que una Teoría del concepto ideológico de necesidad.

Y esta naturalidad en la utilización del concepto de necesidad, representada en el lenguaje típico de la sociología, la psicología y la economía, no es inocente, puesto que “la tautología es siempre la ideología racionalizante de un sistema de poder” (Baudrillard, 1991: 64), resguardando así la integridad de los modos de operar del sistema de producción.

La discusión sobre el concepto de necesidad es importante al transitar a los conceptos de consumo y estilo de vida, no para relativizar las necesidades en su totalidad, sino para dar espacio a la influencia del *habitus* y las dinámicas grupales en las decisiones de consumo.

3.2 Sobre el consumo

Autores como Bauman entienden el consumo como una condición permanente e inamovible de la vida, que no está circunscrita a ni una época ni a un período particular de la historia (Bauman, 2011). Sin embargo, a pesar de considerarse una condición de base en la sociedad, en la actualidad “el consumo está lejos, muy lejos de ser una práctica culturalmente neutra” (Araujo y Martucelli, 2012: 59). Si bien siempre ha estado presente, el rol del consumo en la cultura ha sufrido mutaciones a lo largo de la historia y ha ido cobrando progresiva relevancia hasta el día de hoy.

Marinas (2001) identifica tres circuitos de la cultura del consumo: i) *comprar*, es el primer nivel, limitado a los estudios conductistas, donde se reduce el consumo a la compra, suponiendo que el sujeto compra porque tiene una necesidad y el objeto adquirido la satisficiera con su utilidad, entregando el énfasis al satisfactor; ii) *gastar*, que es un nivel más complejo, porque está situado socialmente. En este nivel, las prácticas sociales de consumo como la pérdida, el despilfarro, o el consumo suntuario, no se ofrecen como anomalías o excepciones “irracionales” de los consumidores, sino como procesos grupales duraderos y abundantes. Aquí el objeto aparece como objeto-signo, y confiere una identidad a quien lo adquiere basada en el reconocimiento, suscitando la dinámica del deseo, lo que puede, finalmente, volcar la racionalidad de los sujetos, yendo en contra de la utilidad; finalmente, iii) *consumir*, es el nivel más complejo y engloba a los anteriores. Refiere a la integración como tensión, apuntando a las condiciones de reproducción

social en una sociedad de consumo. Estos circuitos no son simultáneos, sino que se han ido desarrollando a la par con la sociedad de consumo. Es prudente hacer la salvedad de que en esta investigación, si bien se habla de consumo o de gasto de los hogares, lo cierto es que a la luz de la conceptualización de Marinas, las encuestas de ingresos y gastos solo permiten el acceso al circuito de la compra, puesto que no consultan sobre las motivaciones detrás de la *performance* de los hogares en el mercado.

Entre los primeros autores que relevaron la importancia del consumo en la cultura se encuentra Thorstein Veblen, identificando al consumo ostensible o conspicuo de su época (finales del siglo XIX y comienzos del siglo XX) como el sucesor del ocio ostensible, en su función demostrativa, de potencia pecuniaria y estatus social (Veblen, 1963). El ocio en las sociedades cortesanas preindustriales era profundamente valorado como una (no) ocupación digna del tiempo, en una época en que el trabajo productivo era deshonoroso entre los estamentos nobles, a quienes eran reservadas ocupaciones no lucrativas, como la guerra, el sacerdocio, el gobierno y los deportes, que solo podían ser desempeñados por miembros de la aristocracia. El consumo conspicuo, donde la motivación de adquisición está guiada por una orientación demostrativa, destinada a provocar envidia y resentimiento en el otro (Veblen, 1963), aparece como reemplazo de esta misma lógica de ostentación y despilfarro, y Veblen considera que se populariza más el consumo como estrategia clasificatoria porque tiene más potencia simbólica que el ocio.

La relevancia del consumo en la actualidad deriva fundamentalmente de una constatación antropológica (Douglas e Isherwood, 1990), donde los bienes materiales, los objetos, son necesarios no solo para la subsistencia de los seres humanos, sino también para estabilizar y visibilizar categorías que estructuran su convivencia diaria, siendo el consumo una manera de objetivar y compartir significados con el resto de la sociedad. En los objetos prevalece una lógica del don y del estatus, una lógica simbólica y una lógica de la diferencia. Un objeto adquiere un valor simbólico independiente de su relación técnica con la satisfacción de necesidades y, a la vez, en su condición de objeto-signo, entra en un juego de diferencias y distinciones que expresa socialmente el valor atribuido intersubjetivamente a dichos objetos por quienes se relacionan con ellos mediante el consumo (Baudrillard, 1999). Los objetos, en su calidad de signos desempeñan un papel de discriminantes sociales en las sociedades contemporáneas, y los sujetos los utilizan para representarse a sí mismos ante los otros y para afirmar su propia identidad (PNUD, 2002),

convirtiendo al consumo en una práctica de “producción de sentido”. Ante esto, rescatando los vestigios que deja la producción de los discursos simbólicos en torno al consumo, es posible construir taxonomías de la experiencia cotidiana y los estilos de vida de las personas.

Por otra parte, en detrimento de la política y la dimensión del trabajo, para la sociología chilena de principios de este siglo, aparece el consumo como vía de ingreso a la ciudadanía (Araujo y Martuccelli, 2012; Moulian, 2002; PNUD, 2002, Larraín, 2001), y como el nuevo anclaje al vínculo social (PNUD, 2002; Moulian, 2002). Siendo más específico, Larraín (2001) atribuye al legado de la dictadura en Chile el declive de la acción colectiva y el creciente énfasis en el consumo como base de la construcción de identidades y búsqueda de reconocimiento por parte de los demás. Ahora, frente a las sucesivas recesiones económicas, la crisis energética, la conciencia ecológica, entre otros fenómenos, resulta tentador pregonar un nuevo cambio cultural y el fin de la sociedad de consumo. Sin embargo, para Lipovsky (2003) las personas están destinadas a consumir, aunque sea de una manera distinta, y eso es la sociedad posmoderna, no el más allá del consumo sino que su apoteosis.

El consumo ostentoso, tal como fue descrito por Veblen a principios del siglo XX, tenía un significado completamente diferente del que tiene hoy. Aquel era un consumo exhibicionista, de riqueza sólida y durable, que aumentaba su eficacia simbólica mientras más imperecederas resultaban las propiedades exhibidas (Bauman, 2011). A la sociedad que describe Veblen, Bauman la llama sociedad de productores, y era una sociedad que mostraba un espíritu de prudencia y circunspección, que apostaba al largo plazo. En la actual sociedad de consumo, en franca oposición a las formas anteriores de vida, la felicidad está asociada no solo a la simple satisfacción de los deseos, sino a la respuesta al incremento incesante del volumen e intensidad de los deseos, decantando en una suerte de mito de Sísifo, donde se genera un bucle interminable de satisfacción de deseos y necesidades.

En la cultura “inmediatista” de la sociedad de consumo existe un apremio de los individuos por adquirir y acumular, sin embargo, lo que convierte este apremio en urgencia para Bauman (2011), se encuentra en el revés de este movimiento: la necesidad de eliminar y reemplazar. En este sentido, Fernandez-Villaverde y Krueger (2007), analizando el gasto de los hogares en bienes durables y no durables durante el ciclo de vida, a través de la *Consumer Expenditure Survey* de Estados Unidos, obtienen dentro de sus hallazgos principales que el modelo econométrico que

ellos utilizan fracasa en su predicción del gasto en relación a los bienes durables. En lugar de bajar el gasto en este tipo de bienes una vez alcanzado cierto *stock*, y solo cubrirse la depreciación, el gasto en estos sigue creciendo hasta el final del ciclo vital². La compulsión a la adquisición, eliminación y reemplazo de objetos y prácticas de consumo desafía los preceptos de la racionalidad económica, que en este caso supone a los hogares como austeros o prudentes. Ya en 2007 – fecha de la primera edición de “Vida de Consumo” – Bauman señala que las facturas de las empresas especializadas en la comercialización de bienes durables rara vez presentan un cargo por el envío a domicilio, pero sí un abultado adicional por el servicio de eliminación, o retiro del artículo a reemplazar (Bauman, 2011). Sin ir más lejos, en el caso de Chile, al menos en la capital, es muchísimo más dificultoso deshacerse de este tipo de bienes que adquirirlos. Abandonarlos en la vía pública no está permitido y su retiro debe ser gestionado a través de las municipalidades, fundaciones o empresas privadas, con un costo asociado ya sea en dinero o tiempo para gestionar la tarea.

La dimensión del gasto, en sí mismo, no ha perdido en absoluto importancia y potencia explicativa en la actualidad, frente a la aparición de nuevos temas también muy relevantes para las ciencias sociales que están asociados al fenómeno del consumo y a su dinámica compulsiva impuesta por la cultura del consumo. La conceptualización expuesta aporta a la comprensión de que las motivaciones del gasto son diversas, y escapan a las rigideces del acceso material a los bienes, servicios y prácticas, abriendo un gran abanico de discursos simbólicos posibles en torno al consumo en las sociedades.

3.3 Sobre los estilos de vida

Diversas teorías apoyan la idea de que los individuos pueden ser agrupados en conglomerados, basándose en su compromiso con la cultura material (Simmel, 2002) o su posición en un campo relacional (Bourdieu, 1998). Para Max Weber, el fenómeno de la desigualdad o la diferenciación social estaban ligados a la distribución del poder en la sociedad (Aguilar, 2009), y una de estas

² El modelo econométrico utilizado por los investigadores presupone que los hogares realizarán un gasto mayor en bienes durables cuando el ciclo familiar está en ciernes, a causa de la compra de muebles, electrodomésticos, entre otros bienes necesarios para equipar una vivienda por primera vez, y que el gasto en bienes durables de los hogares en ciclos vitales más avanzados debiera ser marginal, debido a que solo requerirán hacerse cargo del desgaste de estos. Distinto es el caso de los bienes no durables como el vestuario y alimentos, que requieren ser repuestos periódicamente, independiente del ciclo de vida de los hogares.

formas de distribución estaba relacionada con el prestigio asociado a determinadas formas de consumo, que configuran el estilo de vida característico de algunas comunidades que él denomina como grupos estamentales, o grupos de estatus, como fue traducido por la sociología norteamericana. Lo que Weber describe es el fenómeno de transformación de los objetos de consumo en símbolos asociados a una posición, que connota un determinado grado de prestigio social: “los grupos estamentales se definen por un estilo de vida común a partir del consumo” (Aguilar, 2009: 133). La interpretación de Weber de la distribución del prestigio social a través del consumo apunta a los mismos elementos expuestos en el apartado anterior, puesto que el autor, cuando habla del consumo, tampoco está pensando en el acto de consumir como parte del circuito económico tradicional, sino como un acto que transforma a los objetos en significantes, en símbolos de una posición en la estructura social.

Por otra parte, un importante aporte a la comprensión de las dinámicas de reproducción social y los estilos de vida fue hecho por Bourdieu, principalmente en *La distinción* (1998), a través de su concepción del *habitus*. El autor plantea que los sujetos son productores de prácticas objetivamente enclasables, pero también lo son de operaciones no menos objetivas de enclasamiento, que, por su parte, también son enclasables. “El *habitus* es a la vez el principio generador de prácticas objetivamente enclasables y el sistema de enclasamiento (*pincipium divisiones*) de esas prácticas. Es en la relación entre las dos capacidades que definen al *hábitus*- la capacidad de producir unas prácticas y unas obras enclasables y la capacidad de diferenciar y apreciar estas prácticas y estos productos (gusto) donde se construye el mundo social representado, esto es, el espacio de los estilos de vida” (Bourdieu, 1998: 171). Bourdieu señala que el *habitus* es un esquema de disposiciones incorporadas, que está estructurado a la vez que es estructurante. El *habitus* es productor de prácticas y obras con sentido, que hace a dichas prácticas sensatas, en base a un sistema de distinciones que emerge desde las mismas disposiciones. Este esquema puede ser aplicado sistemática y universalmente, más allá del contexto en que fue aprendido. La aplicación de un tipo de prácticas y el desempeño de un tipo de distinciones se asemejará a otro, pero también distará del resto, configurando estilos de vida diferentes. El *habitus* es la fórmula generadora que se encuentra en la base del estilo de vida, el que Bourdieu define como un “conjunto unitario de preferencias distintivas que expresan, en la lógica específica de cada uno de los sub-espacios simbólicos- mobiliario, vestidos, lenguaje o *hexis corporal*-, la misma intensidad expresiva” (Bourdieu, 1998: 173).

Los aportes teóricos de Bourdieu han sido ampliamente utilizados como marco conceptual para el desarrollo de investigación empírica. En Chile el PNUD, en su Informe de Desarrollo Humano del 2002, se propuso realizar una radiografía de los chilenos, desde una mirada culturalista. Dentro de los innumerables aportes a la caracterización de los rasgos culturales de los Chilenos, el PNUD desarrolla perfiles de estilos de vida (ellos los llaman “modos de vida”) en base a una serie de características, dentro de las que se encuentra la relación de los individuos con el consumo, que descomponen en 4 tipologías: i) el consumidor necesitado, o “yo consumo para sobrevivir”; ii) el consumidor de bienestar, o “yo consumo para estar mejor”; iii) el consumidor existencial, o “yo consumo para ser más”; y iv) el consumidor modelo, o “yo consumo para gratificarme”. El tipo de datos con los que se realizó esta clasificación difiere del tipo de datos utilizados en esta tesis, puesto que el Informe de Desarrollo Humano contaba con información sobre actitudes de los entrevistados hacia al consumo. La teoría de Bourdieu es un bastión ineludible para cualquier trabajo que aborde el fenómeno del consumo y los estilos de vida en ciencias sociales, sin embargo, ha sido utilizada con mayor profusión en la investigación sobre consumo cultural en Chile, donde la dimensión simbólica de las transacciones resulta más evidente (Gayo, Teitelboim y Mendez, 2009; CNCA, 2017).

En relación al estado del arte sobre la investigación en consumo y estilos de vida, en Chile no se hallaron investigaciones similares a la que se plantea en esta tesis. Como se indicó en el apartado de Diseño de investigación, y se retomará con más detalle en el de Diseño metodológico de este trabajo, lo que se busca es identificar distintos estilos de vida de los hogares chilenos utilizando solo el reporte de sus gastos de consumo, a partir de la EPF, encuesta destinada a conocer los patrones de consumo de los hogares en el país. En ese sentido, sí se hallaron investigaciones de este tipo en otros países, incluso una de ellas sirvió de modelo para el desarrollo de este trabajo: “*Lifestyles through expenditures: a case-based approach to saving*”, realizada por Keister, Benton y Moody, a partir de la encuesta de ingresos y gastos de Estados Unidos (*Consumer Expenditure Survey*). Este estudio, utilizando un enfoque bourdieuano, construye perfiles de estilos de vida de los hogares norteamericanos a partir de la proporción respecto de su gasto total dedicado a diferentes ítems (salud, alimentación, transporte, etc.) y luego, mediante regresión lineal, estudian el efecto de tener cierto estilo de vida en la propensión al ahorro del hogar, concluyendo que el efecto del estilo de vida en el ahorro es significativo. Keister *et al.* (2016), identifican 8 grupos: *Homeowners, Renters, Health-constrained, Drivers, Givers, Pleasure*

seekers, Studens and their parents, y Crisis managers. La idea central, la estrategia metodológica y la definición de los grupos de Keister, Benton y Moody fueron un parámetro importante para el desarrollo de esta tesis y la valoración de sus hallazgos.

Otra investigación similar en su metodología y objetivos es “*Dominant Patterns of Expenditure Among Older People in the United Kingdom: Segmenting the Older Consumer Using the Living Costs and Food Survey*”, de Hayes y Finney (2014), quienes utilizan la encuesta de ingresos y gastos de Reino Unido para realizar una clasificación de hogares compuestos exclusivamente por personas de 60 años o más. El enfoque teórico de este estudio es más pragmático, el concepto de estilo de vida no está detrás de sus análisis, y se centra en el gasto con un objetivo de política pública: visibilizar que los hogares de adultos mayores tienen una estructura de gastos distinta a la estructura promedio de la población, y que, por tanto, se ven especialmente afectados por el alza de precios en ámbitos como la salud, y mucho menos que el resto de la población en otros como la educación, transporte o vestuario, llamando a pensar en políticas públicas específicas para este grupo etario. Si bien los investigadores no están tras los estilos de vida, construyen perfiles similares a los estilos de vida del estudio de Keister *et al.* (2016): *Conservative customers, Foodies, Smokers, Burdened by bills, Socialites, y Recreation and clothing*. En este caso, son 6 los perfiles construidos, y también servirán como referencia en esta tesis para caracterizar e incluso nombrar a los clusters³ de estilos de vida resultantes de la clasificación.

Tras repasar los diferentes abordajes de los conceptos de necesidad, consumo y estilos de vida, y antecedentes empíricos que permiten visualizar de manera más concreta el objetivo de estetrabajo, se revisará en el siguiente apartado la estrategia metodológica desarrollada en el marco de esta investigación.

³ El sustantivo “cluster”, al ser una palabra en inglés, debiese ser escrito con letra cursiva cada vez, sin embargo, se advierte que en esta tesis se utilizará intensivamente el concepto, por lo que, para evitar la sobrecarga gráfica, se decidió escribirlo con fuente normal. Esto aplicará para el singular y el plural, no así para el concepto *clustering*, que no es nombrado frecuentemente.

4. Diseño metodológico

El presente capítulo expone de manera detallada las distintas etapas implementadas para en el análisis y procesamiento de información de la VIII Encuesta de Presupuestos Familiares, poniendo especial énfasis en las pruebas realizadas, los problemas enfrentados, las decisiones tomadas y su justificación.

En primer lugar, se describirán las características de la VIII EPF, que es una encuesta que reviste cierta complejidad por su estructura y conceptos; en el siguiente apartado se expone un panorama general de las metodologías de *clustering* y clasificación, entre las que se encuentra la metodología a utilizar; y finalmente se trazará el plan de análisis utilizado en esta tesis, mostrando el flujo del proceso desde la exploración inicial de los datos hasta la presentación de los resultados obtenidos.

4.1 Descripción general de la VIII Encuesta de Presupuestos Familiares

Este trabajo se desarrolló en base a una fuente secundaria de datos, la VIII EPF a cargo del Instituto Nacional de Estadísticas de Chile. Esta versión de la encuesta fue levantada entre julio de 2016 y junio de 2017, y publicada en junio del año 2018.

El objetivo principal de esta encuesta es identificar la estructura y características del gasto en consumo final⁴ de los hogares urbanos en las capitales regionales de Chile y algunas de sus zonas conurbadas, con un período de referencia de un año; también cuenta con un objetivo secundario, que es identificar la estructura del ingreso total disponible para la misma cobertura geográfica y periodo de referencia.

El principal uso de esta encuesta a nivel de política pública es servir de insumo para la actualización de la canasta de bienes y servicios, que es la base de monitoreo del Índice de Precios al Consumidor (IPC). También se utiliza como fuente de información para: la

⁴ El gasto en consumo final corresponde al gasto realizado por los hogares en bienes y servicios, excluyendo los gastos en activos fijos en forma de viviendas u objetos valiosos (considerados inversión), así como los gastos de consumo intermedio, es decir, el gasto en bienes y servicios adquiridos para el desarrollo de una actividad productiva (Eurostat, 2003).

actualización de las líneas de indigencia y pobreza en la medición de pobreza por ingresos en el país; la estimación de los diferentes componentes del sector institucional de hogares en la compilación del Sistema de Cuentas Nacionales (SCN) del Banco Central; y para el desarrollo de diversos estudios sobre calidad de vida, nutrición y patrones de consumo, entre otros (INE Chile, 2018a).

Otro aspecto a tener en cuenta es que la EPF permite realizar estimaciones estadísticamente significativas con frecuencia anual (sin desagregación mensual), y en términos geográficos para las áreas de estimación de Gran Santiago, Resto de capitales regionales (en su conjunto) y Total de capitales regionales (Gran Santiago y Resto de capitales regionales). Por lo tanto, en términos de desagregación geográfica, los datos son publicados solamente a estos dos niveles, no contando con la variable comuna o región en la base de datos.

4.1.1 Estructura de los datos

La VIII EPF cuenta con 3 bases de datos, cada una con una estructura particular y que funcionan de forma complementaria⁵.

Base de personas de la VIII EPF: es la base principal de la encuesta, contiene la información de los 15.239 hogares entrevistados, distribuida a nivel de personas en 48.308 observaciones. Contiene, a su vez, 247 variables relativas a los siguientes tópicos:

- i) información sociodemográfica;
- ii) variables sobre actividad económica, ingresos y filiación a sistemas de salud y previsión;
- iii) variables agregadas de gasto por hogar y desagregadas de gasto en educación formal;
- iv) y variables sobre la tenencia, caracterización y existencia de gastos relacionados a la vivienda principal y secundaria.

⁵ Las bases de datos, y toda la documentación asociada a la VIII EPF, están disponibles para su descarga en el micrositio de la encuesta: www.ine.cl/epf

Base de Gastos de la VIII EPF: incluye 13 variables y 1.064.239 observaciones, en este caso, gastos. Esta base permite detallar los gastos de los hogares de la manera más desagregada posible. Una fila en la base de gastos corresponde al gasto mensualizado que realizó un hogar en un tipo de producto. En consecuencia, esta base tendrá tantas filas como tipos de productos en los que un hogar declaró haber realizado gasto en el período de referencia, para cada hogar entrevistado.

Base CCIF VIII EPF: esta base contiene los códigos y glosas de la Clasificación del Consumo Individual por Finalidades (CCIF), que corresponde a la adaptación del clasificador de Naciones Unidas (2001) empleado para codificar los bienes y servicios capturados por la encuesta. Este clasificador tiene 5 niveles de desagregación, siendo los 2 niveles más desagregados (subclase y producto) adaptados para dar cuenta de la realidad nacional.

El clasificador de la VIII EPF contiene 12 divisiones, posibles de desagregar en 59 grupos, 126 clases, 285 subclases y 1.186 productos. En suma, esta base cuenta con un total de 1.668 observaciones que caracterizan cada uno de estos niveles. Se adjunta en el Anexo 1 el clasificador a nivel de división y grupo, para ser consultado en caso requerirse. No se dispusieron clase, subclase y producto por su extensión.

4.2 Técnicas utilizadas

4.2.1 Técnicas de *clustering* y clasificación

Dentro de los llamados métodos de *clustering*, existe una amplia paleta de técnicas “no supervisadas” que tienen la finalidad de encontrar patrones o grupos (clusters) entre un conjunto de observaciones. Todos estos métodos tienen la finalidad muy intuitiva de agrupar a las observaciones que son similares entre sí, separándolas de las que son diferentes de acuerdo a atributos definidos. Se dice que estos métodos son “no supervisados” porque el algoritmo no cuenta con una variable de respuesta que indique a qué grupo realmente pertenece cada observación, lo que sí ocurre en métodos “supervisados” como regresión, árboles de decisión, *Support Vector Machine* (SVM), etc. La literatura suele distinguir entre métodos de *clustering* y

métodos de clasificación, basándose justamente en la supervisión o no supervisión del algoritmo (Amat, 2017).

Para el ejercicio desarrollado en esta tesis no se contaba con una variable de respuesta, los estilos de vida no estaban definidos a priori, por lo que se exploraron técnicas de *clustering* para la tarea. Es posible distribuir los métodos de *clustering* en 3 tipos:

- i) *Partitioning clustering*: requieren que el usuario especifique de antemano la cantidad de clusters que desea construir (K-medias, K-medoids, *Clustering Large Applications*[CLARA]).
- ii) *Hierarchical Clustering*: no requieren que el usuario especifique la cantidad de clusters (*agglomerative clustering*, *divisive clustering*).
- iii) Métodos combinados: que mezclan aspectos de los dos anteriores (*hierarchical K-means*, *fuzzy clustering*, *model based clustering* y *density based clustering*)

Los algoritmos más utilizados en *clustering* y para los que se cuenta con mayor documentación y ejemplos prácticos son K-medias y *hierarchical clustering* (Gorenshteyn, 2018; Roark, 2018). Algunas de las ventajas del método jerárquico son que ofrece la posibilidad de escoger si el algoritmo operará de forma aglomerativa (*agglomerative clustering*, en el inicio cada observación es un cluster por sí misma) o divisiva (*divisive clustering*, en el inicio existe un solo cluster que contiene a todas las observaciones), de escoger entre variadas medidas de distancia entre las observaciones, y la posibilidad de tener un control visual de los pasos de la clasificación mediante los dendrogramas⁶. Sin embargo, una desventaja importante del *hierarchical clustering* es que tiene baja escalabilidad (Berzal, 2019), es decir, que se vuelve poco eficiente o presenta dificultades de computo o análisis cuando el volumen de los datos aumenta. Una de estas dificultades es que cuando se cuenta con muchas observaciones, el control visual mediante los dendrogramas deja de ser manejable. Otra dificultad se encuentra en el procesamiento: se trata de un cálculo costoso, por lo que se suele usar sobre muestras reducidas de un set de datos, para escoger un número de clusters(k) y luego ejecutar algún método de *partitioning clustering* más

⁶ Los dendrogramas son visualizaciones típicas del análisis jerárquico que presentan la información de los diversos pasos del algoritmo mediante una estructura de árbol que se ramifica o que converge hacia el tronco, y que además permite observar la distancia (en base al tipo de cálculo de distancia escogido) que debió recorrer para unir o dividir a los *clusters*, ofreciendo una idea de dónde establecer el corte y determinar la cantidad de grupos óptima.

eficiente. Otra desventaja importante es que, debido a la naturaleza de su algoritmo, una vez que *hierarchical clustering* asigna una observación a un cluster no puede reasignarla, por lo que si realiza una mala división o agrupación en los pasos iniciales, esto redundará en los resultados finales del procedimiento.

K-medias es uno de los algoritmos de *partitioning clustering* más utilizados, debido a su sencillez y rapidez de cómputo, lo que además lo hace muy versátil con grandes conjuntos de datos o cuando se utiliza un gran número de variables o atributos. Sin embargo, también presenta algunas limitaciones: i) requiere un valor para K a priori. Esto puede ser complicado en investigaciones exploratorias, y si bien existen estrategias para identificar potenciales óptimos de K , son solo referenciales; ii) Las agrupaciones son inestables, cambian en cada intento; y iii) no es robusto ante la presencia de *outliers*.

Para enfrentar esta última limitación existe otra técnica de *partitioning clustering* que logra sortear mejor la presencia de datos atípicos: K-medoids. La diferencia entre K-medoids y K-medias es que en el primero cada cluster está representado por una observación presente en el cluster, mientras que en K-medias cada cluster está representado por su centroide, que es el promedio de todas las observaciones. Puede entenderse como una relación similar a la que se da entre mediana y media. K-medoids es más robusto que K-medias en cuanto a la presencia de *outliers*, pero es mucho más costoso en términos computacionales.

Se escogió K-medias para la agrupación de estilos de vida por 2 motivos:

- i) porque se contaba con un set de datos voluminoso: 15.080 observaciones (luego de la depuración) y con un set de 15 variables activas, lo que requería algoritmos eficientes para las múltiples exploraciones.
- ii) Y porque los *outliers* fueron analizados y tratados como parte del proceso de preparación de datos, por lo que la sensibilidad de K-medias al ruido y a los valores atípicos dejaba de ser un problema⁷.

Adicionalmente, para solventar la necesidad de un valor de K a priori se utilizó el coeficiente *Average silhouette width*, que será abordado más adelante en este capítulo, y para resolver el

⁷ Para conocer los detalles sobre las técnicas utilizadas para la detección de *outliers* y los resultados de la depuración, ver Anexo 2.

problema de la inestabilidad de las agrupaciones se utilizaron semillas de aleatorización, para poder anclar la ejecución de K-medias con mejor resultado.

4.2.2 *Clustering* a través de K-Medias

El análisis de clusters o *clustering* es un tipo de análisis exploratorio que busca, básicamente, identificar a las observaciones (casos) de un set de datos que, de acuerdo a sus atributos (variables), son similares unas con otras, y al mismo tiempo son distintas de otras observaciones identificadas en otros grupos, generando agrupaciones con sentido que comparten características comunes (Gorenshteyn, 2018).

Uno de los métodos de *clustering* más utilizado es K-medias. Esta metodología, al igual que *hierarchical clustering*, utiliza la distancia euclídea para cuantificar similitud o disimilitud entre observaciones.

El cálculo de la distancia euclídea funciona de la siguiente manera: en un espacio bidimensional, la distancia euclídea entre dos puntos P_1 y P_2 de coordenadas cartesianas (x_1, y_1) y (x_2, y_2) corresponde a:

$$d_e(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

En el caso de K-medias, en el lugar de P_2 está el centroide, y a diferencia del método de clasificación jerárquico, la distancia euclídea se calcula en relación a los K centroides definidos a priori, y no en relación al resto de las observaciones del set de datos.

A diferencia de otros métodos de clasificación, K-medias requiere asignar un valor al argumento K , que le indica al algoritmo cuántos grupos debe construir. El algoritmo de K-medias funciona siguiendo la rutina que se describe a continuación.

El primer paso en el algoritmo es establecer K puntos en posiciones aleatorias en el espacio conformado por los atributos o variables, los que son llamados centroides de los clusters. Luego se calcula la distancia euclídea de cada observación en relación a los centroides definidos aleatoriamente, asignándolas al grupo del centroide más cercano.

Una vez conformados estos clusters iniciales, el siguiente paso es mover el centroide al punto central o media de los clusters resultantes de esta primera asignación. Posteriormente, se recalcula la distancia euclídea de cada observación con los nuevos centroides, reasignando a las observaciones a su centroide más cercano. Este procedimiento se repite iterativamente hasta que los centroides se estabilizan y las observaciones ya no pueden ser reasignadas a otro cluster.

Esto es equivalente a minimizar la suma cuadrática del error intra estrato (INEGI, 2010). Considerando el conjunto de p variables dado por $\{x_1, x_2, \dots, x_p\}$, el método de K-medias busca la conformación de un conjunto de k clusters S , tal que $S = \{S_1, S_2, \dots, S_k\}$, que minimice la suma de cuadrados del error intra grupo.

Formalmente,

$$\min S = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

Donde k es el número de clusters, n es el número de unidades, $x_i^{(j)}$ indica la unidad i y c_j es el centroide del cluster j .

El posicionamiento de los centroides iniciales se escoge de manera aleatoria, y esto redundante en que los resultados de la clasificación no son necesariamente los mismos al repetir el procedimiento. Al partir de un punto diferente cada vez, el proceso iterativo puede finalizar asignando a las observaciones que se encuentran en el límite de 2 clusters a uno o a otro. Para obtener clasificaciones estables que permitan la investigación reproducible, lo que se sugiere en la literatura es plantar una semilla aleatoria, ya que con esto se asegura que al volver a ejecutar el código, el algoritmo tendrá el mismo punto de inicio.

4.2.3 Técnicas para determinar el valor óptimo de K

Existen criterios teóricos y/o empíricos para determinar el valor de K que se le entregará al algoritmo de K-medias para desarrollar la clasificación, y también es posible asumir criterios matemáticos o estadísticos, basados en coeficientes, cuando se trata de una investigación exploratoria donde no se cuenta con antecedentes o parámetros externos para determinar un valor de K . Un ejemplo de un valor de K definido teóricamente es la estratificación socioeconómica

(INEGI, 2010). Muchas veces la cantidad de grupos a generar viene dada y se trata de encontrar la mejor diferenciación entre un grupo “bajo”, uno “medio” y uno “alto”, de acuerdo a las variables utilizadas. Otro ejemplo de este tipo de clasificaciones en que se cuenta con un parámetro es el utilizado por Gorenshteyn (2018), que, con fines didácticos, busca a través de K-medias identificar la pertenencia de un set de 22 observaciones a un equipo de futbol, en base a la información de la posición de cada una de estas 22 observaciones en un plano de coordenadas que representa un campo de futbol previo al inicio de un juego. En este caso, el investigador tiene un parámetro para el valor de K , sabe que hay 2 equipos en cancha.

Para la investigación abordada en esta tesis, se revisaron algunas experiencias similares que ofrecen pistas sobre cuántos grupos se debiera esperar, pero dado su carácter exploratorio, no existe un número de grupos fijo para realizar la clasificación, y K-medias necesita ese argumento para operar.

El coeficiente *Average Silhouette Width* (ASW) ofrece una aproximación matemática al número óptimo de clusters presentes en un conjunto de datos para un conjunto de variables dado, y es ampliamente utilizado en investigación social con este fin (Madero-Cabib y Fasang, 2016; Madero-Cabib, Gauthier y Le Goff, 2016; Degirmenci y Özbakir ,2018; Roark, 2018; Gorenshteyn, 2018). A continuación se profundizará en las características del algoritmo de cálculo del ASW, aportando antecedentes para una correcta valoración de sus resultados.

4.2.3.1 *Average silhouette width*

El *Silhouette Width* es un coeficiente utilizado para determinar el ajuste de la asignación de cada observación a un cluster (Gorenshteyn, 2018). Opera en base a distancias euclídeas, aunque utilizando la función K-medoids, que es similar a la de K-medias. Se articula en 2 pasos: i) calcula el promedio de la distancia de una observación con el resto de las observaciones que fueron asignadas al mismo cluster (C); y ii) calcula el promedio de la distancia entre la misma observación respecto de las observaciones pertenecientes al cluster más cercano, o vecino más cercano (N). Para definir a este vecino más cercano se calcula la distancia de la observación respecto a todos los clusters, y se define como más cercano al que presenta la menor distancia promedio.

Formalmente,

$$s(i) = \begin{cases} 1 - C(i)/N(i), & \text{if } C(i) < N(i) \\ 0 & , \text{ if } C(i) = N(i) \\ N(i)/C(i) - 1, & \text{if } C(i) < N(i) \end{cases}$$

Entonces, si la distancia promedio con las demás observaciones del cluster es menor que la distancia con el vecino, el valor del coeficiente estará entre 0 y 1 (bien clasificado en su cluster), si son iguales será 0 (en el límite exacto entre dos clusters) y si la distancia promedio con el vecino más cercano es menor, el coeficiente será negativo (la observación estaría mejor clasificada en el cluster vecino). El coeficiente *Average Silhouette Width* es el promedio del *Silhouette Width* para cada observación, por lo que puede suceder que el ASW esté aceptablemente sobre cero, indicando una buena clasificación general, y al interior de los clusters haya observaciones con un *Silhouette Width* negativo.

Dado que este trabajo busca identificar diferentes perfiles de gasto entre los hogares, y no se cuenta con un parámetro para argumentar un valor de *K* a priori, este coeficiente se utilizó intensivamente como indicador de ajuste y también la revisión de los grupos conformados a través de estadísticos descriptivos.

4.2.4 Análisis de Componentes Principales

La reducción de dimensionalidad, es decir, resumir la información contenida en un cierto número de variables, tratando de concentrar la mayor cantidad de varianza explicada posible pero en un menor número de nuevas variables (o componentes), es recomendada como un paso previo a implementación de un método de *clustering*, para mejorar sus resultados (Ding y He, 2014; Gorenshteyn, 2018; Roark, 2018; INEGI, 2010). Con este objetivo en este trabajo se utilizará el análisis de componentes principales (ACP), que es probablemente el más popular para estos efectos. Este método tiene fundamentalmente 3 objetivos (Roark, 2018): i) encontrar combinaciones lineales en las variables sobre las que se aplica, creando a partir de ellas un número menor de componentes; ii) mantener la mayor cantidad posible de varianza en los nuevos componentes, concentrándose más varianza explicada en el primero y reduciéndose progresivamente hasta el último; y iii) obtener componentes iniciales que, además de concentrar mayor varianza que las variables originales de forma individual, no estén correlacionados o sean ortogonales unos con otros.

El ACP no es un método de clasificación, su foco principal está puesto en las variables, no en los casos, y es utilizado tradicionalmente para identificar la estructura de relaciones entre un set de variables, y descubrir, asimismo, dimensiones subyacentes para un determinado set de datos.

Formalmente, con un conjunto de variables originales tal como:

$$\{x_1, x_2, \dots, x_p\}$$

Los componentes principales tendrán la forma:

$$y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p$$

$$y_m = a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mp}x_p,$$

Los componentes principales están contruidos de manera que la varianza explicada va decreciendo, tal que $v(y_1) \geq v(y_2) \geq \dots \geq v(y_m)$ y $m \leq p$, donde $\{y_1, y_2, \dots, y_m\}$ corresponde al conjunto de componentes principales.

Los componentes principales dependen de los valores de las variables originales (x) y de los ponderadores lineales (a). Los ponderadores se obtienen del conjunto de datos, aplicando la técnica algebraica llamada descomposición espectral, que se aplica sobre la matriz de varianzas o covarianzas de las variables originales, centradas o no en sus medias (o escaladas) ($x_p - \bar{x}_p$). Los resultados son distintos de acuerdo a si esto último se hace o no. Finalmente, el resultado de esta descomposición espectral son los *eigenvalues* y *eigenvectors*.

Los *eigenvalues* muestran la información de la varianza que es explicada por cada componente, y los *eigenvectors* permiten conocer cómo se relaciona cada variable original con cada uno de los componentes conformados, mostrando de paso también que aunque los componentes se vean representados en mayor medida por la importancia de ciertas variables, o incluso de una, estos resumen información de todas.

Si bien intuitivamente se tenderá a pensar que una reducción de variables originales en un número menor de componentes implicará un sacrificio de información cuando se ejecuta como preparación para una técnica de clasificación, existe investigación que muestra lo contrario (Ding y He, 2004). Utilizando ACP antes de realizar una clasificación no solo disminuye el costo computacional de los cálculos. Ding y He (2004) abordan la relación entre ACP y los resultados de la clasificación mediante K-medias, y muestran cómo también mejora la precisión de una

clasificación reduciendo un conjunto original de 1.000 dimensiones a 60, a 40, a 20, a 10 y a 6 dimensiones o componentes, solo empeorando la precisión al pasar a 5. Por otro lado, para el desarrollo de algunas metodologías de clasificación en grupos es imperativo reducir la dimensionalidad del problema. La metodología de estratificación de Dalenius-Hodges (INEGI, 2010) permite realizar cortes en el continuo de una única variable, por lo que el ACP es fundamental para lograr que dicha variable contenga la mayor cantidad de información posible.

4.3 Plan de análisis

El plan de análisis trazado para esta investigación contempló principalmente el uso de las 3 técnicas revisadas (K-medias, ASW y ACP). La estrategia implementada de reducir dimensionalidad mediante ACP, para luego implementar un K-medias, o alguna otra técnica de *clustering*, está sugerida en investigación aplicada (INEGI, 2010) y también en documentación metodológica (Ding y He, 2014; Gorenshteyn, 2018; Roark, 2018). Sin embargo, el paso a paso tal como fue ejecutado para la realización de esta tesis fue una combinación de elementos de esta literatura y también de otros estudios sobre gastos o estilos de vida a partir del consumo (Hayes y Finney, 2014; Keister *et al.*, 2016), y se puede resumir en los siguientes puntos:

- i) **Construcción de las variables activas en la clasificación.** Esta actividad tuvo por objetivo preparar y tratar las variables activas, desde su forma original en la fuente de datos secundaria, hasta que se encontraron listas para ser ingresadas al modelo de K-medias definitivo, y se compuso de las siguientes tareas:
 - a. Definir un nivel de desagregación del gasto en las variables a utilizar, el que se sustentó a partir de un análisis de correlación y de afinidad conceptual entre distintos tipos de gasto de acuerdo al clasificador utilizado en la EPF.
 - b. Establecer una definición o tipo del gasto a utilizar. Para esta definición se utilizó el coeficiente ASW, pruebas de K-medias para los valores sugeridos de *K* y revisión de la calidad de las clasificaciones mediante estadísticos descriptivos.

- c. Reducción de dimensionalidad. Una vez operacionalizado el gasto (desagregación y definición) se buscó optimizar los resultados de la clasificación, reduciendo la dimensionalidad del set de variables inicial con un análisis de componentes principales (ACP).
- ii) **Construcción, identificación y descripción de los estilos de vida de los hogares chilenos.** Esta actividad se compuso de tres tareas:
- a. Construcción de los clusters definitivos mediante K-medias, a partir de los componentes generados por el ACP que fueron seleccionados.
 - b. Identificación o denominación de los clusters de estilos de vida generados con K-medias, a partir de sus características, principalmente en las variables de gasto o activas en el modelo.
 - c. Descripción de los hogares que ostentan los distintos estilos de vida configurados, principalmente a partir de variables sociodemográficas y socioeconómicas no incluidas en el modelo (variables suplementarias), que fueron construidas para ese fin, y también a partir de gastos que se encuentran en un nivel más desagregado que las variables activas, pero que resultan significativos para cada cluster.

El plan de análisis descrito sintetiza la estructura de exposición de los resultados de esta investigación, donde se describe detalladamente cada proceso.

5. Resultados

5.1 Sobre la construcción de una clasificación

Las clasificaciones no vienen dadas por la naturaleza, son construcciones. Incluso las clasificaciones que pretenden o aparentan objetividad, al desprenderse de atributos “objetivos”, como ciertas características sociodemográficas, el ingreso, o los patrones de consumo, etc., dependen de criterios arbitrarios, como la selección de ciertas variables y la exclusión de otras. Como señalan Joignant y Güell, “las clasificaciones no se «descubren», sino que se «proponen»” (2009: 11), y ante esto es necesario asumir la tarea de elaborar una clasificación, en sí misma, como la construcción de un objeto de estudio.

En la sociología de Durkheim y Bourdieu, el concepto de clasificación adquiere particular relevancia. Ambos autores cuentan en su bibliografía trabajos epistemológicos: Las reglas del método sociológico, en la de Durkheim, y El oficio del sociólogo, en la de Bourdieu, sin embargo, quizás sus aportes más importantes a la comprensión de la construcción de conocimiento en las ciencias sociales, los hacen a través de sus trabajos de investigación: La clasificación primitiva (Durkheim y Mauss, 1969) y La distinción (Bourdieu, 1998). En ambos trabajos se pone el foco en cómo los sujetos administran la realidad a través de sus sistemas de pensamiento, los cuales, al ser construidos socialmente, dan lugar a sistematizaciones, clasificaciones y categorizaciones diferentes de la realidad, de acuerdo a la sociedad o grupo dentro de la sociedad al cual pertenezcan. Desde ambos, a partir del estudio de la forma en que los sujetos clasifican u ordenan la realidad, se logra establecer un puente con la forma en que el investigador clasifica u ordena a su objeto de estudio, a través del tamiz de sus propias preconcepciones. En este sentido, Durkheim y Bourdieu asumen una posición epistemológica internalista, al considerar que no existen esquemas conceptuales independientes del sujeto que observa (Herrera, 2006).

El ejercicio clasificatorio no está restringido a las metodologías cuantitativas o cualitativas, tiene una importancia capital en ambas, y no solo en investigación empírica o aplicada, también en la investigación teórica. Las clasificaciones ordenan la realidad, como se ha mencionado, de una manera no desprovista de arbitrariedad. En ese sentido las clasificaciones basadas en métodos

estadísticos han tomado cierta ventaja en lo que refiere a la ejecución de la clasificación, la agrupación misma de elementos en conjuntos. Una vez escogidas y tratadas las variables sobre las cuales se realizará una clasificación, escogido el algoritmo y los parámetros a utilizar, los criterios y decisiones quedan expuestos en el código y son analizables, criticables e incluso (debieran ser) replicables. Por otra parte, el ejercicio de interpretar, valorar, y sobretodo "nombrar" objetos resultantes de una clasificación, tanto cualitativa como cuantitativa, es un ejercicio subjetivo, donde se ven reflejados intereses, experiencias personales y acervos teóricos. Incluso el haber estudiado cierto *paper* en relación al tema y haber desestimado el estudio de otro, hará que el investigador modifique sus parámetros y la forma en que valora sus resultados.

Bourdieu señala que los científicos sociales tienden a asumir con demasiada facilidad que la relevancia de sus investigaciones viene garantizada por la importancia o centralidad sociopolítica de su objeto (la política, la pobreza, la desigualdad, etc.), desatendiéndose la construcción del objeto de estudio -mediada por el método-, que pasa a segundo plano. Sin embargo, sostiene el autor, lo que en verdad importa es la construcción del objeto de estudio, que tiene la potencialidad de transmutar objetos socialmente insignificantes en objetos científicos: "Sólo en función de una determinada construcción del objeto tal método de muestreo, tal técnica de recolección o análisis de datos, etc., se vuelven imperativos" (Bourdieu y Wacquant, 2008: 279).

Se retorna, entonces, siempre a la importancia de los diseños de investigación en ciencias sociales. Ningún objeto es trivial o central endógenamente, así como su ruta de acceso no es única e inopinable. En ese sentido, uno de los esfuerzos principales de este trabajo es plasmar sistemáticamente el proceso de delimitación del constructo a medir—esto es, de los estilos de vida—, y la secuencia de indagaciones y pruebas desplegadas detrás del diseño metodológico, aspirando con esto a lograr una construcción del objeto prístina para la revisión de las decisiones teóricas y metodológicas puestas a prueba. Los métodos de clasificación no son un recetario, la clasificación es un "arte" que está colmado de decisiones, y lo importante es que estas decisiones sean explícitas, para que sea posible someterlas a escrutinio y evaluación, volviendo, de alguna manera, objetivable el punto de vista del sujeto que observa y que clasifica.

5.2 Construcción de variables activas en la clasificación

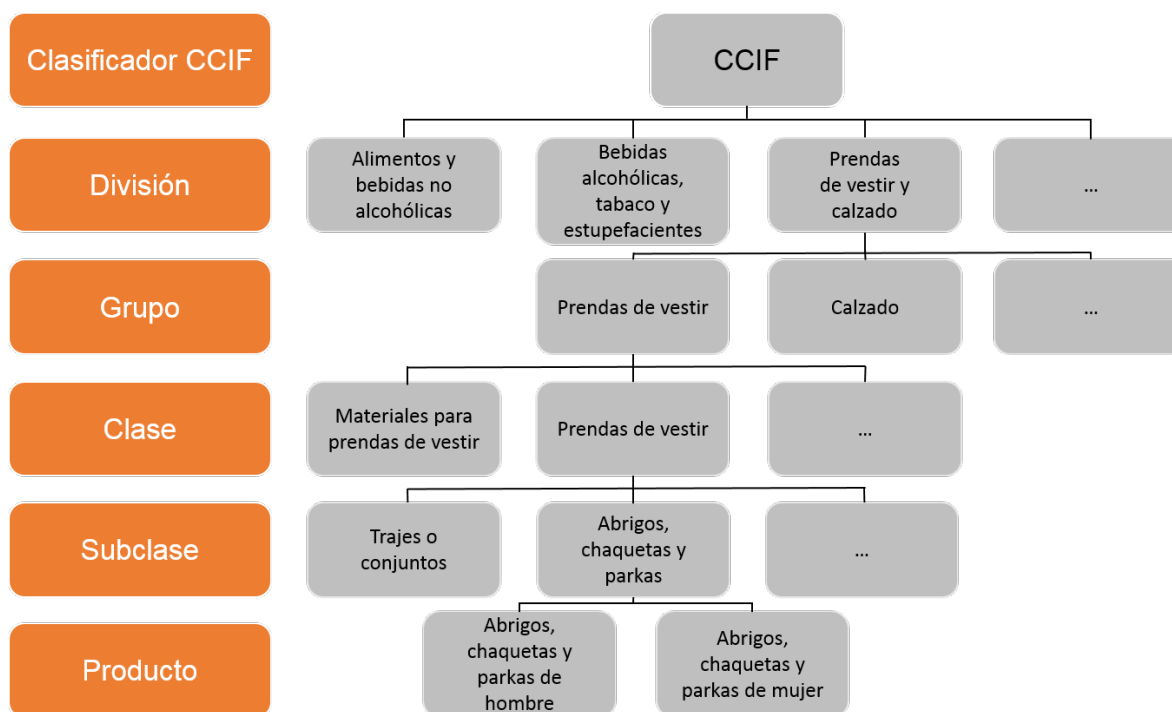
Un aspecto central en la construcción de una clasificación es determinar qué variables serán incluidas en el modelo, es decir, en base a qué atributos o características las observaciones de un set de datos (en este caso hogares) serán similares o distintas unas con otras. Estas variables son llamadas “variables activas”, dado que participan del cálculo de K-medias, y cualquier modificación que se ejecute sobre ellas (agregar o quitar variables, truncarlas, recodificarlas, etc.) generará cambios en el resultado de la clasificación. Por otra parte, se utilizan también “variables suplementarias”, que serán descritas más adelante. Las variables suplementarias, al contrario de las variables activas, no afectan el desarrollo de la clasificación, pero sí son útiles para contrastar sus resultados, dado que permiten conocer características de los clusters obtenidos más allá de los atributos por los que se construyen. Esta nomenclatura (variables activas y suplementarias), si bien es propia del análisis de componentes principales (Lê, Josse y Husson, 2008), es perfectamente aplicable para la investigación a través de K-medias.

Las variables activas en la clasificación son variables de gasto en diversos ítems del consumo de los hogares. Como se indicó en relación a las características de la fuente secundaria de datos, la VIII EPF, existen varios niveles de desagregación del gasto y muchas formas de tratar estas variables para incluirlas en el modelo.

5.2.1 Nivel de desagregación del gasto

La EPF considera 5 niveles de desagregación del gasto: división, grupo, clase, subclase y producto. El nivel de producto es el nivel más desagregado. Cada producto pertenece a una subclase, que a su vez pertenece a una clase, dicha clase a un grupo y este a una de las 12 divisiones existentes en la Clasificación del Consumo Individual por Finalidades (CCIF). En la figura 1 se expone un ejemplo de la interacción entre los niveles jerárquicos del clasificador, utilizando una muestra de códigos reales.

Figura 1: Niveles del Clasificación del Consumo Individual por Finalidades (CCIF)



Fuente: Instituto Nacional de Estadísticas (2018a).

La primera decisión importante a tomar es qué nivel de desagregación del gastoutilizar para clasificar a los hogares. Respecto a este punto, Degirmenci y Özbakir (2018) trabajando con la *Household Budget Survey* de Turquía, utilizan el nivel de producto para producir clusters de hogares. Esto significaría, en el caso de la EPF, calcular similitud entre hogares en función de 1.186 atributos diferentes, lo que generaría no solo un costo computacional elevado, sino también una dificultad a la hora de interpretar, mediante estadísticos descriptivos, cuáles son las características que le dan identidad a cada cluster.

Se consideró que el nivel de subclase (285 variables) e incluso de clase (126), seguían siendo difíciles de manejar en términos estadísticos, pero por sobre todo conceptuales. Las primeras pruebas, por lo tanto, se realizaron utilizando el gasto de los hogares para cada grupo (59 variables), sin embargo, no tuvieron buenos resultados en términos estadísticos. Al analizar el éxito de la clasificación mediante el coeficiente *Average Silhouette Width* no se encontró un buen ajuste para ningún valor de K , por lo que se decidió seguir explorando en el nivel más agregado, el de división, tal como lo hacen los estudios de Hayes y Finney (2014) y el de Keister *et al.* (2016), y estudiar la desagregación a nivel de grupo de algunas de estas divisiones.

5.2.1.1 *Análisis de afinidad y de correlación de los grupos al interior de las divisiones de gasto*

Además, siguiendo la metodología de Keister *et al.* (2016), se realizó un análisis sobre los grupos al interior de cada división. El primer análisis fue conceptual, con el objetivo de relevar ciertos gastos que podrían ser indicativos de un estilo de vida en particular y que no se veían reflejados a nivel de división, pero sí a nivel de grupo. Para este análisis además se atendió a las decisiones de desagregación tomadas en los estudios mencionados. El segundo análisis fue estadístico, y permitió optimizar los resultados de la clasificación, identificando divisiones en las que los grupos que las componían estaban poco correlacionados, o correlacionados inversamente, sugiriendo que la mejor opción era ingresar al modelo no la división completa sino que sus grupos de forma independiente. Como resultado de esta revisión se tomaron las siguientes decisiones:

- i. De la división 2, “Bebidas alcohólicas, tabaco y estupefacientes”, que cuenta con 4 grupos, se ingresó de forma independiente el grupo 1, “Bebidas alcohólicas”, y el grupo 2, “Tabaco”. Se desechó el grupo 3, “Estupefacientes”, y el grupo 4, que es un residual de la división⁸, a raíz de su escaso aporte al gasto de los hogares. Es necesario considerar que de acuerdo a la experiencia en medición de gastos a nivel nacional e internacional, la división 2 es la que presenta la mayor subdeclaración (en relación al parámetro del sistema de cuentas nacionales, que registra las ventas), de manera que es de los pocos gastos que se ponderan por un factor de multiplicación al momento de la construcción de la canasta de bienes del IPC.

Ingresar los gastos del grupo 1 y grupo 2 por separado se sustenta en que entre ellos hay una correlación baja (0.16), y en que en el estudio de Keister *et al.* (2016), los gastos en alcohol y en tabaco se comportaron de manera independiente. En dicho trabajo se halló una clara relación entre el gasto elevado en salud y el elevado gasto en productos relacionados con el tabaco, en el cluster que denominaron *Health-Constrained*.

⁸Cada división de gasto cuenta con un grupo que funciona como un residual. Este grupo contiene gastos “no desglosados”, es decir, que se sabe que pertenecen a la división, pero que no fue posible clasificarlos en algún grupo en específico, porque incluían gastos de más de un grupo en un mismo registro, asociados a un gasto imposible de separar.

- ii. La división 7, que cuenta con 4 grupos, se dividió en 2 tipos de gastos. El primero contiene la fusión de los gastos del grupo 1, “Adquisición de vehículos”, y el grupo 2, “Funcionamiento de equipo de transporte”. La correlación entre estos dos no es muy alta (0,11), pero conceptualmente están relacionados; quienes han adquirido o ya poseen un vehículo realizarán mayor gasto en gasolina, peajes, estacionamientos, mantención, etc. Por otro lado, la correlación entre estos dos grupos y el grupo 3, “Servicios de transporte” (básicamente el transporte público), es inversa; entre el gasto en el grupo 3 y el grupo 2, por ejemplo, la correlación es de -0.23. Esto es de sentido común, puesto que en términos de transporte, para muchas personas el vehículo individual y el transporte público operan como bienes sustitutos, son medios distintos para un mismo fin.

El grupo 4, que es el residual de la división, fue desechado por no aportar información útil para la interpretación, dado que contiene mezclados gastos relativos a grupos que se propone incluir en el modelo de forma independiente.

- iii. El último ajuste fue realizado al interior de la división 12. Como indica su nombre, “Bienes y servicios diversos”, este grupo contiene categorías de gasto muy amplias y de difícil interpretación a nivel agregado, desde servicios de Prostitución, pasando por Impresiones y fotocopias, hasta Servicios funerarios.

Se decidió agrupar los gastos del grupo 1, “Cuidado personal”, con los del grupo 3, “Efectos personales”, por un lado, y por otro, los gastos contenidos en el grupo 5, “Seguros”, con los del grupo 6, “Servicios financieros”. Los gastos de los grupos 1 y 3 no estaban correlacionados (0.02), pero sí tienen una relación conceptual, mientras que los de los grupos 5 y 6, además de su vinculación conceptual, presentaban una correlación considerable de 0.26. La primera fusión está además apoyada por la experiencia del estudio de Keister *et al.* (2016), que utiliza una categoría única de cuidado personal.

Se revisó también la posibilidad de fusionar divisiones de gasto. En particular, la fusión de la división 4, “Alojamiento, agua, electricidad, gas y otros combustibles”, con la división 5, “Muebles, artículos para el hogar y para la conservación ordinaria del hogar”, por ser ambos gastos relacionados con la vivienda. Sin embargo, su correlación es inversa (-0,15), lo que se explica porque disputan su importancia relativa respecto al mismo gasto total. La otra fusión evaluada fue la de la división 9, “Recreación y cultura”, con la división 11, “Restaurantes y

hoteles”, bajo el supuesto de que ambas contienen gastos primordialmente suntuarios. Se encontró una alta correlación al analizar no el gasto proporcional sino que el total en ambas divisiones (0,48), y en otros trabajos siempre el más alto gasto en estas dos divisiones está asociado a un mismo cluster: al grupo llamado *Pleasure seekers* en Keister et al. (2016), y los *Socialites*, en Hayes y Finney (2014).

A pesar de estos antecedentes, se decidió no realizar estas agregaciones, para mantener un set de variables más comparable con los estudios revisados. Por otro lado, estas agregaciones no ofrecían un mejor ajuste en términos de distancias intra y entre grupos para ningún valor de K, lo que fue evaluado mediante el coeficiente ASW.

La tabla 1 resume las decisiones descritas y presenta el set de variables definitivo sobre el que se desarrolló la clasificación⁹.

Tabla 1: Set de variables inicial

Gasto	Nombre variable	Glosa teórica
División 1	D01	Alimentos y bebidas no alcohólicas
División 2 – grupo 1	G02_1	Bebidas alcohólicas
División 2 – grupo 2	G02_2	Tabaco
División 3	D03	Prendas de vestir y calzado
División 4	D04	Alojamiento, agua, electricidad, gas y otros combustibles
División 5	D05	Muebles, artículos para el hogar y para la conservación ordinaria del hogar
División 6	D06	Salud
División 7 – grupo 1 y 2	G07_1_2	Compra y mantención de vehículos
División 7 – grupo 3	G07_3	Transporte público
División 8	D08	Comunicaciones
División 9	D09	Recreación y cultura
División 10	D10	Educación
División 11	D11	Restaurantes y hoteles
División 12 – grupo 1 y 3	G12_1_3	Cuidado personal
División 12 – grupo 5 y 6	G12_5_6	Servicios financieros y seguros

Fuente: Elaboración propia en base a VIII EPF-INE.

⁹ Se recomienda la revisión del Anexo 1 para obtener una visión más general de cada una de las divisiones y grupos que componen la CCIF.

5.2.2 Definición de gasto para la clasificación

Además del nivel de desagregación del gasto a utilizar para el modelo de clasificación, es necesario decidir cómo operacionalizar este gasto. No es trivial para los modelos de clasificación basados en distancias, como K-medias, que una variable sea incluida como continua, ordinal o dicotómica, o que esta sea convertida en una proporción o razón. Si se decidiera, por ejemplo, dicotomizar los gastos de los hogares, convirtiendo en “1” los que son iguales o superiores a la media y en “0” los que están bajo la media, la probabilidad de encontrar hogares que tengan exactamente los mismos atributos en las 15 variables seleccionadas para el modelo (y tengan, por tanto, una distancia igual a 0) es más elevada que si en vez de usar dicotomías se usaran tramos de gasto. Por supuesto, es más improbable encontrar hogares idénticos cuando las variables incluidas en el modelo son continuas, como sería si se usara el monto del gasto mensual, que puede ir desde 0 hasta varios millones de pesos.

En la bibliografía consultada donde se buscaba alcanzar objetivos similares a los de este trabajo, se realizaba una transformación del gasto a proporciones respecto del gasto total (Keister *et al.*, 2016). Además se expresaba la importancia de que las variables estuvieran en una misma unidad de medida y, de no estarlo, de escalarlas o normalizarlas, para que las variables que tengan un rango más amplio no monopolicen la varianza y, en consecuencia, la clasificación. Sin embargo, no se aportaron antecedentes que permitieran descartar el uso de las variables gasto en el formato que las personas que contestan la EPF las reportan: como montos. Es por esto que se decidió probar 2 definiciones de las 15 variables seleccionadas: i) como montos (en pesos), y ii) cada una de las variables como un porcentaje (entre 0% y 100%) respecto del gasto total reportado por el hogar.

Ahora, las variables de monto no deben ser ingresadas en su forma original, puesto que la magnitud de los montos reportados se ve afectada por la cantidad de miembros del hogar. Por este motivo se realizó un tratamiento a las variables para aislar el efecto del tamaño del hogar: se transformó el gasto en “gasto equivalente”. El concepto de gasto equivalente es una adaptación del de ingreso equivalente utilizado en la medición de pobreza por ingresos desde la versión 2013 de la Encuesta de Caracterización Socioeconómica Nacional (Casen), y que mediante un cálculo muy similar utilizan Hayes y Finney(2014) para ajustar los gastos en su estudio sobre los estilos

de vida a través del consumo entre hogares de adultos mayores en Reino Unido. El Ministerio de Desarrollo Social utiliza una escala de equivalencia para ajustar los ingresos de los hogares, para así incluir el efecto del tamaño del hogar en el cálculo del ingreso per cápita, considerando que existen economías de escala en el consumo al interior de los hogares (CMP, 2014). Esto redundaría en que cada miembro adicional requeriría menos que un incremento proporcional del ingreso para obtener el mismo nivel de bienestar.

Entonces, para lograr el gasto equivalente se siguió una metodología similar a la utilizada por el Ministerio de Desarrollo Social:

$$g_{eq} = \frac{g_t}{n} \times 0.7$$

Donde g_{eq} es el gasto equivalente de un hogar en algún ítem, g_t es el gasto total de ese hogar en ese ítem y n es el número de personas en ese hogar. El ponderador único de 0.7 es el propuesto por la Comisión para la Medición de la Pobreza para reflejar las economías de escala al interior de los hogares (CMP, 2014). Se aplicó este tratamiento a 14 de las 15 variables, por considerarse que el gasto en la división 5, “Muebles, artículos para el hogar y para la conservación del hogar”, se veía menos afectado por las economías de escala que las demás.

5.2.2.1 Análisis y depuración de outliers o valores atípicos

Por otra parte, atendiendo a los pasos sugeridos por la bibliografía de trabajos empíricos, así como también textos y materiales metodológicos de *clustering* revisados (Hayes y Finney, 2014; Everitt *et al.*, 2011), se realizó una detección y depuración de *Outliers* o casos atípicos.

Las clasificaciones basadas en distancias son altamente sensibles a los casos atípicos, que estiran agresivamente los rangos de las variables y empeoran el rendimiento de los algoritmos de clasificación, que, o recorren gran distancia para vincular una observación a algún grupo (incrementando la varianza al interior de los clusters), o generan clusters de una sola observación (Hayes y Finney, 2014), ambas alternativas poco auspiciosas para la evaluación posterior de la clasificación.

La estrategia de análisis y tratamiento de *outliers* implementada en este trabajo tuvo dos aristas. La primera fue un tratamiento de los gastos atípicos empleando una mezcla de dos métodos

estadísticos:El método de Hadi y el método del Rango intercuartil. Para todas las variables de gasto, salvo la división 1 (“Alimentos y bebidas no alcohólicas”), se implementó el método de Hadi, mientras que para tratar la división 1 se utilizó la metodología de Rango intercuartílico. La explicación detallada de estos dos métodos, así como el detalle de sus hallazgos, se dispuso en el Anexo 2.

Las técnicas de análisis de *outliers* en general tienen problemas para detectar casos atípicos en la cola inferior de la distribución (Hadi, 1992), por lo que la segunda arista del tratamiento de estos casos tuvo un enfoque normativo, para evitar la presencia de hogares con gastos muy bajos que distorsionaran la clasificación: se eliminaron de la muestra los hogares que presentaron un gasto total per cápita menor a la canasta básica alimentaria deflactada al mes inicial de la encuesta (julio de 2016), avaluada en \$40.118¹⁰ (MDS, 2018b), por considerarse un indicio de subdeclaración de gastos.

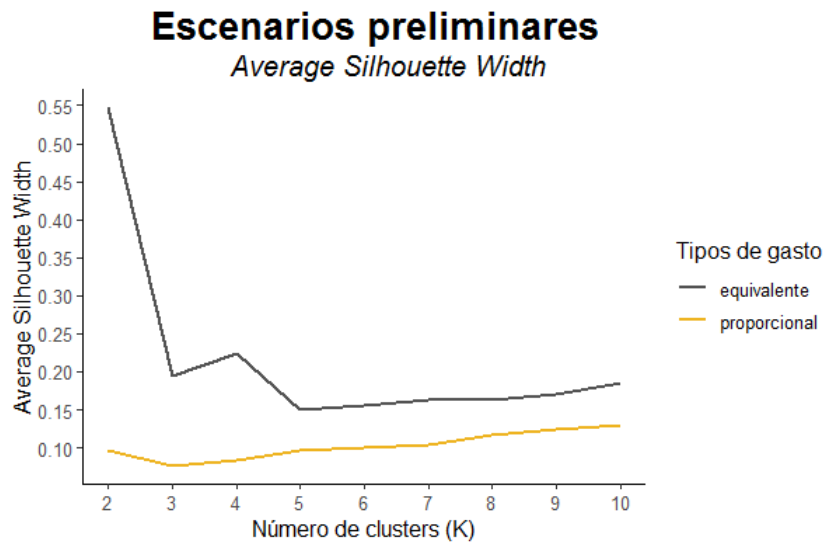
Como consecuencia de estos procedimientos, se truncaron en promedio 8,4 observaciones (hogares) por cada variable de gasto, por tener un valor muy elevado, y se eliminaron 159 hogares que declararon un gasto total bajo el parámetro establecido.

5.2.2.2 *Pruebas para la definición del tipo de gasto para la clasificación*

La primera aproximación a los datos se hizo mediante el coeficiente ASW, luego, para profundizar, se realizaron pruebas de clasificación con K-medias con distinta cantidad de grupos (valor de K), de acuerdo a lo sugerido por el ASW, para finalmente revisar estadísticos descriptivos y evaluar la calidad de las agrupaciones. Estas pruebas se realizaron utilizando como variables activas el set de 15 variables expuesto en la tabla 1.

¹⁰ Esto significaría estar incluso muy por debajo de la línea de pobreza extrema de acuerdo a la medición actual de pobreza por ingresos.

Gráfico 1: Pruebas para definición de tipo de variable con ASW. Escenarios preliminares



El vector representado en el gráfico 1 con el color gris oscuro muestra los coeficientes ASW asociados a las distintas soluciones de K para las variables de tipo gasto equivalente, y la línea amarilla representa los coeficientes ASW asociados a las distintas soluciones de K para las variables de tipo gasto proporcional. La utilidad del ASW para analizar estos escenarios preliminares se encuentra en que entrega en un formato sencillo un panorama sobre el ajuste matemático de los grupos simulando iterativamente que se ejecutara un K -medias para 2, 3, 4 o 5 grupos.

Se observa en el gráfico 1 que al utilizar las variables de gasto equivalente se alcanza un coeficiente cercano a 0,55 (de un rango de entre -1 y 1 del ASW) para la solución de 2 clusters, que sería un ajuste más que aceptable al confrontarlo con otras investigaciones consultadas (Degirmenci y Özbakir, 2018; Madero-Cabib y Fasang, 2016; Madero-Cabib, Gauthier y Le Goff, 2016). Sin embargo, si se tiene a la vista el objetivo de identificar diferentes estilos de vida entre los hogares chilenos, la solución de dos grupos es poco informativa, puesto que lo que está haciendo es separar a los que gastan más dinero de los que gastan menos dinero en general, lo que, dada la alta correlación entre ingreso y gasto, es básicamente separar entre ricos y pobres. Por otra parte, se observa que luego de la solución de dos grupos los coeficientes caen drásticamente.

Sin embargo, a pesar de la caída del ASW después de la solución de 2 clusters, los coeficientes del set de variables de gasto equivalente se mantienen constantemente sobre el de las variables de

gasto proporcional. Al ejecutar un K-medias con un valor de $K = 4$ (4 clusters) para el set de variables de gasto equivalente, que tiene un ASW de 0.22, se encontró que los grupos conformados tenían tamaños excesivamente diferentes, con cerca del 90% de los hogares (los de menor gasto) agrupados en un mismo cluster y el cluster de menor tamaño contenía solo 1 hogar, situación que se repetía de forma similar para otras soluciones de K superiores. Por otra parte, al estudiar las medias de gasto de los clusters en las variables activas, las diferencias se encontraron en la magnitud del gasto y no en la estructura. Es decir, algunos clusters gastan más y otros gastan menos en todo, mientras que lo que se espera es encontrar grupos que sean configurados por sus preferencias y gustos, y no solamente por su capacidad de gasto o acceso. En conclusión, el set de variables de gasto equivalente, si bien presenta mejores coeficientes, sufre un serio desequilibrio entre los tamaños de los clusters resultantes y no ofrece información sustantivamente distinta a la de la estratificación socioeconómica.

Por otro lado, se observa que el vector que representa al set de variables de gasto proporcional tiene un comportamiento más conservador y estable en torno al coeficiente ASW. Al ejecutar K-medias para distinta cantidad de grupos (7, 8, 9, 10), se encontró que el tamaño de los grupos conformados era más equilibrado y, lo más importante, los grupos efectivamente mostraban estadísticos descriptivos que permitían identificar estilos de vida: un grupo que se congregaba en torno a su alto gasto proporcional en alimentación, otro en torno a su alto gasto en arriendo y cuentas en general, etc. Dado lo anterior, y considerando que fue la definición del gasto utilizada en el estudio de Keister *et al.* (2016) que se tomó como referencia principal, se determinó seleccionar el set de variables de gasto proporcional, anteponiendo como criterio la calidad sustantiva de la clasificación lograda, frente a la recomendación matemática del ASW. Sin embargo, se hacía necesario implementar una estrategia para mejorar el rendimiento de la clasificación. Para esto la literatura sugiere utilizar el Análisis de Componentes Principales (Ding y He, 2014; Gorenshteyn, 2018; Roark, 2018).

5.2.3 Reducción de dimensionalidad con ACP

Uno de los hallazgos empíricos obtenidos de las pruebas realizadas con ASW para el set de variables de gasto proporcional, es que el valor óptimo de K sugerido va en directa relación con la cantidad de variables activas incluidas en el modelo. En términos simples, al utilizar 12 variables originales de gasto para el ASW, la cantidad de clusters sugeridos es 12, con 15

variables originales la cantidad de clusters sugeridos es 15. Al ejecutar K-medias con el valor de K sugerido y revisar las medias de gasto de los clusters en las variables ingresadas al modelo, los clusters tienden a agrupar, cada uno, a los hogares que gastan más proporcionalmente en cada una de las variables ingresadas. Esta clasificación no es deseable, pues no representa la complejidad de los estilos de vida de los hogares, lo que hizo imperativo explorar alguna manera de optimizar o recomponer la información contenida en las variables originales, para dar lugar a configuraciones más complejas de los clusters. Esto se logró mediante la implementación del análisis de componentes principales.

El ACP fue aplicado sobre las 15 variables de gasto por hogar resultantes de las definiciones previas, resumidas en la tabla 1. El algoritmo que se utilizó operó sobre una matriz de variables centradas en sus medias.

A continuación se presentan los *eigenvalues* resultantes del ACP:

Tabla 2: <i>Eigenvalues</i> resultado del Análisis de Componentes Principales				
Componente	<i>Eigenvalue</i>	Diferencia	Proporción	Acumulado
Comp1	1,8554	0,5215	0,1237	0,1237
Comp2	1,3339	0,0532	0,0889	0,2126
Comp3	1,2807	0,0936	0,0854	0,2980
Comp4	1,1871	0,0908	0,0791	0,3771
Comp5	1,0963	0,0559	0,0731	0,4502
Comp6	1,0405	0,0221	0,0694	0,5196
Comp7	1,0184	0,0455	0,0679	0,5875
Comp8	0,9728	0,0385	0,0649	0,6523
Comp9	0,9344	0,0040	0,0623	0,7146
Comp10	0,9303	0,0423	0,0620	0,7767
Comp11	0,8880	0,0354	0,0592	0,8359
Comp12	0,8527	0,0252	0,0568	0,8927
Comp13	0,8275	0,0563	0,0552	0,9479
Comp14	0,7712	0,7604	0,0514	0,9993
Comp15	0,0108		0,0007	1,0000

Fuente: Elaboración propia en base a VIII EPF-INE.

Como es posible observar en la tabla 2, la proporción de varianza explicada por los primeros componentes no es muy alta, y no es tan drástica la caída de la varianza explicada hacia los últimos componentes, como es característico (INEGI, 2010; Roark, 2018).

Tabla 3: *Eigenvectors* resultado del Análisis de Componentes Principales

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9	Comp10	Comp11	Comp12	Comp13	Comp14	Comp15
D01	-0,527	0,179	-0,310	-0,099	0,106	-0,023	-0,259	-0,109	-0,203	-0,197	0,052	-0,041	-0,275	-0,305	0,491
G02_1	-0,031	0,325	0,026	-0,471	-0,021	0,141	0,167	0,411	0,164	-0,062	0,444	-0,119	-0,225	0,395	0,064
G02_2	-0,175	0,275	0,150	-0,376	0,148	0,052	0,108	0,147	-0,010	0,605	-0,279	0,115	0,361	-0,272	0,102
D03	0,066	0,249	0,158	0,449	-0,201	0,228	-0,199	-0,095	0,175	0,482	0,223	-0,428	-0,213	-0,101	0,122
D04	-0,269	-0,543	0,308	-0,037	-0,455	0,134	0,140	0,052	0,202	0,056	0,035	0,197	0,015	0,057	0,448
D05	0,180	-0,008	-0,424	0,119	-0,287	0,235	0,046	0,491	-0,171	-0,077	-0,450	-0,297	0,135	0,080	0,206
D06	0,151	-0,246	-0,361	0,100	0,370	-0,183	0,616	-0,117	0,079	0,242	0,178	-0,157	-0,068	-0,027	0,296
G07_1_2	0,411	-0,107	-0,039	-0,362	0,124	0,155	-0,402	-0,426	-0,020	0,054	-0,029	-0,119	0,201	0,323	0,377
G07_3	-0,290	0,282	0,270	0,287	0,174	-0,139	0,183	-0,151	-0,061	-0,037	-0,399	-0,031	-0,045	0,611	0,186
D08	-0,017	-0,033	0,258	0,294	0,483	0,484	0,005	0,219	-0,089	-0,278	0,248	0,023	0,382	-0,104	0,161
D09	0,250	0,281	0,066	0,109	-0,350	-0,128	0,197	-0,087	-0,636	0,040	0,313	0,304	0,135	0,003	0,205
D10	0,227	-0,090	0,234	0,103	0,184	-0,630	-0,349	0,485	0,042	0,043	0,029	0,003	-0,047	-0,038	0,281
D11	0,241	0,322	0,261	-0,113	-0,167	-0,134	0,281	-0,182	0,346	-0,453	-0,156	-0,236	0,107	-0,371	0,209
G12_1_3	0,000	0,283	-0,417	0,253	-0,088	-0,042	-0,128	-0,003	0,534	0,004	0,106	0,496	0,290	0,120	0,122
G12_5_6	0,366	0,071	0,090	0,014	0,169	0,327	0,050	0,058	0,034	0,034	-0,286	0,475	-0,609	-0,121	0,126

Fuente: Elaboración propia en base a VIII EPF-INE.

En la tabla 3 se puede observar el aporte de cada variable original a cada componente. De esta tabla se desprende que las variables que tienen un mayor aporte al componente 1 son la división 1 y la división 7 grupo 1 y 2. Las variables que más aportan al componente 2 son la división 4, la división 2 grupo 1 y la división 11; observando los *eigenvectors* de esta manera es posible conocer también el aporte de las variables originales en el resto de los componentes. No se profundizará en este apartado en la descripción de los componentes, puesto que el objetivo de la implementación del ACP en esta investigación es la optimización mediante la reducción de dimensionalidad.

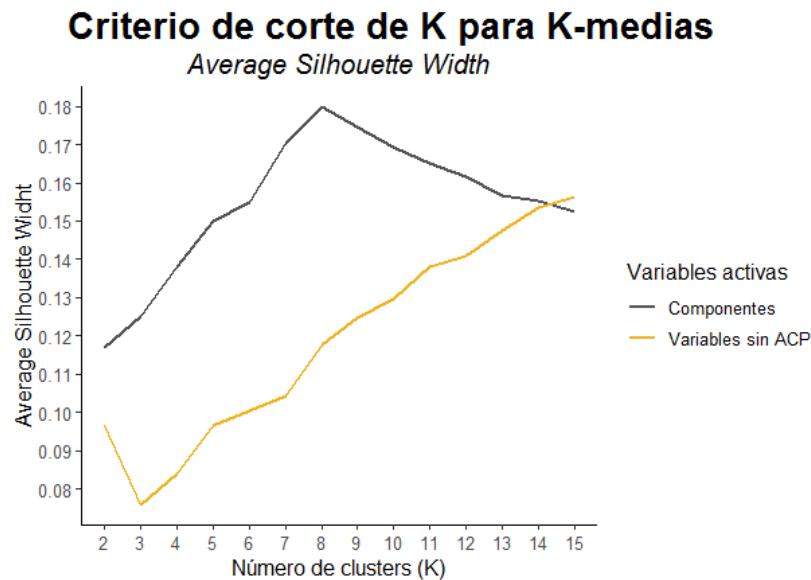
A partir de un proceso iterativo, agregando componentes uno a uno y calculando el coeficiente *Average Silhouette Width*, se decidió ingresar 7 componentes para el K-medias, dado que con este conjunto se lograba una solución óptima de *K* razonable para distinguir estilos de vida que tuvieran cierta complejidad, y porque la solución resultaba matemáticamente aceptable. Utilizar 7 componentes implicó considerar un 58,8% de la varianza explicada por el set de variables originales (atendiendo a la última columna de la tabla 2). Por otra parte, mediante la visualización de los componentes con *biplot*¹¹, fue posible asegurar que todas las variables originales tuvieran

¹¹ El biplot es un tipo de gráfico bidimensional propio del ACP, que permite representar el peso y dirección de las correlaciones entre cada una de las variables originales con cada par de componentes principales.

participación en los 7 componentes iniciales, lo que además se ve reflejado en los resultados de la clasificación definitiva con K-medias.

En el gráfico 2 se expone el comportamiento del coeficiente ASW para distintos valores de K , comparando el desempeño del set de 15 variables de gasto originales ya presentado en el gráfico 1 (vector de color amarillo) con el de los 7 primeros componentes resultantes del ACP (vector de color gris)¹².

Gráfico 2: Comparación de ASW entre variables originales y componentes principales.



Como se puede apreciar, el ASW para el set de variables con ACP presenta mejores coeficientes para todas las soluciones de K hasta el valor 14, donde se cruzan ambos vectores. El set de variables originales logra su mejor coeficiente cuando se calcula para 15 grupos (0,156), lo que coincide con la cantidad de variables. Esto ya se indicó previamente, y al probar con 12 variables (sin desagregar divisiones) el mejor coeficiente se logró al clasificar en 12 grupos.

El valor de K que obtiene el coeficiente más alto para el set de variables con ACP es 8, con un ASW de 0,18. Ante el valor de este coeficiente en otras investigaciones, como las de Madero-Cabib y Fasang, que obtienen un ASW de 0,45, o Madero-Cabib, Gauthier y Le Goff (2016) que obtienen un ASW de 0,27, pareciera ser un resultado deficiente. Sin embargo, el coeficiente

¹² Se hicieron modificaciones a los ejes entre el gráfico 1 y gráfico 2 para adaptarse a las características de los vectores y el contenido de la exposición, pero el vector amarillo representa en ambos casos al vector de coeficientes ASW de las variables de gasto proporcional sin ACP.

obtenido para 8 grupos no resulta insatisfactorio al compararlo con el obtenido en otras investigaciones que utilizaron datos similares (muchas variables y continuas), como el del estudio de Degirmenci y Özbakir (2018), que presenta un mejor coeficiente ASW de 0,216.

Por otra parte, la solución de 8 grupos resultó satisfactoria al evaluar los estadísticos descriptivos del gasto en las variables originales por grupo, conformándose clusters con sentido y con ciertos rasgos parecidos a los grupos encontrados en investigaciones similares.

Como intuición desarrollada a partir del proceso de exploración, es posible concluir, además, que el coeficiente ASW es sensible a la cantidad de variables que se utilizan para la clasificación. En el proceso de probar el comportamiento del coeficiente, ingresando ya sea variables originales o componentes de uno en uno, los coeficientes para cada punto de corte bajaban cada vez que se agregaba una variable al set, resultando mejores coeficientes para el cálculo con un solo componente (que contiene solo el 12,4% de toda la varianza). Ante estos antecedentes se decidió generar la clasificación vía K-medias con un K igual a 8, y un ASW de 0,18.

5.3 Construcción, identificación y descripción de los estilos de vida de los hogares chilenos

Como fue descrito en los apartados anteriores del capítulo de resultados, mediante la evaluación de diferentes alternativas de construcción de variables activas para la clasificación, y su evaluación definitiva a través del ASW, fue posible llegar a un óptimo de K de 8 grupos para el conjunto de datos utilizado.

A continuación se presenta, en primer lugar, las variables suplementarias que se utilizarán para la caracterización de los grupos, y luego se hará una descripción general de los clusters en base a algunas de estas variables. En tercer lugar, se analizará el comportamiento de los clusters en relación a las variables activas, lo que permitirá “bautizar” a estos grupos con un nombre representativo de sus estilos de vida, y finalmente se expondrá un relato acerca de cómo son los estilos de vida de los hogares chilenos, en función de sus prácticas de consumo, y quienes los ostentan, en términos socioeconómicos y sociodemográficos.

5.3.1 Variables suplementarias para la caracterización de los clusters

El concepto de variable suplementaria, al igual que el de variable activa, fue tomado de la nomenclatura del análisis de componentes principales (Lê, Josse y Husson, 2008), por ser una forma intuitiva de entender el rol que tienen en el marco de una aplicación de *clustering*. Las variables suplementarias son variables que no participan como argumentos de un K-medias, y que se construyen para contrastar sus resultados, permitiendo caracterizar a los grupos obtenidos.

En este apartado se describirán las variables suplementarias utilizadas en esta investigación, justificando su relevancia para el estudio de los estilos de vida, ya sea por lo sugerido en la literatura o en base a los hallazgos particulares de este trabajo. Para su exposición fueron clasificadas por su carácter de variables sociodemográficas o socioeconómicas.

5.3.1.1 Variables sociodemográficas

- **Zona:** es una variable categórica, que corresponde a la zona geográfica a la que pertenece cada hogar según la representatividad de la VIII EPF. La referencia geográfica resultó relevante en el estudio *Lifestyles through expenditures* (Keister *et al.*, 2016), donde se explicaba el comportamiento del cluster de los *drivers* por su residencia en zonas suburbanas, lo que los obligaba a manejar grandes distancias y tener mayores gastos en vehículos de transporte personal.

Categorías: 1. Gran Santiago; 2. Resto de capitales regionales.

- **Tenencia de la vivienda principal:** es una variable categórica que indica cuál es la situación de los hogares encuestados en torno a la propiedad o tipo de ocupación de la vivienda. Esta variable se utiliza en la medición del objetivo 11 de los Objetivos de Desarrollo Sostenible (ODS) de Naciones Unidas (ONU, 2019), como parte de los componentes del indicador de "tenencia segura de la vivienda", que permite la identificación de hogares vulnerables por pertenecer a barrios marginales, pudiendo ser considerada como un *proxy* de bienestar.

Categorías: 1. Propia, totalmente pagada; 2. Propia pagándose; 3. Arrendada; 4. Cedida, herencia, sucesión y otros.

- **Edad de la jefatura de hogar:** esta variable se utiliza en dos formatos, como numérica discreta (edad) y como ordinal (edad en tramos). Se generó como un *proxy* de la edad de los miembros del hogar, que a su vez se debiera reflejar en los estilos de vida (Hayes y Finney, 2014; Keister *et al.*, 2016).

Categorías: para la variable edad en tramos, 1. 15 - 24 años; 2. 25 - 34 años; 3. 35 - 44 años; 4. 45 - 54 años; 5. 55 - 64 años; 6. 65 o más.

- **Hogar adulto mayor:** esta es una variable dicotómica, que identifica a los hogares en los cuales todos sus miembros son personas de 60 años o más. La construcción de esta variable fue tomada del IPC de adulto mayor que publica Clapes UC, y justifica su inclusión también por su utilización en el estudio “*Dominant patterns of expenditure among older people in the United Kingdom: Segmenting the older consumer using the Living Costs and Food Survey*” (Hayes y Finney, 2014).

Categorías: 0. Hogar no adulto mayor; 1. Hogar adulto mayor.

- **Estructura del hogar:** esta es una variable categórica, que corresponde a una tipología de hogares utilizada para estudios de Cepal, publicada en la investigación “Tipos de hogares urbanos y etapas del ciclo de vida familiar” (Arriagada, 1997). Esta variable combina en su interior varios aspectos que se han visto relacionados con los hábitos de consumo en esta investigación: la cantidad de miembros en el hogar, núcleos al interior del hogar y las relaciones de parentesco.

Categorías: 1. Hogares unipersonales; 2. Hogares monoparentales; 3. Hogares nucleares; 4. Hogares extensos; 5. Hogares compuestos; 6. Hogares sin núcleo.

- **Ciclo familiar:** esta es una variable categórica, y al igual que la estructura del hogar, es una tipología utilizada por Cepal (Arriagada, 1997). Además de la cantidad de miembros del hogar, núcleos al interior del hogar y sus relaciones de parentesco, utiliza las edades de los hijos y de sus madres para su construcción. Se ha estudiado el efecto del ciclo familiar en investigaciones sobre consumo de bienes durables en Fernandez-Villaverde y Krueger (2007), bajo la hipótesis de que estos gastos debieran decrecer en etapas de consolidación familiar.

Categorías: 1. Hogar no familiar; 2. Pareja joven sin hijos; 3. Inicio de la familia; 4. Expansión; 5. Consolidación; 6. Desmembramiento; 7. Nido vacío.

- **Escolaridad de el/la sustentador/a principal en tramos:** esta es una variable ordinal, a nivel hogar, que indica el mayor nivel educativo logrado por la persona que aporta más al presupuesto del hogar. Se incluye esta variable dado que se ha estudiado profusamente la relación entre el nivel educativo y el bienestar económico de las personas y sus hogares, así como también su relación con los niveles de endeudamiento, que agregan una restricción material al despliegue de los estilos de vida (2018; Cox, 2006; SBIF, 2018; Banco Central, 2018).

Categorías: 1. Menos que media completa; 2. Media completa; 3. Superior incompleta; 4. Superior completa; 5. Postgrado.

- **Tipo de estudiante:** esta es una variable categórica, construida *ad hoc* a los objetivos de esta investigación, que indica, para todos los miembros del hogar, si es que asisten a la educación formal y, en caso de asistir, a cuál, en términos agregados. Su inclusión permite revisar fácilmente la cantidad de estudiantes asistiendo a la educación superior en cada cluster.

Categorías: 1. No estudia; 2. Escolar; 3. Superior.

- **Dependencia administrativa del establecimiento educacional:** esta es una variable categórica que indica la dependencia administrativa del establecimiento educacional al cual asisten las personas que declararon asistir a la educación superior, permitiendo caracterizar de mejor manera el gasto educacional.

Categorías: 1. Universidad privada; 2. Universidad del consejo de rectores; 3. Centro de formación técnica (CFT) o Instituto profesional (IP).

- **Sistema de salud:** esta es una variable categórica que corresponde al sistema de salud al cual se encuentra afiliado el jefe o la jefa de hogar, y en el caso de que esta persona no tenga un sistema de salud, el de su pareja o cónyuge. Esta permite estudiar la relación entre el sistema previsional de salud de los hogares y sus gastos en este ítem.

Categorías: 1. No tiene; 2. FONASA; 3. FFAA; 4. ISAPRE; 5. Otro.

5.3.1.2 Variables socioeconómicas

- **Gasto e ingreso total:** ambas son variables numéricas continuas, y corresponden, respectivamente, al gasto en consumo final mensualizado de los hogares en todos los ítems de gasto considerados por la VIII EPF, y al ingreso disponible total mensual de los hogares, de acuerdo a la definición de ingreso utilizada por la misma encuesta. Se incluyen por su capacidad de resumir información socioeconómica de los clusters y para analizar la relación entre gasto e ingreso en cada uno de ellos.
- **Quintil de ingreso:** esta es una variable ordinal, que indica la posición ordenada que ocupan los hogares en la distribución del ingreso disponible mensual per cápita de los hogares. Para construir esta variable se ordena a los hogares según la variable de ingreso escogida, de forma ascendente y luego se establecen cortes para dividir la muestra en 5 porciones, cada una de alrededor de un 20% del total. Esta variable es el *proxy* más utilizado de estratificación socioeconómica en encuestas de hogares y en este estudio es importante para revisar si efectivamente la clasificación realizada mediante K-medias con variables de gasto aporta información distinta que simplemente utilizar cuantiles de ingreso para agrupar hogares.

Categorías: 1. Quintil 1; 2. Quintil 2; 3. Quintil 3; 4. Quintil 4; 5. Quintil 5.

- **Deuda mensual total y por tipo:** se trata de 8 variables numéricas continuas, la deuda mensual total y los tipos de deuda desagregadas que la componen: hipotecaria, de consumo, automotriz, educacional, en cuentas corrientes, de tarjetas de crédito bancarias, y de tarjetas de crédito de casas comerciales¹³. La EPF no tiene como objetivo estimar el endeudamiento, pero sí capturar el gasto en servicios financieros de los hogares (intereses, costos administrativos, comisiones y seguros asociados a pasivos financieros), por lo que recoge información de montos de deudas y cuotas para validar dicho gasto.

La relevancia del endeudamiento en los estilos de vida tiene una dimensión material y una simbólica o moral. La primera dimensión tiene que ver con que el endeudamiento se presenta como un doble fondo en relación a las posibilidades de acceso a bienes y servicios, es un *backup* que ofrece un nuevo horizonte al traspasar la barrera del ingreso disponible, pero que

¹³ El detalle de cada una de estas variables puede encontrarse en el Anexo 3, que ofrece un cuadro resumen que incluye a todas las variables suplementarias utilizadas en esta investigación.

está determinado por este y por otras variables (SBIF, 2018; Banco Central, 2018), ampliando en mayor medida el acceso de quienes más tienen (Matus *et al.*, 2010). Por otra parte, el endeudamiento también tiene una dimensión moral (Mayol *et al.*, 2013; Araujo y Martuccelli, 2012; Moulian, 2002), una “resaca”, que tiene que ver con el juicio social frente a la situación de endeudamiento entre los chilenos. Ambas dimensiones justifican incluir una variable para cuantificar el nivel de endeudamiento de los hogares en cada uno de los clusters, lo que ciertamente influye en sus estrategias y experiencias en torno al consumo de bienes y prácticas.

Por otra parte, el argumento para desagregar las deudas por tipo y no solo utilizar la deuda total, es que no es lo mismo endeudarse con un banco por un crédito hipotecario que hacerlo por un consumo en el corto plazo con una casa comercial. Según el Banco Central (2018), dos tercios de los hogares chilenos (66%) declara tener alguna deuda, lo que lo hace un atributo bastante común, sin embargo, encontrarse endeudado con una casa comercial es ciertamente una situación más desventajosa para los hogares, considerando las altas tasas de interés que estas entidades cobran¹⁴ (Castillo, 2011).

- **Carga financiera total y por tipo:** son 8 variables numéricas continuas, la carga financiera total y las cargas financieras por tipo de deuda, que son las mismas mencionadas anteriormente: hipotecaria, de consumo, automotriz, educacional, en cuentas corrientes, de tarjetas de crédito bancarias, y de tarjetas de crédito de casas comerciales. La carga financiera es definida como la proporción que alcanza la deuda mensual respecto al Ingreso disponible mensual total del mismo hogar (SBIF, 2018), y puede ser mayor que 100% cuando la deuda mensual es mayor que el ingreso disponible mensual total.

La importancia de este indicador es que proporciona una medida relativa de la situación de endeudamiento, y sobre cuán “ahogados” están los hogares en relación a sus compromisos financieros.

¹⁴ Según Castillo (2011), las altas tasas de interés que reportan las casas comerciales por sus créditos se justifican por sus escasas barreras de acceso, compensando el mayor riesgo de no pago de sus clientes con este mayor cobro. El riesgo tiene que ver con que las casas comerciales no tienen el mismo nivel de acceso a la información financiera de sus potenciales clientes para evaluar su capacidad de pago, mientras que los bancos cuentan con un sistema que les provee esa información.

- **Sobre endeudamiento:** existe discusión respecto a cómo debiera calcularse el sobre endeudamiento y cuál debiera ser su umbral (Anderloni y Vardone, 2008; Finlay, 2006; SBIF, 2017; SBIF, 2018). La medida que se utiliza en Chile se construye a partir de la carga financiera de los hogares o personas, pero incluso la SBIF propone dos umbrales diferentes entre el 2017 (Más del 40% del ingreso destinado al pago de deudas) y el 2018 (más del 50%). En este trabajo se tomó la recomendación del *Bank for International Settlements* (BIS, 2014), que propone un 40% de carga financiera como límite para el sobre endeudamiento.

La justificación del uso de este indicador para estudiar los estilos de vida guarda relación con lo ya expuesto sobre los dos grupos de variables anteriores, pues se trata de una radicalización de las consecuencias del endeudamiento. Araujo y Martuccelli (2012) proponen que, ante la ausencia de un discurso colectivo, es el endeudamiento, como método de coerción externo, el que pone los límites a las posibilidades de consumo, y el sobre endeudamiento no solo se explica por trayectorias o decisiones individuales, pues es inseparable de movimientos macrosociales, que le entregan un contexto y un marco de responsabilidad colectiva.

Categorías: 0. Hogar no sobre endeudado; 1. Hogar sobre endeudado.

5.3.2 Descripción general de los clusters obtenidos con K-medias

En este apartado se revisan algunas variables suplementarias con el propósito de introducir a los resultados del K-medias. En la tabla 4 se expone el tamaño de cada uno de los clusters conformados. Se muestran los valores muestrales y expandidos (con cada hogar ponderado por su factor de expansión), sin embargo, en adelante se utilizarán siempre los valores expandidos, que son los que permiten describir a la población objetivo y no solamente a los hogares de la muestra.

Tabla 4: Distribución de hogares en cada cluster muestral y expandido

Cluster	n muestral	% Muestral	n expandido	% Expandido
1	1242	8,24	267.918	8,01
2	1360	9,02	330.530	9,88
3	1825	12,10	396.273	11,85
4	1976	13,10	437.141	13,07
5	1017	6,74	221.825	6,63
6	2566	17,02	607.509	18,17
7	3068	20,34	631.501	18,88

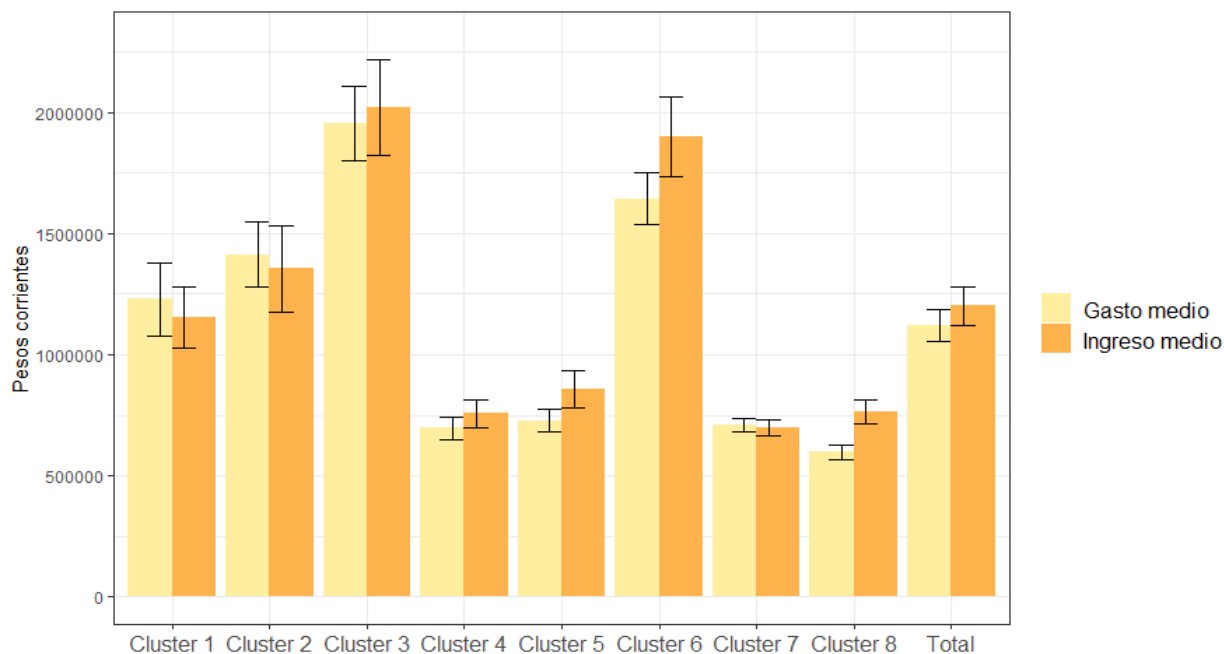
8	2026	13,44	451.478	13,50
Total	15080	100	3.344.176	100

Fuente: Elaboración propia en base a VIII EPF-INE.

Los clusters resultan bastante equilibrados. Entre estos, el cluster más numeroso es el 7 y el menos numeroso el cluster 5. Como se mencionó anteriormente, cuando se clasifica en base a montos absolutos, como el gasto equivalente, los grupos conformados tienden a ser más bien desequilibrados, generándose desde clusters que agrupan a alrededor del 90% de los hogares, hasta clusters que contienen solo una observación, lo que despoja por completo de sentido la tarea de caracterizarlos.

A continuación, en el gráfico 3 se expone la media del gasto mensual total por hogar y el ingreso disponible promedio mensual por hogar en cada cluster y para toda la muestra, además de los intervalos de confianza, desplegados sobre cada barra.

Gráfico 3: Medias e intervalos de confianza del gasto mensual total e ingreso disponible mensual total por hogar



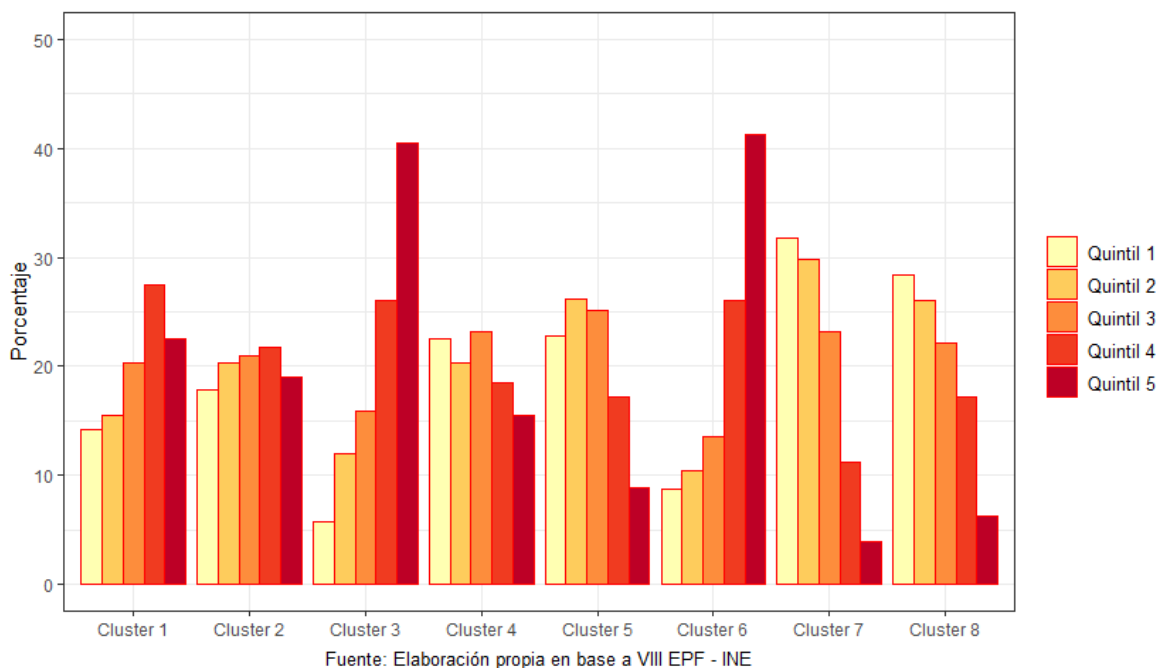
Fuente: Elaboración propia en base a VIII EPF - INE

El cluster 3 es el que presenta mayor nivel de gasto e ingreso, seguido por el cluster 6, ambos muy por encima del ingreso y gasto medio a nivel de todos los clusters. En el nivel inferior, los clusters 4, 7 y 8 son los de menor ingreso, seguidos por el cluster 5. No es posible afirmar cual tiene los ingresos medios más bajos, dado que sus intervalos de confianza se solapan.

En la tarea de describir estilos de vida, se vuelve relevante la constatación de que el ingreso es significativamente mayor que el gasto en el cluster 8 y en el cluster 5. También es llamativo que la barra que representa al gasto medio se eleva por sobre la del ingreso medio en los clusters 1 y 2, sin embargo, no es un rasgo determinante, dado que esta diferencia no es significativa.

El siguiente gráfico muestra la estructura de los clusters a nivel de quintiles de ingreso. El ordenamiento de los hogares se realizó en función del ingreso disponible mensual per cápita por hogar. Puesto que se trata de quintiles, no es necesario agregar como parámetro las columnas de quintiles a nivel total, pues el valor para cada uno es aproximadamente 20%.

Gráfico 4: Distribución de hogares en cada cluster por quintil de ingreso



Lo primero que se observa es que cada cluster tiene una estructura diferente en relación al ingreso disponible de los hogares que los componen. Esto es natural, puesto que ciertos estilos de vida requieren un cierto nivel de ingreso que los sustenten, haciendo más probable la presencia de hogares de cierto nivel de ingresos en particular. Ahora, más allá de esto, este gráfico permite observar que en cada cluster, más allá de las diferencias, hay presencia de todos los quintiles de ingreso. Este punto es importante puesto que apunta en la dirección de la hipótesis planteada en este trabajo, que propone que identificar grupos de hogares utilizando como insumo variables de gasto resultará en una clasificación distinta a la estratificación socioeconómica.

Estos resultados muestran clusters de hogares que están siendo agrupados de acuerdo a sus restricciones presupuestarias, pero también de acuerdo a gustos y preferencias, los que se ven influenciados por sus características socioeconómicas y sociodemográficas, como proxy de su posición en el espacio social. De lo anterior se desprende que hogares con un nivel similar de ingreso, incluso gasto, pueden desarrollar estilos de vida diferentes, y es por eso relevante indagar en qué atributos tienen en común los hogares pertenecientes a cada uno de los grupos identificados.

Muy relacionado con lo mostrado por el gráfico 3, el cluster 3 y el cluster 6 tienen mayor presencia de hogares del quinto quintil, que corresponde el 20% de los hogares con ingresos más altos en la población. También son los clusters con la menor presencia del quintil de más bajo ingreso. Ahora en relación a los clusters que mostraron un menor ingreso medio en el gráfico 3, el cluster 4, 7 y 8, se observa una diferencia importante que ocultaba la media entre la estructura del cluster 4, en comparación con la de los clusters 7 y 8. La estructura del cluster 4 es muy equilibrada, con una presencia de cada quintil relativamente cercana al 20%, mientras que los clusters 7 y 8 tienen una estructura escalonada descendente hacia el quinto quintil, lo que implica que tienen menos presencia de hogares de ingresos altos.

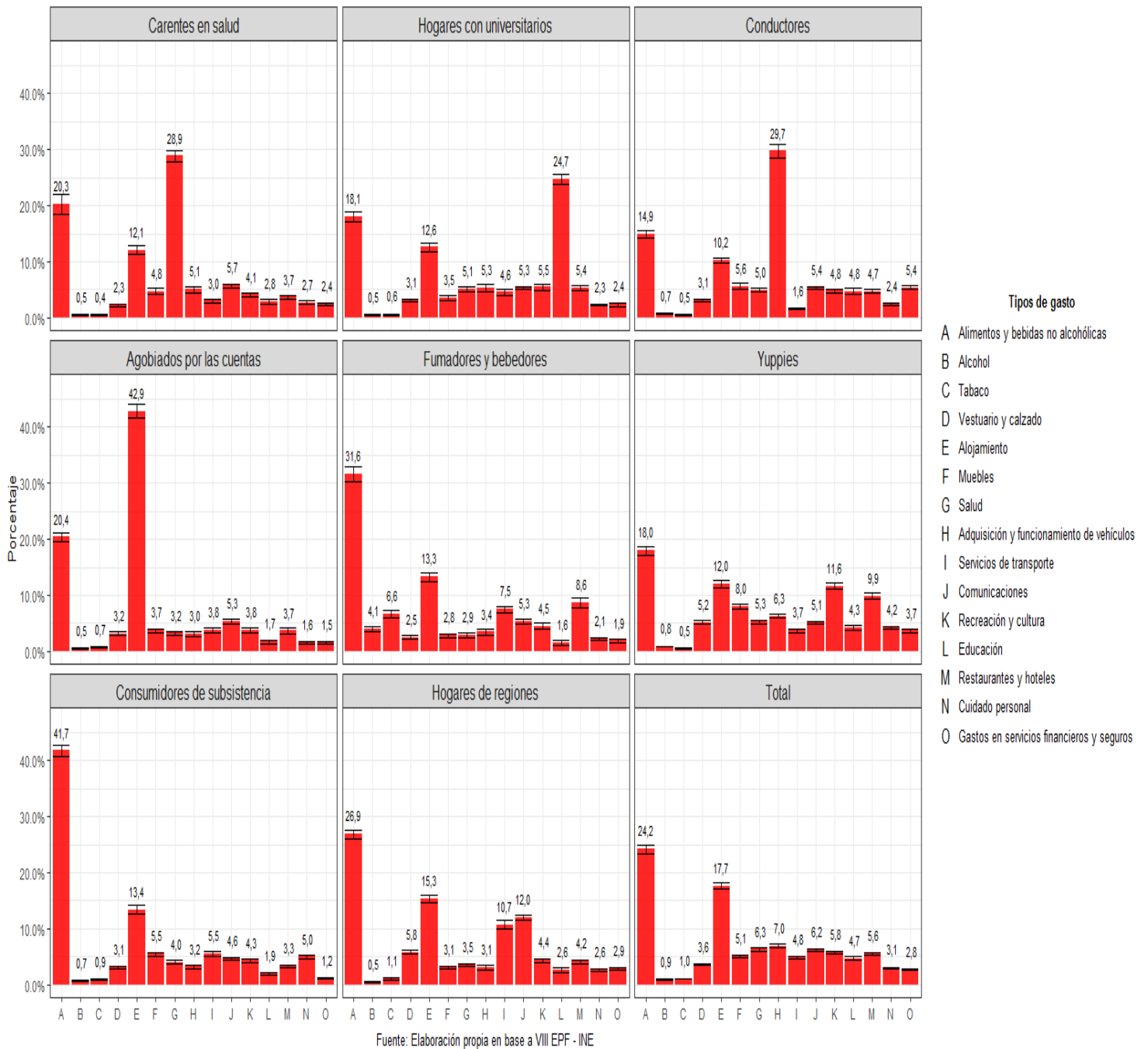
5.3.3 Denominación de los clusters

La descripción en base a las variables activas utilizadas para la clasificación es un paso básico en la interpretación de los resultados de un K-medias (Hayes y Finney, 2014; Keister *et al.*, 2016). Revisando las características de los clusters en relación a estas mismas variables, es posible identificarlos y denominarlos de alguna manera significativa (etiquetarlos), aunque claramente siempre se trata de una reducción. Si bien este es el insumo principal, también se utilizan las variables suplementarias para guiar esta denominación, para ser más precisos o cuando los descriptivos de las variables incluidas en el modelo no resultan suficientes para esta tarea.

El gráfico 5 muestra la estructura del gasto para cada uno de los 8 estilos de vida identificados, junto con el nombre que les fue asignado de acuerdo a sus características predominantes de gasto y también según información suplementaria, junto con la estructura promedio para el total de los grupos en las 15 variables de gasto utilizadas para el modelo, que servirá como parámetro para

revisar la estructura de gasto de cada cluster. En todos los casos se agrega sobre cada barra el intervalo de confianza.

Gráfico 5: Estructura del gasto total y por cluster



En la última faceta, en la esquina inferior derecha, se aprecia que en relación al gasto total, el gasto en Alimentos y bebidas no alcohólicas es el más importante, y le sigue el gasto en alojamiento, el que, en términos generales, incluye arriendos, gasto común y gastos periódicos en servicios de la vivienda, como electricidad, gas, agua, entre otros, excluyendo los gastos en servicios de internet y telefonía, que pertenecen a la división 8, Comunicaciones.

Esta información funcionará como parámetro de comparación para entender las especificidades de la estructura del gasto en cada uno de los clusters de hogares que se revisarán a continuación. La hipótesis detrás de K-medias indica que al agrupar a los hogares en función de estas variables, debieran conformarse grupos que tengan estilos de vidas similares internamente, a la vez que distintos de los que ostentan los demás grupos.

El cluster 1 presenta un elevado gasto en Salud, 22,6% por sobre la media del gasto en Salud para todos los perfiles. Dentro de este cluster se encuentran, entre otros hogares con elevados gastos regulares en salud, los hogares que han tenido eventos importantes en este ítem, como son hospitalizaciones, intervenciones ambulatorias o tratamientos, los cuales, aun mensualizados siguen siendo significativos. Por este motivo fueron denominados “*Carentes en salud*”. El cluster 2 es el que presenta el más significativo gasto proporcional en Educación (20% sobre la media total), sin embargo, se les señaló con el nombre de “*Hogares con universitarios*”, porque este gasto no se realiza en cualquier tipo de educación, como se verá más adelante. La denominación del cluster 3, “*Conductores*”, fue tomada del estudio de Keister *et al.*(2016), donde al grupo con mayor gasto en Adquisición y mantención de vehículos se les llamó *drivers*. Este cluster presenta un gasto en Adquisición y mantención de vehículos 21,8% por sobre la media total, y al mismo tiempo el gasto más bajo en Servicios de transporte, donde se encuentra el gasto en transporte público.

La denominación otorgada al cluster 4, “*Agobiados por las cuentas*”, fue tomada de la investigación de Hayes y Finney(2014), donde se llamó *Burdened by Bills* a los hogares que, al igual que los del cluster 4, tienen un importante gasto proporcional en la división 4, que incluye fundamentalmente “las cuentas” asociadas al costo periódico de la vida en alojamiento y servicios básicos. Este grupo pudo ser llamado también “*Arrendatarios*”, de acuerdo a sus características, pero se prefirió una denominación más general.

En el cluster 5 fueron agrupados los hogares denominados aquí como “*Fumadores y bebedores*”. Si bien el gasto en Alcohol y el gasto en Tabaco pertenecen ambos a la división 2, y fueron disociados para tener la posibilidad de encontrarlos exacerbados en grupos diferentes, como ocurrió en los estudios revisados, aunque acá aparecen claramente relacionados al mismo grupo. Por otra parte, el cluster 5 figura como el segundo de mayor gasto en Restaurantes y hoteles, que

es la división que incluye los gastos en consumo de alcohol en locales establecidos (pubs, bares discotecas, restaurantes, etc.).

Se hallaron en el cluster 6 muchos rasgos similares a los que caracterizaron al grupo de los *socialites* en el estudio de Hayes y Finney (2014). Principalmente, el cluster 6 presenta un alto gasto proporcional en productos y servicios culturales y recreativos (5,9% por sobre la media total), un alto gasto en Restaurantes y hoteles (4,3% sobre la media total) y, al mismo tiempo, un gasto superior al del resto de los grupos en Equipamiento del hogar, que en este caso es de un 2,8% por sobre la media total de 5,2%. Atendiendo a la similitud los gastos de los *socialites* con los de los hogares del cluster 6, habría estado justificado utilizar la misma etiqueta para identificarlos, sin embargo, se decidió investirlos con un nombre diferente, que según sus características sociodemográficas y socioeconómicas es más preciso: *yuppies*. La palabra *yuppies* es un acrónimo de Joven Profesional Urbano (*Young Urban profesional*). Si bien es un concepto proveniente del periodismo y ensayismo en Estados Unidos, y no de la tradición sociológica, resulta útil para sintetizar sus características principales, las que se irán desarrollando a lo largo de este capítulo.

Los hogares del cluster 7 fueron denominados “*Consumidores de subsistencia*”. En el estudio de Hayes y Finney (2014) el grupo con mayor gasto en alimentación¹⁵ fue denominado *Foodies*, “sibaritas”, sin embargo, el cluster 7 es un grupo de bajo ingreso, que destina a la compra de alimentos una proporción importante de su dinero probablemente por una restricción presupuestaria.

Finalmente, el cluster 8 a simple vista no presenta ningún gasto especialmente sobre la media total que permita darles un sentido. Estos hogares muestran la mayor proporción de gasto en Vestuario y calzado, en Servicios de transporte y en Comunicaciones, y la segunda mayor en alojamiento, antecedentes que no resultaron suficientes para gatillar alguna intuición sobre qué etiqueta pudiera darles identidad. Entonces, fue necesario hacer uso de variables suplementarias, y al revisar la composición geográfica de este cluster por zona (si pertenecen a Gran Santiago o

¹⁵ Hay una importante distinción en el clasificador CCIF en relación con el gasto en alimentos para el consumo en el hogar y fuera del hogar. La división 1 no contempla todo el gasto en alimentación y bebidas no alcohólicas, solo corresponde al gasto en supermercados, ferias, almacenes y establecimientos de venta productos no listos para el consumo. La comida fuera de casa, en restaurantes, fuentes de soda, bares, pubs, entre otros, que podría ser más una práctica sibarita, se consigna en la división 11 de Restaurantes y hoteles.

alguna capital regional) fue posible observar que tiene una presencia de hogares *de regiones*¹⁶ mayor a la media, e incluso mayor que la de hogares de la Región Metropolitana, que tiene un peso más elevado en la cobertura de la encuesta. Este cluster presentó un 52,4% de hogares *de regiones*, mientras que la proporción de hogares *de regiones* a nivel nacional es 44,1%.

Al tabular estos gastos comparando su importancia relativa *en regiones* y en la Región Metropolitana, fue posible comprobar que efectivamente los hogares del resto de las capitales regionales presentan una proporción de gasto mayor y de forma significativa en estos ítems. Gastan más en servicios de transporte, pero no en cualquier transporte: en taxi colectivo; gasta más en alojamiento, pero no en arriendos (la mayoría son propietarios): en electricidad, gas y otros combustibles. Dado lo anterior, se les dio una denominación en base a una variable suplementaria, a diferencia de los otros clusters.

Esta es una descripción preliminar de los grupos. En el siguiente apartado se seguirá profundizando en sus características a través del cruce con variables suplementarias, para develar quienes ostentan estos estilos de vida, en términos sociodemográficos y socioeconómicos, y también profundizar en aspectos relevantes de estos estilos de vida, haciendo un *zoom* en algunos de sus gastos representativos.

5.3.4 Los estilos de vida de los hogares chilenos

En los dos apartados anteriores en este capítulo, se describieron los clusters obtenidos con K-medias de forma general, definiéndolos en función de su tamaño, medias de ingreso y gasto, y quintil de ingreso, y luego en función de sus gastos a nivel de división, y en algunos casos grupo (las variables activas del modelo).

Este último apartado busca responder básicamente dos preguntas: ¿cómo son los estilos de vida configurados y quiénes los ostentan? Preguntas que se desprenden de los objetivos específicos 1.2 y 1.3, respectivamente. Para responder la segunda parte de la pregunta, se utilizaron las variables suplementarias descritas anteriormente, las que permitieron caracterizar a los grupos, distinguirlos uno de otro, así como también desarrollar un discurso respecto a la relación entre

¹⁶ Aun teniendo claridad sobre la identidad de este grupo, fue extremadamente complejo hallar una expresión sintética para agrupar en un mismo conglomerado a los hogares que son de regiones distintas a la metropolitana. Se utiliza la expresión *de regiones* en cursiva para hacerse cargo de que es una distinción falaz- dado que la región Metropolitana también es una región-, pero que en términos comunicativos funciona.

ciertas características de los hogares y sus estilos de vida a partir del consumo. Para responder a la primera parte de la pregunta, adicionalmente al análisis de las variables activas, se profundizó en el tipo de gastos que realizan los hogares de cada cluster al nivel más desagregado, de clase, subclase y producto, dado que en la exploración de sus estadísticos descriptivos fue posible reconocer detalles que aportaban mayor riqueza a los análisis.

Como se expuso en el diseño metodológico de este trabajo, realizar una clasificación utilizando los últimos niveles de desagregación del clasificador CCIF presentaba dificultades para el tratamiento computacional del algoritmo de K-medias (especialmente para evaluar el valor óptimo de K) y también para la interpretación de los resultados, debido a la gran cantidad de variables involucradas. Sin embargo, la revisión de estos gastos más específicos aporta elementos concretos sobre los estilos de vida de los hogares. A mayor nivel de agregación se ocultan preferencias o situaciones que delinean los estilos de vida y que se encuentran al interior de cada tipo de gasto. Por ejemplo, los *Carentes en salud*, ¿tienen un porcentaje de gasto en salud más alto porque algún miembro del hogar estuvo hospitalizado, o porque gastan mucho en medicamentos o en consultas médicas? Este tipo de distinciones quedan ocultas al analizar el gasto por división, pues solo indica la magnitud del gasto en salud como un todo. Lo mismo sucede con los gastos en alimentación y bebidas no alcohólicas. Ciertamente el estilo de vida de un hogar es diferente si consumen más o menos vegetales, o según el tipo de pan o carne que consumen. Esto sucede con todos los gastos que fueron utilizados para la clasificación, en mayor o menor medida, pero también es claro que no resulta deseable hacer una descripción exhaustiva de las diferencias en el gasto entre los 8 clusters de hogares conformados a nivel de producto, recordando que para este nivel existen 1.186 variedades diferentes, por lo que solo se destacaron los más llamativos o interesantes.

Dado lo anterior, a continuación se hará una descripción para cada uno de los estilos de vida, que combina atributos sociodemográficos y socioeconómicos de los clusters, con un *zoom* en ciertos gastos representativos o simbólicos, identificados a partir de la revisión de matrices con estimaciones de gasto absoluto y proporcional a nivel de clase, subclase y producto, generadas para ese fin.

5.3.4.1 *Cluster 1: Carentes en salud*

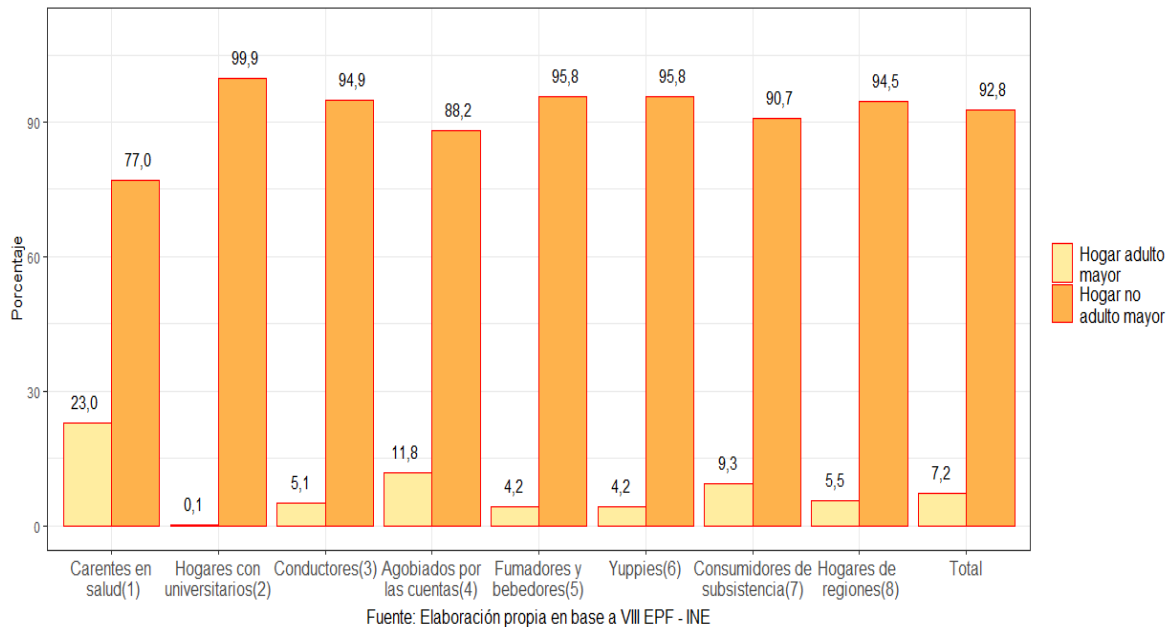
El cluster 1 de los *Carentes en salud* corresponde al 8% de los hogares y se caracteriza y denomina por el elevado porcentaje de su presupuesto que dedica al gasto en Salud (22,6% sobre la media de todos los grupos). Para los hogares que han sufrido la hospitalización de alguno de sus miembros durante al año anterior a la aplicación de la encuesta, este gasto es especialmente importante, sin embargo, el más importante para este grupo es el gasto en medicamentos, que asciende a \$99.238 en promedio, el triple del gasto en medicamentos del cluster que le sigue (los *Conductores*, con \$33.354).

Los *Carentes en salud* son quienes tienen un mayor gasto en prácticamente todos los ítems de gasto en salud (medicamentos, consultas médicas, artefactos de corrección médica, etc.). Es interesante observar que en los únicos 2 ítems donde otros clusters gastan más es en cirugía y procedimientos médicos dermatológicos (*Conductores*) y cirugías plásticas (*Yuppies*), ambulatorias y con hospitalización, justamente los gastos más relacionados a aspectos suntuarios de la salud, y que no reflejan necesariamente una “carencia” o necesidad en salud.

En relación a otros ítems de gasto, presentan el segundo gasto proporcional más bajo en Vestuario y calzado. Tienen un gasto bajo la media en Servicios de transporte y también en Adquisición y mantención de vehículo, es decir, en toda la división de transporte, lo que podría indicar que muchos de estos hogares no se desplazan demasiado o al menos no cotidianamente. Uno de sus rasgos sociodemográficos principales es que son el cluster de mayor edad. Tienen una media de edad de sus jefaturas de hogar de 58,6 años, el 40,9% de estas jefaturas son adultos mayores (se encuentran en el tramo de 65 años o más) y el 23% de los hogares del cluster son conformados exclusivamente por adultos mayores¹⁷, como se puede apreciar en el gráfico 6.

¹⁷ El indicador presentado en el gráfico 6 fue construido tomando la metodología del IPC de adulto mayor publicado mensualmente por Clapes UC, quienes consideran hogar adulto mayor a los que cuentan entre sus miembros exclusivamente a personas de 60 años o más.

Gráfico 6: Distribución de hogares de adultos mayores en cada cluster



Según el estudio de Hayes y Finney (2014) sobre el consumo de los hogares de adultos mayores en Reino Unido, el porcentaje del gasto en transporte entre los hogares del tramo de 50 a 55 años decrece de 18% a 7% hacia el último tramo de 80 años y más, así como también el gasto en Vestuario y calzado baja de 6% a 3% entre los mismos tramos de edad. Tienen un perfil de consumo muy diferente a los *Yuppies* en lo que refiere a actividades recreativas, culturales, viajes o salidas a comer fuera de casa. Esto indicaría que ser adulto mayor explica en gran medida el estilo de vida de los *Carentes en salud*.

En términos socioeconómicos, los *Carentes en salud* se encuentran en el promedio de ingreso y gasto, y aunque no es significativo, es uno de los dos clusters que presenta una media de gasto estimado por sobre el ingreso. Por otra parte, en mayor medida que el resto de los grupos, son dueños de sus viviendas, por lo que en la mayoría de los casos no pagan arriendo y además cuentan con activos.

Otras de sus características sociodemográficas interesantes son: que se trata del grupo con mayor proporción de mujeres por hogar, tienen una presencia mayor a la media de hogares

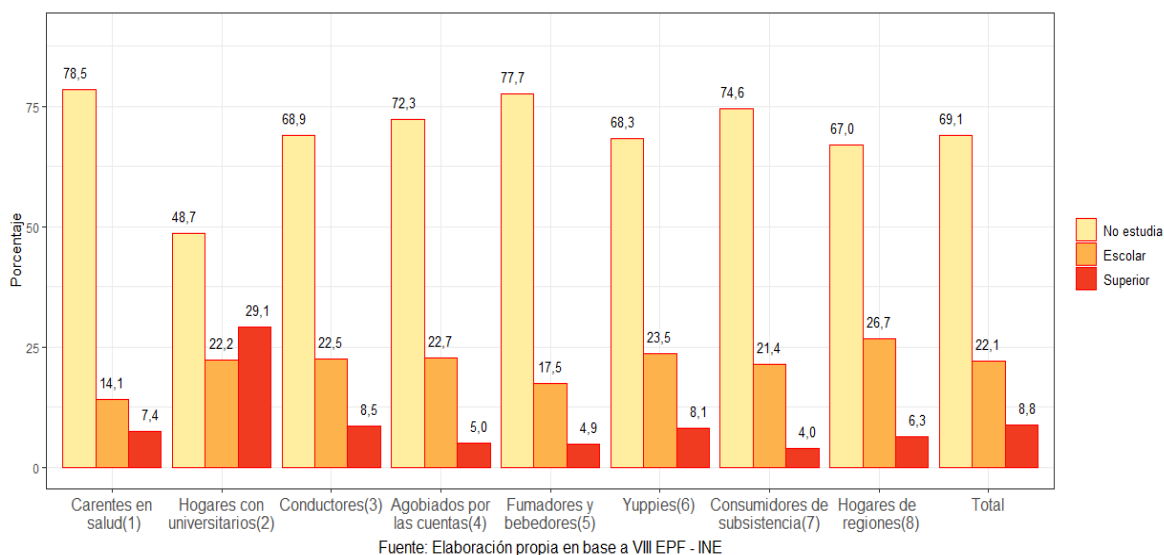
unipersonales y el 77,1% de los hogares unipersonales de adultos mayores están conformados por mujeres. Destacan los hogares con nido vacío en relación al ciclo familiar, entendiéndose por nido vacío un hogar donde hay una pareja mayor y no hay hijos residiendo con ellos, y, consecuentemente, es el cluster con menor proporción de personas asistiendo a la educación formal (78,5% de las personas no estudia).

Algo que llama la atención respecto a este grupo es que siendo justamente los que gastan mayor proporción de su presupuesto en salud, tengan un mayor porcentaje de afiliados a los sistemas de previsión de Fuerzas Armadas respecto de la media de los demás cluster (6,7%, que equivale a un 4,2% por sobre la media total), los cuales son criticados por ser más ventajosos que los sistemas disponibles para el resto de la población, y el menor porcentaje de personas afiliadas a FONASA (78,4%, mientras que la media de los grupos es 84,7%), que como sistema público debiera otorgar mayores garantías.

5.3.4.2 Cluster 2: Hogares con universitarios

El cluster 2 de los *Hogares con universitarios* está conformado por el 9,9% de los hogares y su nombre se debe a que son hogares con un alto gasto en promedio en la división 10 (20% por sobre la media del total de los grupos), que incluye gastos en educación en todos los niveles, formal y no formal, desagregado en aranceles, matrícula, cuotas de centros de padres, etc. Si bien este grupo presenta gastos educativos más altos que el resto de los grupos en todos los niveles, es especialmente relevante el gasto en educación terciaria o superior (63,4% del gasto en educación de este cluster), y especialmente en educación universitaria, que asciende al 83,4% de los gastos de los *Hogares con universitarios* en educación superior. Por otra parte, como se puede observar en el gráfico 7, los *Hogares de regiones*, *Conductores*, *Agobiados por las cuentas* y *Yuppies*, tienen un mayor porcentaje de escolares, por lo que habría sido erróneo apresurarse y llamar "*Hogares con estudiantes*" a este grupo.

Gráfico 7: Distribución de personas en cada cluster por tipo de estudiante



Un último aspecto a destacar de este cluster en relación a sus gastos en educación superior, es que la dependencia administrativa que se impone es universidad privada, casi duplicando en su grupo a quienes asisten a las universidades del consejo de rectores. Un 46,1% de los estudiantes de educación superior en este cluster asisten a universidades privadas, mientras que un 27,6% lo hacen a universidades del consejo de rectores y un 26,3% a centros de formación técnica o institutos profesionales.

En relación a otros ítems de gasto, presentan gastos proporcionales más bajos que la media total en alimentación y bebidas no alcohólicas, cercanos a los de los *Conductores* y *Yuppies*, que tienen una media de ingresos significativamente más alta. Esto también ocurre en el estudio de Keister *et al.* (2016), en el cluster *students and their parents*, y resulta llamativo, dado que existe una correlación inversa entre el ingreso y el gasto porcentual en alimentación y bebidas no alcohólicas, lo que se puede observar, por ejemplo, al cruzar la participación de la división 1 con el PIB *per cápita* ajustado por poder adquisitivo de los países (INE Chile, 2018b).

En términos socioeconómicos, los *Hogares con universitarios* se encuentran dentro del intervalo de confianza de la media de ingresos del total de los grupos, con una media superior a los *Carentes en salud*, aunque no significativamente, pero significativamente inferior a la de los

Conductores y los *Yuppies*. Este grupo, al igual que los *Carentes en salud*, tienen un gasto medio superior a sus ingresos medios, aunque no de forma significativa.

En relación a sus características sociodemográficas, es un cluster joven. La media de edad de sus jefaturas de hogar es de 46,8 años. Además es el cluster con menor proporción de adultos mayores jefes de hogar (8,3%, respecto de una media total de 21,5%) y con menor proporción de hogares exclusivamente de adultos mayores (0,12%). Tienen la más alta proporción de hogares sin núcleo (8,33%), que es la categoría donde son clasificados en las encuestas, entre otros, los hogares de estudiantes. También pertenece a este cluster la mayor proporción de hogares en desmembramiento (58,1%), que son los hogares con hijos en edad universitaria residiendo en el hogar, y la menor proporción de hogares con nido vacío.

Por otra parte, los *Hogares con universitarios* presentan la mayor proporción de sustentadores o sustentadoras principales que reportan tener educación “superior incompleta” como el mayor nivel alcanzado (17,4%), Esto se debe a que existen personas que se encuentran estudiando y además trabajan, ya sea para mantenerse a sí mismas o a sus familias.

Finalmente, otros rasgos que caracterizan a este cluster son: que adquieren más bicicletas que el resto de los grupos, lo que permite derivar que es un método de transporte importante para ellos; que pesar de no ser el grupo con mayores gastos en tecnología, sí son quienes adquieren más computadores de escritorio y portátiles; y también que gastan más en actividades deportivas como suscripciones a gimnasios y clubes deportivos, aunque menos que los *Yuppies*.

5.3.4.3 Cluster 3: *Conductores*

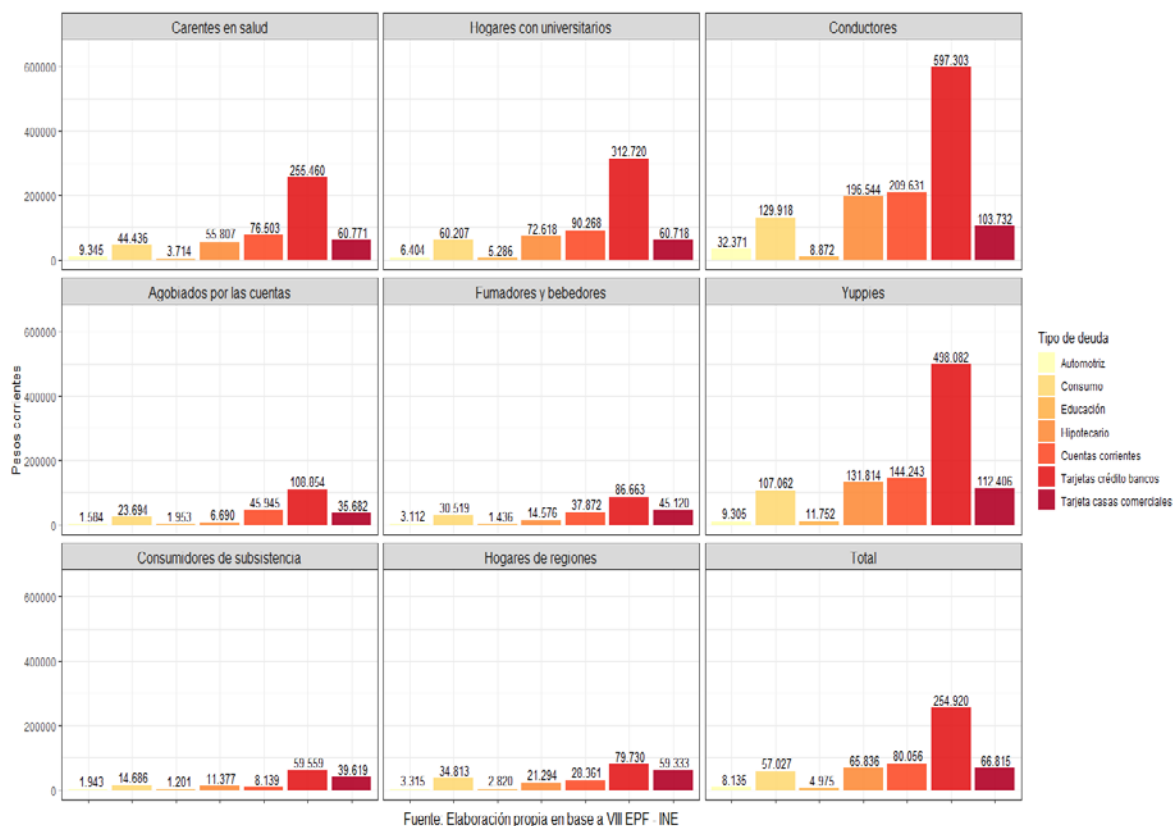
El cluster 3 de los *Conductores* corresponde al 11,9% de los hogares y fue denominado de esa manera por su elevado gasto en adquisición y funcionamiento de vehículos (21,8% sobre la media de todos los grupos) y su más bajo gasto en servicios de transporte (1,61%), que es el gasto en transporte público, tomándose su nombre del estudio de Keister *et al.*, (2016), que describieron un grupo con características similares al que llamaron los “*Drivers*”.

Los *Conductores*, en términos socioeconómicos, son los que tienen la media más alta de ingresos, aunque no es significativamente mayor que la media de ingresos del cluster que le sigue, los *Yuppies*. Por otra parte, su media de gasto si es significativamente la más alta de todos los grupos.

El grupo de los *Conductores* es el que presenta el menor gasto proporcional en alimentos y bebidas no alcohólicas, lo que de acuerdo a datos empíricos se relaciona con que tienen un mayor bienestar económico en general (INE Chile, 2018b). También tienen la proporción más baja de gasto en alojamiento, que se explica en gran medida por que un 71,8% de los hogares de este grupo son propietarios de sus viviendas. Si bien la mayoría de estos hogares está aún pagando su vivienda, es necesario destacar que los pagos en dividendo no son sumados al gasto en la EPF, por ser considerados inversión y no consumo final. Tienen el segundo gasto proporcional más alto en equipamiento del hogar y el más alto gasto en servicios financieros y seguros, aspecto que se ve directamente relacionado con sus altos niveles de endeudamiento y sobre endeudamiento.

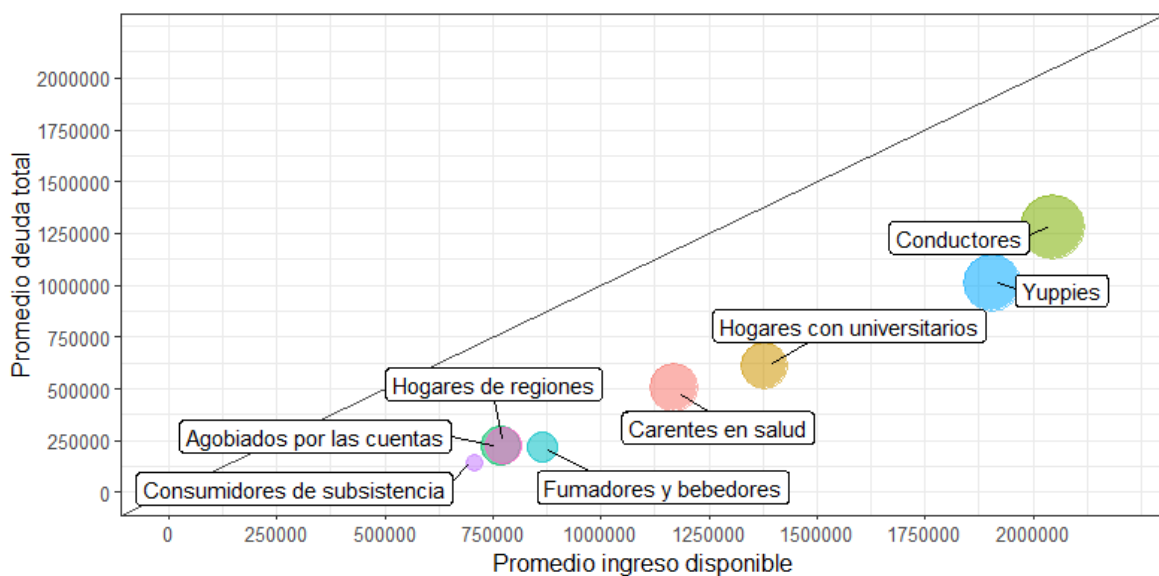
Como se indicó previamente, al describir las variables suplementarias, la Encuesta de Presupuestos Familiares captura el endeudamiento de forma indirecta al registrar los gastos en servicios financieros, lo que permite contar con esta información para cada hogar. En el gráfico 8 es posible observar el promedio del monto que los diferentes clusters pagan por concepto de deudas mensualmente.

Gráfico 8: Promedio deuda mensual por tipo de deuda según cluster y total



Quienes dedican en promedio mayores ingresos al pago de deudas son los *Conductores*, seguidos de los *Yuppies*, quienes solo superan a los primeros en el pago de deudas universitarias y de tarjetas de crédito de casas comerciales, y ambos se encuentran bastante alejados en el monto destinado al pago de deudas respecto a los demás clusters. Ahora bien, se indicó que los *Conductores* son también el cluster con mayores ingresos, por lo que para analizar su situación es necesario revisar su deuda sobre su ingreso, es decir, su carga financiera. El gráfico 9 resume en un plano de coordenadas la relación entre la deuda total promedio (eje y) y el promedio del ingreso disponible mensual (eje x) de cada uno de los estilos de vida. El gráfico muestra también, a través del tamaño de los círculos, la carga financiera de los hogares. La línea diagonal que cruza el gráfico marca el umbral de la carga financiera de 100%, es decir, donde el pago de deudas es igual al ingreso disponible.

Gráfico 9: Relación entre ingreso disponible y monto deuda según cluster



Fuente: Elaboración propia en base a VIII EPF - INE

Los *Conductores* tienen el promedio más alto de ingreso, pero también de deuda, lo que redundaría en una alta carga financiera. Se puede advertir en el gráfico 9 una relación casi lineal entre el ingreso y la deuda, salvo por los *Consumidores de subsistencia* y *Fumadores y bebedores*, que tienen una deuda proporcionalmente más baja que el resto de los clusters en relación a su ingreso.

Por otra parte, los *Conductores* se desmarcan de su escolta, los *Yuppies*, al revisar sus niveles de sobre endeudamiento. El 42,2% de los hogares del cluster de los *Conductores* está sobre

endeudado, es decir, tienen una carga financiera que supera el 40% de su ingreso mensual, mientras que entre los *Yuppies* un 30,7% de los hogares está sobre endeudado.

El cluster 3 pudo ser llamado, frente a estos antecedentes, los “Endeudados” o “Sobre endeudados”, en lugar de *Conductores*, sin embargo, se decidió evitar esta etiqueta, porque el concepto de endeudamiento se percibe entre los chilenos con una carga negativa (Mayol *et al.*, 2013; Araujo y Martuccelli, 2012; Moulian, 2002). Dado que este es el cluster de más altos ingresos, el alto nivel de endeudamiento no alcanza a ser una restricción para el despliegue de sus estilos de vida y aspiraciones de consumo, situación que sí hubiera sido determinante para otros clusters. Como indica Félix Parrado, ex Superintendente de la SBIF, el nivel de “sanidad” de una carga financiera depende del nivel de ingresos absolutos que tengan las familias. Si una familia cuenta con ingresos elevados, aunque su carga financiera sea alta, el porcentaje disponible para el resto de los gastos del hogar seguirá siendo una suma importante en términos absolutos, mientras que a una familia de bajos ingresos o ingresos medios, una carga mensual de 50% les deja un margen estrecho para otros gastos (Arroyo, 2017).

Ahora, en términos sociodemográficos, este grupo es el tercer cluster con jefatura de edad más joven (48,1 años promedio) y también el segundo cluster con mayor proporción de jefaturas de hogar adultos jóvenes (un 44,7% tienen entre 25 y 44 años). Respecto a la estructura del hogar, tienen la mayor cantidad de hogares nucleares (66,2%) y, junto con los *Yuppies*, la mayor cantidad de hogares con hijos menores a 18 años (37,2%). Son probablemente el cluster que más se acerca a la idea de familia tradicional. En términos educacionales, son el segundo cluster con mayor nivel educativo, un 46,5% de sustentadores o sustentadoras principales tienen estudios superiores completos como mayor nivel alcanzado y además un 6,4% completó un postgrado. También tienen la mayor cantidad de jefatura de hogar masculina (75%).

Al observar los gastos absolutos (en monto) de los *Conductores*, es posible observar otros aspectos de su estilos de vida. Son quienes gastan más en arriendo de estacionamientos y garajes para el uso del hogar, son quienes tienen un gasto absoluto mayor en cuentas y servicios periódicos del hogar (suministro de agua, electricidad, gas de cañería y gastos comunes), salvo en arriendo. Los *Conductores* tienen el segundo mayor gasto en servicio doméstico, después de los *Yuppies*, y gastan más en todos los ítems de adquisición y mantenimiento de vehículos, salvo en la adquisición de bicicletas, donde los superan los *Hogares con estudiantes*. Si bien el rasgo

principal de este grupo se explica en parte por la adquisición de algún vehículo durante el último año previo a la aplicación de la encuesta (y por eso tienen un gasto mensual promedio de \$432.989 en ese ítem), también gastan más que el resto de los clusters en mantención y funcionamiento de vehículos, por lo que es claramente el uso de vehículo es un rasgo importante de su estilo de vida.

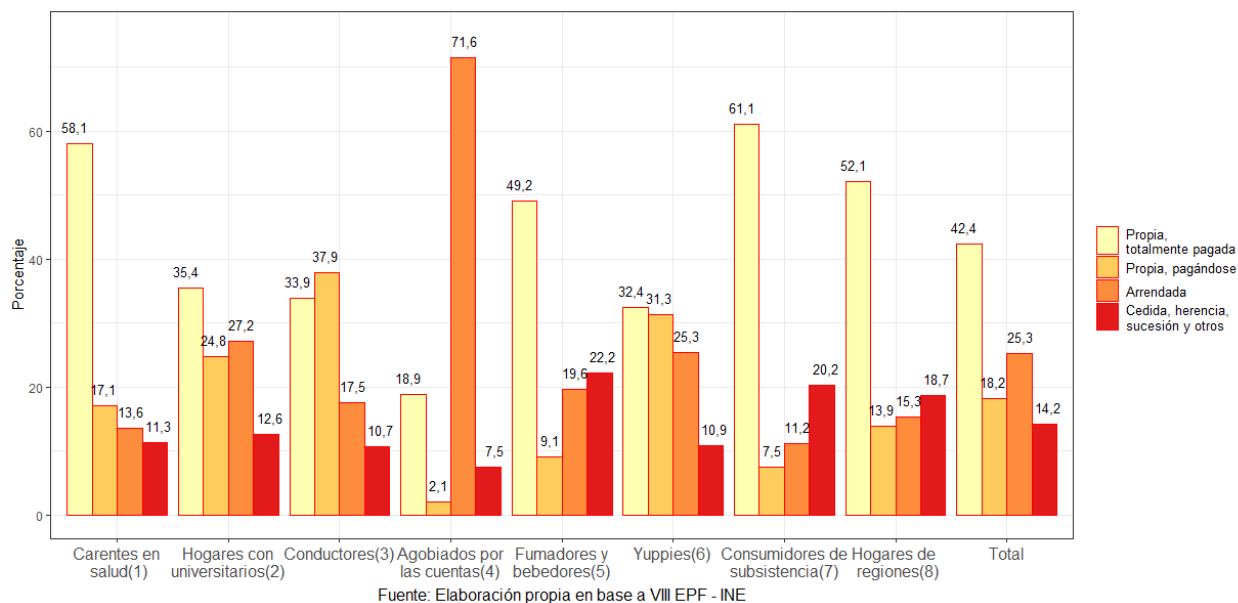
5.3.4.4 *Cluster 4: Agobiados por las cuentas*

El cluster 4 de los *Agobiados por las cuentas* corresponde al 13,1% de los hogares. Su nombre lo obtienen por su semejanza a los “*Burdened by bills*” del estudio de Hayes y Finney(2014), y su característica principal es su alto gasto en “cuentas” asociadas al costo periódico de la vida en alojamiento y servicios básicos del hogar (división 4).

En términos socioeconómicos, tienen un ingreso medio por hogar más bien bajo (\$757.060 mensuales), aunque no es significativamente diferente a los ingresos de los otros cluster de bajo ingreso, y es bastante equilibrada su composición por quintil de ingreso disponible, aunque, por supuesto, tienen una menor presencia de hogares del cuarto y quinto quintil que del resto de los quintiles.

Los *Agobiados por las cuentas* tienen un bajo gasto proporcional en la mayoría de los ítems de gasto: en equipamiento del hogar, en salud, en adquisición y mantención de vehículos, en educación, restaurantes y hoteles y en cuidado personal, debido a que su gasto representativo (en “cuentas”) asciende a casi el 50% de su gasto mensual. También tienen un bajo gasto proporcional en servicios financieros y seguros, lo que se relaciona, como muestra el gráfico 10 con que la mayor parte de los hogares que conforman este cluster son arrendatarios (71,6%) y solo un 2,1% se encuentra pagando su vivienda y, por lo tanto, tienen un compromiso hipotecario con alguna institución financiera.

Gráfico 10: Distribución de hogares en cada cluster por tipo de tenencia de la vivienda principal



En términos sociodemográficos, este cluster tiene la mayor proporción de hogares unipersonales (25,4%), la mayor cantidad de hogares compuestos (2,59%), es decir que incluye padre o madre del jefe de hogar además de otros no parientes, y también tiene una importante participación de hogares sin núcleo (4,6%). En relación al ciclo familiar, los *Agobiados por las cuentas* tienen la mayor presencia de hogares no familiares (30,1%) y de parejas jóvenes sin hijos (9,7%), lo que permite suponer que este grupo se aleja un poco de la concepción tradicional de hogar familiar compuesto por una pareja e hijos.

En relación a su sistema de salud, son quienes tienen la mayor prevalencia de jefatura de hogar y su pareja sin sistema de salud (6,8%), que es la situación más precaria en este ámbito de la seguridad social.

En términos educacionales, el 5% de las personas que componen este cluster asiste a la educación superior, y de ese 5% la mayoría (48%) asiste a universidades del consejo de rectores, a pesar de

que la tendencia general entre los estudiantes de educación superior es asistir a universidades privadas.

Si bien el nombre de este cluster refiere a su alto gasto en “cuentas” en general, lo que constriñe su presupuesto para realizar otro tipo de consumos, lo cierto es que en términos absolutos lo que destaca es su gasto específicamente en el arriendo de la vivienda (\$177.925 mensuales promedio), que triplica al promedio de gasto en arriendo del cluster que le sigue en este ítem, y que resulta especialmente alto considerando que hay un 28,4% de hogares en este grupo que no paga arriendo (son propietarios, su vivienda es cedida, heredada, o tienen otro tipo de tenencia). En relación al resto de las “cuentas”, quienes más gastan, como se indicó previamente, son los *Conductores*.

5.3.4.5 Cluster 5: *Fumadores y bebedores*

El cluster 5 de los *Fumadores y bebedores* es el grupo más pequeño, está conformado por el 6,6% de los hogares, y debe su nombre a su alto gasto proporcional en la división 2, específicamente en alcohol y tabaco. Los hogares de este cluster tienen el mayor gasto proporcional en alcohol (4,1%) y al mismo tiempo en tabaco (6,6%), a pesar de que estos gastos fueron incluidos por separado en el modelo, lo que permite intuir una fuerte relación entre beber y fumar.

Los *Fumadores y bebedores* se encuentran entre los clusters de ingresos medios-bajos, con una media de ingresos de \$857.167. Tienen la particularidad de ser uno de los dos clusters, junto con los *Hogares de regiones*, que tiene una media de ingresos significativamente mayor su media de gasto mensual.

Los hogares del cluster de los *Fumadores y bebedores* tienen el segundo gasto proporcional más elevado en alimentos y bebidas no alcohólicas (31,6%), tienen el segundo gasto más bajo en vestuario y calzado (2,5%), el gasto más bajo en equipamiento del hogar (2,8%), el más bajo en educación (1,6%), el segundo más alto en servicios de transporte (7,5%), y el segundo más alto en restaurantes y hoteles (8,6%), que es el ítem de gasto donde se registra el gasto en bares, pubs, cafés, entre otros lugares donde se expende alcohol. También tienen el gasto más bajo en salud (2,9%), lo que se contrapone con los hallazgos de la investigación de Keister *et al.*(2016), donde el cluster con mayores gastos en salud (*Health constrained*) fueron además quienes tenían mayores gastos en tabaco y productos relacionados.

En términos sociodemográficos, los *Fumadores y bebedores* son los más metropolitanos (66,3% de los hogares), tienen la mayor proporción de viviendas con tenencia irregular: cedidas, heredadas, o en otra situación (22,2%), que son hogares no propietarios pero que no pagan arriendo.

En relación al ciclo familiar, tienen la segunda mayor prevalencia de hogares en desmembramiento (52,7%), es decir, hogares con hijos mayores de 18 años. Este cluster tiene un bajo nivel educativo, la escolaridad promedio de la jefatura de hogar alcanza los 10,3 años, un 42,2% alcanzó menos que educación media completa y un 40% alcanzó a completar la educación media. El 4,9% de las personas de este cluster se encuentran asistiendo a la educación superior, y entre ellos se encuentra la segunda mayor proporción de estudiantes que asisten a CFT o IP (47,1%).

Los *Fumadores y bebedores* tienen un alto gasto proporcional en alimentos y bebidas no alcohólicas, sin embargo, en términos absolutos, el único ítem en que este cluster gasta más que los demás es en bebidas gaseosas (\$16.772 promedio mensual), lo que podría encontrar sentido en su mayor gasto en bebidas destiladas (\$6.563 promedio mensual), que triplica al gasto del cluster que le sigue. Los *Fumadores y bebedores* también tienen el mayor gasto en vinos (\$9.213 promedio mensual) y sobretodo en cervezas (\$19.061 promedio mensual), donde triplican al cluster que les sigue. Al mismo tiempo que el alto gasto descrito en alcohol, este cluster presenta un gasto promedio mensual de \$41.022 en tabaco, casi 6 veces el gasto promedio mensual del cluster que tiene el segundo mayor gasto en tabaco. En relación a su consumo en restaurantes, pubs, bares, etc., su gasto no es especialmente alto en comparación con otros clusters como los *Yuppies*. Este antecedente permite indicar que los hábitos de consumo de alcohol de los *Fumadores y bebedores* es más domiciliario, dado que no destaca su gasto en alcohol consumido en establecimientos comerciales, sino más bien la compra para consumir en privado.

Por último, el cluster de los *Fumadores y bebedores* tiene el gasto promedio mensual más elevado en transporte multimodal (\$21.930), que es el que incluye a los servicios de transporte interconectado, como el Transantiago y el sistema de transporte urbano en Valparaíso. Esto se relaciona con la mayor presencia de hogares de la región Metropolitana en este cluster.

5.3.4.6 Cluster 6: *Yuppies*

El cluster 6 de los *Yuppies* es el segundo más grande de todos los clusters y agrupa a un 18,2% del total de los hogares. Su denominación proviene de un concepto acuñado en ensayos periodísticos a comienzos de los 80's en Estados Unidos, que tiene una connotación generalmente negativa, y es usado para identificar a un incipiente grupo de jóvenes profesionales urbanos (*Young urban profesional*), que ejercen sus profesiones, tienen ingresos medios-altos, visten a la moda, tienen una gran valoración por lo material y tienden a mantenerse a la vanguardia en tecnología, entre otras características (Burnett y Bush, 1986). Los *Yuppies* tienen un alto gasto proporcional en recreación y cultura (5,9% sobre la media total), en restaurantes y hoteles (4,3% sobre la media total) y en equipamiento del hogar (2,8% sobre la media de los grupos).

En términos socioeconómicos, son el segundo cluster con mayores ingresos medios (\$1.896.552), aunque no es significativamente menor que el ingreso medio de los *Conductores*. Su gasto medio es de \$1.643.160, es inferior a su ingreso, pero sus intervalos de confianza se solapan, por lo que no es posible afirmar que ganan más de los que gastan. Además, tienen la mayor proporción de hogares del cuarto y quinto quintil de ingresos.

En relación a sus características sociodemográficas, los *Yuppies* presentan la media de escolaridad más alta entre sus sustentadores o sustentadoras principales del hogar (14,04 años) y, como se observa en la tabla 5, la mayor proporción de personas con educación superior completa (48%) y postgrados (7,7%). Por otra parte, se encuentran dentro de los clusters con jefaturas de hogar más jóvenes, con una media de edad de 46,4 años, y la mayor cantidad de jefaturas de hogar entre 25 y 44 años (49,7%). Sobre su sistema de salud, son quienes tienen en mayor proporción cobertura de salud privada (15%).

Tabla 5 : Distribución de hogares en cada cluster por tramos de escolaridad del sustentador principal

Cluster	Tramos de escolaridad				
	Menos que media completa	Media completa	Superior incompleta	Superior completa	Postgrado
1	27,3	31,2	7,9	30,1	3,6
2	15,3	30,9	17,4	32,6	3,9
3	15,9	24,3	6,9	46,5	6,4
4	34,5	28,7	9,6	25,5	1,6

5	42,2	40,5	5,0	11,7	0,6
6	16,7	21,4	6,2	48,0	7,7
7	54,8	29,8	3,9	11,0	0,5
8	35,1	42,8	6,4	15,3	0,5
Total	31,0	30,2	7,5	28,1	3,2

Fuente: Elaboración propia en base a VIII EPF-INE.

El estilo de vida de estos hogares es el más globalizado, sofisticado y orientado al placer (similar al grupo de los *pleasure seekers* identificados en el estudio de Keister *et al.*[2016]). Por ejemplo, en relación a su gasto en alimentación y bebidas no alcohólicas, son quienes presentan mayor consumo de leches vegetales (de soya, almendras, etc.). Tienen el segundo gasto más elevado en alcohol, en destilados, vinos y cerveza, y son también quienes consumen más estupefacientes, tanto marihuana y cocaína, como otras drogas sintéticas. Gastan más en vestuario y calzado, prácticamente en todos los productos relacionados a dicha división.

Los *Yuppies* también tienen el gasto más elevado en servicio doméstico. Gastan más que los *Carentes en salud* en solo un ítem de salud: cirugías plásticas. Son quienes más viajan en avión, quienes más gastan en hoteles y quienes más dinero gastan en paquetes turísticos nacionales e internacionales, en resumen, los que más viajan.

La relación con la tecnología también es importante dentro de su estilo de vida. Tienen el segundo mayor gasto en servicios de telefonía móvil, y son quienes más presupuesto invierten en equipos de audio, audífonos, cámaras y accesorios de fotografía y video, computadores de escritorio, portátiles y tablets, sin embargo, no son quienes más gastan en reparación de artículos de tecnología, lo que sugeriría que prefieren consumir objetos nuevos en lugar de reparar los antiguos.

Los *Yuppies* también tienen el mayor gasto en equipamiento y en actividades deportivas como gimnasios y clubes. Tienen mascotas, lo que se refleja en su mayor gasto que el resto de los grupos en alimentos y accesorios para mascotas y en atención veterinaria. También gastan más en libros.

Los *Yuppies* dedican más recursos que el resto de los clusters, incluso que los *Conductores* que tienen una media de ingreso mayor, en salir a comer fuera de casa. Los primeros gastan en

promedio \$134.845 mensuales en restaurantes, cafés y similares, mientras que el grupo de los *Conductores* alcanza una media de gasto en este mismo ítem de \$83.880 mensuales. Por último, los *Yuppies* son quienes más gastan en promedio en productos y servicios dedicados al cuidado personal.

5.3.4.7 Cluster 7: Consumidores de subsistencia

El cluster 7 de los *Consumidores de subsistencia* es el de mayor tamaño, agrupando a un 18,9% de los hogares. Lo que los identifica es su alto gasto proporcional en alimentación y bebidas no alcohólicas, que sobrepasa en 17,5% la media total de los clusters en esta división de gasto, y su nombre la obtienen por oposición a la denominación que Hayes y Finney (2014) le otorgan al grupo que tiene mayor gasto en alimentación en su estudio sobre el gasto de los hogares de adultos mayores en reino unido: los *Foodies* (sibaritas). Los *Consumidores de subsistencia* tienen la media de ingreso más baja de los 8 grupos, son los más pobres, y como se mencionó anteriormente, a menor poder adquisitivo, mayor participación de la división de alimentos y bebidas no alcohólicas en el gasto total del hogar. Por otro lado, la expresión *foodie* refiere a una predilección por el buen comer y por tanto, un gasto mayor que se explica por la preferencia por alimentos de variedades de precio más elevado y por las salidas a restaurantes o similares, pero ese no es el caso del cluster 7.

En términos socioeconómicos, como se mencionó en el párrafo anterior, los *Consumidores de subsistencia* son los que tienen la media más baja de ingreso (\$698.829 mensuales), sin embargo, no es significativamente menor que la media de ingresos de los *Agobiados por las cuentas* y los *Hogares de regiones*. Por otra parte, este cluster registra la mayor proporción de hogares del primer quintil (31,8%) y la menor proporción de hogares del quinto quintil (3,9%), y un aspecto socioeconómico muy relevante, es que tienen la mayor carga financiera entre todos los clusters debido a deudas de tarjetas de casas comerciales (32,8%, casi el doble que la media total), que como se señaló anteriormente, es una “deuda mala”, por sus altas tasas de interés, explicadas por sus escasas barreras de acceso (Castillo, 2011).

En términos sociodemográficos, los *Consumidores de subsistencia* son quienes tienen la mayor proporción de hogares que residen en una vivienda propia ya pagada (61,1%), y al mismo tiempo tienen la segunda proporción más baja de hogares que se encuentran pagando dividendo (7,5%).

Debido a la baja proporción de créditos hipotecarios que se encuentran en proceso de pago en este cluster, es que sus gastos en servicios financieros y seguros son los más bajos (\$1.596 promedio mensual).

Los *Consumidores de subsistencia* son uno de los clusters de mayor edad. La edad promedio de la jefatura de edad en este grupo es la segunda más alta, con 57,5 años, así como también se encuentra la segunda proporción más alta de hogares liderados por una persona de 65 años o más (34,5% de los hogares del grupo). En el mismo sentido, un 9,3% de los hogares están compuestos exclusivamente por adultos mayores. Por otra parte, en relación a la estructura de los hogares de este cluster, presenta la mayor cantidad de hogares extensos (31,8%), que incluyen a padre y/o madre y otros parientes dentro del hogar. En relación al ciclo familiar, los *Consumidores de subsistencia* tienen una alta participación de hogares en los últimos tramos del ciclo, con un 48,8% en desmembramiento y un 15,4% ya con el nido vacío.

Los *Consumidores de subsistencia* son quienes presentan niveles de educación formal más bajos entre todos los grupos, con una media de escolaridad de 9,49 años, es decir, más de 2 años menor que la cantidad de años de escolaridad obligatorios en la actualidad, que son los 12 años equivalentes a la educación media completa. Al analizar la distribución según mayor nivel educativo alcanzado por el sustentador o sustentadora principal del hogar, es posible observar que el 54,8% de estas personas no alcanza la educación media completa, lo que significaría para estos hogares ser considerados carentes según el Índice de Pobreza Multidimensional (IPM) actualmente vigente en Chile. Por otra parte, este cluster tiene la menor proporción de personas que asisten a la educación superior (3,97%) y de entre ellos un alto porcentaje asiste a la educación técnico profesional (46%).

En relación a su previsión en salud, el cluster de los *Consumidores de subsistencia* es uno de los dos con mayor participación en FONASA, con un 90,2%, a la vez que el menor porcentaje de personas con cobertura privada.

Ahora, yendo al detalle de los gastos de este grupo, es interesante conocer que a pesar de ser un cluster de ingresos bajos, dedican un mayor monto promedio mensual que los otros grupos a la compra de carbohidratos como pan (\$27.995), arroz (\$2.861), y pastas sin cocer (\$4.366), lo que indica que estos productos son importantes en la dieta diaria de este grupo y característicos de su estilo de vida. Los *Consumidores de subsistencia* tienen también un gasto mayor que los otros

clusters en carne (\$47.800 promedio mensual), pero al analizar el tipo de carne que consumen es posible observar que se trata de carne molida, distintos cortes de pollo, vienasas, patés y otros embutidos, que son las variedades más accesibles, y no son grandes consumidores de pescados y mariscos, típicamente más caros. También consumen más legumbres y hortalizas que el resto de los grupos, gastando \$35.871 promedio mensual en este ítem.

Por otra parte, los *Consumidores de subsistencia* son los que menos viajan en avión y los que menos presupuesto dedican a alojamientos y hospedajes, por lo que se puede suponer que, al contrario de los *Yuppies*, este es un cluster que viaja poco. Este cluster también presenta el mayor gasto relativo en cuidado personal (5% promedio de su gasto total), sin embargo, no se podría declarar que sea un grupo caracterizado por su preocupación por la salud y estética, dado que sus mayores gastos en este ítem corresponden a pañales y papel higiénico.

5.3.4.8 Cluster 8: Hogares de regiones

Finalmente, el cluster 8 de los *Hogares de regiones* está conformado por un 13,5% de los hogares de la muestra y fue necesario recurrir a sus características sociodemográficas para denominarlos, ya que su configuración de gastos por sí sola no permitía intuir una identidad adecuada para este cluster. Estos hogares presentan el gasto proporcional más elevado en vestuario y calzado (5,8%), en servicios de transporte (10,7%) y en comunicaciones (12%), y el segundo más alto en alojamiento. Al analizarlos a la luz de las variables suplementarias fue posible observar que este cluster no solo tiene más hogares *de regiones* que la media de los clusters, sino que también tiene más hogares *de regiones* que de la región Metropolitana, a pesar de ser mucho mayor la cobertura de la encuesta en esta última. Como se mencionó anteriormente, los ítems en que estos hogares gastan más proporcionalmente, se condicen con los ítems en que gastan más los hogares de regiones distintas de la metropolitana en la encuesta (taxi colectivo, electricidad, gas y otros combustibles, etc.).

En términos socioeconómicos, el cluster de los *Hogares de regiones* se encuentra entre los de menor ingreso, con una media de ingreso mensual de \$765.764, pero que no es significativamente diferente a la media de los ingresos mensuales de los *Agobiados por las cuentas*, los *Fumadores y bebedores*, y los *Consumidores de subsistencia*. Los *Hogares de regiones* son uno de los dos grupos que, junto con los *Fumadores y bebedores*, tienen un ingreso significativamente más alto

que su gasto. Un fenómeno llamativo es que a pesar de, aparentemente, no gastar más de lo que ganan, tienen una alta carga financiera debido a deudas por tarjetas de casas comerciales (28,6%).

De acuerdo a sus características sociodemográficas, lo que más destaca es la mayor proporción de hogares *de regiones* (52,4%), que supera casi 10 puntos el promedio de hogares *de regiones* en el resto de los grupos. Este cluster tiene la mayor proporción de hogares monoparentales (17,4%) y en relación al ciclo familiar, destaca la proporción de hogares en consolidación (16,4%) que son los que tienen hijos entre 13 y 18 años, y los hogares en desmembramiento (47%), con hijos de 18 años o más. Además, este cluster tiene la mayor proporción de jefatura de hogar femenina, con un 46%.

La escolaridad de los *Hogares de regiones* es más bien baja, con 10,9 años, se encuentra bajo la barrera de 12 años que impone el Estado para la educación obligatoria, aunque en este caso, a diferencia de los *Consumidores de subsistencia*, sólo el 35,1% de los hogares resultaría carente en educación, de acuerdo a los criterios del Índice de Pobreza Multidimensional. Por otra parte, en relación a su sistema de salud, los *Hogares de regiones* son quienes tienen un mayor porcentaje de jefatura de hogar o su pareja que están afiliados al sistema público de salud, con un 90,6%.

Ahora, entre otros atributos que caracterizan el estilo de vida de los *Hogares de regiones* se encuentra su alto gasto promedio mensual en taxi colectivo (\$13.450) y en microbús (\$11.927). Estos medios de transporte no son especialmente importantes en la región metropolitana, pero en regiones son cotidianamente utilizados para el desplazamiento. Por otra parte, si bien se registra un gasto proporcional importante en telecomunicaciones, el más alto de todos los grupos, al analizar el monto dedicado a este ítem es posible observar que no es mayor que el que gastan los *Conductores* o los *Yuppies*. Sin embargo, a la luz del menor ingreso de este cluster, si se vuelve una pista importante sobre las preferencias asociadas al estilo de vida de los *Hogares de regiones*. Esto se relaciona además con que este cluster gasta cerca del doble que el cluster que le sigue en aparatos celulares, lo que indica de manera aún más clara su preferencia por este tipo de tecnología. Otro aspecto importante de los *Hogares de regiones* es que salen poco a Restaurantes, cafés y similares, y cuando viajan lo hacen de forma austera, prefiriendo el transporte por carretera frente al transporte aéreo y reduciendo el pago en alojamiento, que es lo que se puede observar a través de sus gastos en dichos ítems.

6. Conclusiones

En este último capítulo se presentan las conclusiones obtenidas del proceso investigativo de este trabajo de tesis, ofreciéndose reflexiones en cuatro niveles: i) metodológicas, que se derivaron del estudio de las técnicas de análisis, pero sobre todo del ensayo y error con los datos utilizados; ii) sustantivas, referidas a los resultados obtenidos; iii) sobre el carácter construido de la clasificación; y sobre iii) potenciales líneas investigativas que se visualizan a partir de este trabajo.

6.1 Reflexiones metodológicas

A partir del proceso de exploración del coeficiente *Average Silhouette Width* utilizado para el estudio del ajuste matemático¹⁸ del modelo de K-medias implementado, es posible sostener que existe una tensión entre este ajuste matemático y la agrupación obtenida, en términos sustantivos.

Se observó que el ASW disminuye paulatinamente (es decir, va empeorando) cada vez que se incluye una variable o componente nuevo al modelo, dado que, en general, al incluir mayor variabilidad incrementan las distancias euclídeas entre las observaciones y la clasificación genera mayores residuos.

Por otra parte, también es importante considerar cuál es el objetivo de la investigación y el tipo de resultado esperado al valorar el ajuste del ASW. Para cierto tipo de datos (especialmente

¹⁸ El ajuste matemático refiere a la minimización de la suma cuadrática del error intra estrato.

cuando son variables continuas con rangos amplios), matemáticamente la solución más óptima se encuentra al dividir a las observaciones en 2 clusters, sin embargo, esto puede resultar en una clasificación de perogrullo o sin un valor analítico, por ejemplo: ricos y pobres.

Para sortear estas dificultades se sugiere:

- a) Valorar el coeficiente ASW o cualquier otro coeficiente que se utilice para encontrar un valor óptimo del parámetro k , entendiéndolo como una referencia y no como una recomendación determinística.
- b) Comparar los resultados obtenidos para el ASW con los de investigaciones que hayan usado datos de características similares (cantidad y tipo de variables, distribución de las variables, etc.), pues este coeficiente presenta valores más elevados en sets de variables dicotómicas, categóricas u ordinales (como puede observarse en las investigaciones de Madero-Cabib y Fasang [2016] y Madero-Cabib, Gauthier y Le Goff [2016]), que en sets de variables continuas (lo que puede apreciarse en Degirmenci y Özbakir [2018]). Entonces, comparar los coeficientes ASW de investigaciones que utilizan datos con diferentes características puede llevar a una subestimación o sobre estimación del éxito del ajuste.
- c) Por último, siempre debe primar la observación de estadísticos descriptivos de las variables activas para asegurar que haya cabida para una interpretación conceptualmente interesante de los resultados de la clasificación.

6.2 Reflexiones sobre los estilos de vida

6.2.1 Sobre la flexibilidad de los estilos de vida

Bourdieu popularizó el estudio de los campos mediante representaciones, de alguna manera reproduciendo estos campos en espacios donde sus elementos (individuos, consumos culturales, prácticas, estilos de vida, etc.) se distribuyen de acuerdo a su relación con ejes analíticos, generando posiciones observables. En cierta forma este trabajo buscó algo similar: construir un campo analítico que está determinado en estricto por el gasto de consumo, generando a su vez posiciones en el espacio multidimensional del consumo que son enlazadas por esta particular concepción del estilo de vida. En ese sentido, es necesario valorar los resultados de este ejercicio en cuanto a su diseño y ejecución: estos estilos de vida son ciegos a prácticas distintas al gasto de

consumo y a las actitudes o valores con que las personas respaldan su performance en la sociedad de consumo. Sin embargo, resulta en un ejercicio muy interesante en cuanto a que estos estilos de vida son recogidos de un rastro objetivo que dejan los hogares, del cual muchas veces ni siquiera son conscientes. Dado eso, el ejercicio de autoclasificación podría concluir en un resultado imprevisto.

Siguiendo esa idea, esta clasificación invita a preguntarse: ¿los estilos de vida a partir de los gastos de consumo son una condición estable o una posición, en el sentido de un estado temporal, que permite fluir entre diferentes estilos de vida?

Se sostiene que los estilos de vida contruidos a partir de los gastos de consumo son más bien una posición fluctuante, que dependerá de muchos factores, internos y externos, entre los cuales quizás la variable que mejor resume estos cambios es el ciclo familiar, aunque hay otras variables que son importantes para grupos en específico. El ciclo familiar se utilizó como variable suplementaria en este trabajo, y resume aspectos como la presencia de hijos en el hogar y sus edades y la edad de la pareja que compone el núcleo familiar. Como se expuso en el capítulo de resultados, la presencia de adultos mayores en el hogar vuelve más probable formar parte del cluster de los *Carentes en salud*, así como la presencia de miembros del hogar en edad de asistir a la educación superior aumenta la probabilidad de formar parte del cluster de los *Hogares con universitarios*. Para ciertos clusters otras variables aparecen como importantes, como el caso de los *Agobiados por las cuentas* en relación al tipo de tenencia de la vivienda: el solo hecho de que un hogar deje de ser arrendatario vuelve muy probable que este transite a otra categoría de estilo de vida.

En ese sentido, existen variables sociodemográficas y socioeconómicas que presentan una rigidez más estructural para el tránsito de un estilo de vida a otro, y otras variables más coyunturales que permiten un tránsito más fluido. Se considera dentro de las primeras el ingreso disponible del hogar, dado que está muy relacionado con una posición en el mercado del trabajo (posición de clase), que es generalmente estable. Ciertos estilos de vida son más probables en ciertos estratos, por ejemplo, es más probable que un hogar pobre sea parte de los *Consumidores de subsistencia* que de los *Yuppies*, dado que los segundos requieren de cierto ingreso para poder optar por ese estilo de vida y que en los primeros, al tener poco ingreso, el gasto en alimentación (que es ineludible) se vuelve muy importante. También es más estructural la edad de los miembros del

hogar. Una vez que un hogar alcanza la adultez mayor, es muy probable que sus gastos en educación, vestuario y transporte disminuyan y sus gastos en salud aumenten (Hayes y Finney, 2014).

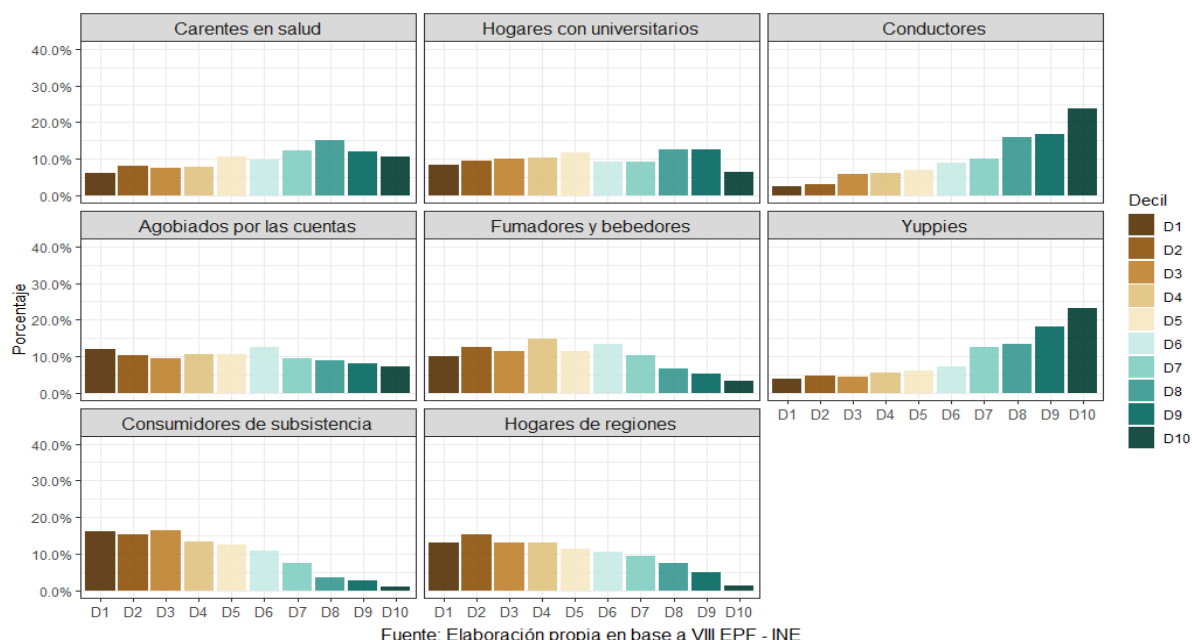
Ahora, son menos rígidas las variables de presencia o ausencia de miembros del hogar en edad escolar o de asistir a la universidad, el pago o no pago de arriendo, o el mismo hecho de fumar o beber alcohol, entre otras, las cuales o son temporales, o son situaciones que podrían cambiar. En fin, estos estilos de vida se relacionan con fases del ciclo vital y familiar de los hogares, por lo que si bien un mismo hogar difícilmente podría transitar por todos, si es altamente probable que pase por varios de ellos y muy poco probable que permanezca en una misma categoría durante todo su ciclo.

6.2.2 Estilos de vida y estratificación socioeconómica

La hipótesis de esta investigación sugiere que es posible identificar distintos estilos de vida, relacionados con aspectos materiales (ingresos, *shocks* externos como una enfermedad, etc.) y también con gustos y preferencias, que se verán influenciados por variados sistemas de valores y *habitus*, y que estos estilos de vida exceden la distinción según estrato socioeconómico.

A través del análisis de la distribución por quintil de los clusters expuesta en el gráfico 4, fue posible advertir que si bien hay una relación entre ingreso y estilo de vida, esta no es determinante, puesto que en todos los clusters hay hogares de los 5 quintiles, incluso en los que tienen medias de ingreso muy altas (*Conductores, Yuppies*) o muy bajas (*Consumidores de subsistencia*). Ahora, al abrir más esta distribución, a nivel de decil, se puede observar que en los clusters de mayor ingreso quienes están en el quintil más rico tienden a ser mayoritariamente del decil más rico también, fenómeno que no se refleja de la misma manera en el primer quintil de los cluster de menor ingreso, los que se dividen de forma pareja entre el primer y segundo decil. Aun así, el gráfico 11 permite confirmar la hipótesis de que los estilos de vida construidos no son un espejo de la estratificación socioeconómica.

Gráfico 11: Distribución de hogares en cada cluster por decil de ingreso



6.3 Sobre el carácter construido de la clasificación

Como se puede desprender de las reflexiones anteriores y retomando los conceptos revisados en el apartado 5.1 sobre la construcción de una clasificación, es posible sostener que las clasificaciones en ningún caso son objetivas, y que derivan de un conjunto de decisiones, basadas en marcos conceptuales, preconcepciones e incluso intuiciones de quien realiza una clasificación.

Bajo el mismo título de esta investigación: “Estilos de vida a partir del consumo en Chile”, se podría haber llegado a un resultado completamente diferente. Se podría pensar, por ejemplo, ¿qué grupos se habrían obtenido si en lugar de utilizar todos los gastos reportados se hubieran seleccionado algunos por considerarse más simbólicos?, ¿o si solo se hubieran incluido los consumos culturales? Por otra parte, ¿qué hubiera resultado de utilizar como variables activas montos absolutos (como el gasto total por ítem) en lugar de gastos proporcionales?, ¿o si se hubiera escogido un método de *clustering* diferente?

Si bien pareciera que el proceso de etiquetar a los clusters con nombres significativos (*Carentes en salud*, *Agobiados por las cuentas*, *Yuppies*, etc.) es el más arbitrario, no lo es mucho más que el resto de las decisiones adoptadas durante el trazado investigativo. Es por esto que el primer objetivo específico de esta tesis buscaba desarrollar una propuesta metodológica adecuada para el

problema de clasificación propuesto. En ese sentido, se enfrentó la tarea de crear una clasificación, en sí misma, como la tarea de construir un objeto de estudio. Si bien clasificar en un “arte” y no existe una única o mejor forma para hacerlo, en esta tesis se intentó al menos dejar escrita la receta para lograr esta paleta de estilos de vida de los chilenos, y así facilitar la valoración de las decisiones tomadas por parte de otros investigadores que pudieran estar interesados en el tema.

6.4 Potenciales líneas de investigación

Bourdieu (1998) señala que el uso de datos secundarios, como las encuestas, presenta limitaciones para el estudio de los estilos de vida que no pueden ser sobrepasadas. El autor menciona a una encuesta realizada por el INSEE (*Institut National de la statistique et des études économiques*) en la cual se busca conocer las intenciones de compra de las personas. Sin embargo, deja ver poca información sobre la frecuencia y ocasión de los consumos, o sobre la calidad de los objetos consumidos, lo que en algunos casos constituye toda la diferencia; Por ejemplo, una sola categoría “teatro” cubre tanto el teatro de vanguardia como el teatro boulevard, o en relación a la asistencia a museos y exposiciones, Bourdieu indica que estas categorías son agregadas, mientras que la frecuentación a museos corresponde a una práctica más ascética, y es más común entre los profesores (según datos de otra fuente del INSEE), y la asistencia a exposiciones la define como más “mundana” y común entre las profesiones liberales y la vieja burguesía patronal (Bourdieu, 1998). Esto también ocurre cuando se utiliza la EPF como fuente secundaria de datos. Por ejemplo, para los objetivos de la encuesta no es importante distinguir entre productos *light* y *no light*, y consecuentemente el clasificador no los distingue, pero sí distingue entre tipos de pan y entre prendas de vestir de hombre y de mujer. Sin embargo, a pesar de estas limitaciones que es necesario tener en cuenta, la EPF ofrece muchas posibilidades de análisis de los estilos de vida en diversas áreas a un nivel muy desagregado (1.186 tipos de productos). En esta tesis se optó por hacer una exploración general analizando el gasto en todos los ítems reportados, no obstante, es posible delinear algunas vetas de investigación que podrían beneficiarse tanto del trazado metodológico propuesto para esta tesis, como de la intuición de analizar aspectos de los estilos de vida de las personas a partir de sus gastos de consumo. Por ejemplo:

- a) **Estilos de vida y alimentación saludable:** como se mencionó en el párrafo anterior, la EPF no distingue entre alimentos *light* o no *light*, sin embargo, es posible hacer un *match* razonablemente exitoso entre el clasificador de alimentos utilizado por la Encuesta Nacional de Consumo Alimentario (ENCA) y la clasificación utilizada en la EPF (CCIF). Esto permitiría atribuir a cada producto alimentario capturado por la EPF una valoración sobre cuán saludables son y así realizar un ejercicio clasificatorio similar al expuesto en esta tesis, pero solo considerando el consumo de productos alimentarios y así caracterizar los estilos de alimentación de los hogares en Chile utilizando conceptos de salud y nutrición.
- b) **Estilos de vida y consumo cultural:** es posible estudiar las prácticas de consumo cultural de los hogares en Chile derivando la presencia de ciertos hábitos cuando registran gastos a causa de ellos. Por ejemplo, si un hogar gastó en libros, se podría considerar que dicho hogar practica la lectura; lo mismo se puede suponer en relación a la asistencia al cine, teatro, conciertos, fotografía (hay un código de gasto específico para “artículos de fotografía”), etc. Investigaciones como la de Bourdieu en “La distinción” (1998) y la de Gayo, Teitelboim y Méndez (2009) utilizan instrumentos que consultan directamente sobre las prácticas y hábitos de consumo cultural, sin embargo, a partir de las encuestas de gastos se pueden alcanzar *proxies* interesantes, y aprovechando información que ya fue producida y se encuentra disponible.
- c) **Otras investigaciones:** una estrategia habitual en la investigación con *clustering* es generar una variable de clasificación para luego usarla como predictor en el estudio de otros fenómenos, como el ahorro en el estudio de Keister *et al.* (2016). Esta clasificación resume información sobre los gastos en una sola variable que puede ser utilizada para predecir fenómenos como el endeudamiento, entre otros posibles de analizar a partir de los datos de la EPF.

7. Bibliografía

Alguacil, J. (2000). *Calidad de vida y praxis urbana: Nuevas iniciativas de gestión ciudadana en la periferia social de Madrid*. Madrid, España: Centro de Investigaciones Sociológicas.

Alonso, L. E. (2009). *Prácticas económicas y economía de las prácticas*. Madrid: Catarata.

Amat, J. (2017). *Clustering y heatmaps: aprendizaje no supervisado*. RPub de RStudio. Obtenido el 15-04-2019. Recuperado de : https://rpubs.com/Joaquin_AR/310338

Anderloni, L. & Vardone, D. (2008). *Households over-indebtedness in the economic literature*. Working Paper, N° 46, Dipartimento di Scienze Economiche Aziendali e Statistiche. University of Milan.

Araujo, K., & Martuccelli, D. (2012). *Desafíos comunes: Retrato de la sociedad chilena* (Tomo I y II). Santiago: LOM.

Arriagada, I. (1997). *Políticas sociales, familia y trabajo en la América Latina de fin de siglo*. Serie Políticas Sociales N°21. Publicaciones CEPAL. Santiago, Chile.

Arroyo, C., (15 de diciembre de 2017). *Interés de tarjetas y líneas de crédito explica alta carga financiera en grupos medios*. Recuperado de: <https://www.latercera.com/noticia/interes-tarjetas-lineas-credito-explica-alta-carga-financiera-grupos-medios/>

Azocar, C., & Mayol, A. (2011). *Politización del malestar, movilización social y transformación ideológica: el caso "Chile 2011"*. Polis (30), 1-20.

Banco Central de Chile (2018). Encuesta Financiera de Hogares 2017: principales resultados. Santiago: Banco Central.

Bank for International Settlements [BIS] (2014). *84° Informe Anual*. Obtenido de http://www.bis.org/publ/arpdf/ar2014_es.htm

Baudrillard, J. (1991). *Crítica de la economía política del signo*. Madrid: Siglo XXI.

Bauman, Z. (2011). *Vida de Consumo*. Buenos Aires: Fondo de Cultura Económica.

Berzal, F. (2019). *Clustering Jerárquico*. Departamento de Ciencias de la Computación e I.A. Obtenido el 15-04-2019. Recuperado de: <https://elvex.ugr.es/idbis/dm/slides/42%20Clustering%20-%20Hierarchical.pdf>

Bourdieu, P. (1998). *La distinción: Criterio y bases sociales del gusto*. Madrid: Taurus.

Bourdieu, P., Passeron, J. C. (1977). *Reproduction in Education, Society, and Culture*. Traducido por R. Nice. Beverly Hills, CA: Sage.

Bourdieu, P. y Wacquant, L. (2008). *Una invitación a la sociología reflexiva*. Buenos Aires: Siglo XXI.

Burnett, J. & Bush, A. (1986). *Profiling the Yuppies*. Journal of advertising research. Nº 26 (Abril - Mayo 1986): 27 – 35.

Castillo, A. (25 de julio de 2011). *Créditos de casas comerciales*. Recuperado de: <https://noticias.uai.cl/columna/creditos-casas-comerciales/>

Catalán, C. (2009). *Consumo y segmentación: algunas consideraciones conceptuales y empíricas*. En: *El arte de clasificar a los chilenos: enfoque sobre modelos de estratificación en Chile*. Santiago: UDP, pp.37-60.

Comisión para la Medición de la Pobreza [CMP] (2014). Informe Final Comisión para la Medición de la Pobreza. Santiago.

Consejo Nacional de la Cultura y las Artes [CNCA] (2017). *Encuesta Nacional de Participación Cultural 2017*. Santiago: CNCA.

Cox, P., Parrado, E., y Ruiz-Tagle, J. (2007). *The Distribution of Assets, Debt and Income among Chilean Households*. Proceedings of the IFC Conference on Measuring the financial position of the household sector, BIS, Basilea, 30 y 31 de Agosto 2006 – Vol 2.

Degirmenci, Tugba, Özbakir, Lale (2018). *Differentiating households to analyze consumption patterns: a data mining study on official household budget survey*. WIREs Data Mining Knowledge Discovery 2018, 8:e1227. doi: 10.1002/widm.1227.

Ding, C., & He, X. (2004). Principal Component Analysis and Effective K-means Clustering. *4th SIAM International Conference on Data Mining*, (p. 5). Lake Buena Vista, FL.

Douglas, M., Isherwood, B. (1990). *El mundo de los bienes*. México: Grijalbo.

Doyal, L.; Gough, I. (1994). *Teoría de las necesidades humanas*. Madrid: Icaria/FUHEM.

Durkheim, E. y Mauss, M. (1963). *Primitive Classification*. USA: The University of Chicago Press.

Eurostat (2003). *Household Budget Surveys in the EU. Methodology and recommendations for harmonisation*. Obtenido de <http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-BF-03-003--N-EN.pdf>

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis (5th Edition)*. Wiley.

Fernandez-Villaverde, J., Krueger, D. (2007). *Consumption over the life cycle: facts from consumer Expenditure survey data*. The Review of Economics and Statistics, Vol. 89, pp. 552–565.

Finlay, S. (2006). *Predictive Models of Expenditure and Over-Indebtedness for Assessing the Affordability of New Consumer Credit Applications*. The Journal of the Operational Research Society, Vol. 57, N° 6, pp. 655-669.

Galtung, J. (1977). El desarrollo, el medio ambiente y la tecnología, hacía una tecnología autónoma. Ginebra: Programa de las Naciones Unidas para el Medio Ambiente (PNUMA).

Gayo, M., Teitelboim, B., y Méndez, M. L. (2009). *Patrones culturales de uso del tiempo libre en Chile. Una aproximación desde la teoría bourdieuana*. Universum 24-2, pp. 42-72.

GfK Chile (2019). *Estilos de vida de los grupos socioeconómicos de Chile*. Obtenido de https://www.gfk.com/fileadmin/user_upload/country_one_pager/CL/GfK_GSE_190502_FINAL.pdf

Gorenshteyn, D. (2018). *Cluster Analysis in R*. Curso E-Learning disponible en la plataforma de aprendizaje: www.datacamp.com

Joignant, A., Güell, P. (2009). *Organizando la heterogeneidad: clasificaciones, taxonomías y convenciones sobre los chilenos y sus diferencias*. En: *El arte de clasificar a los chilenos: enfoque sobre modelos de estratificación en Chile*. Santiago: UDP, pp.37-60.

Hadi, Ali (1992). *Identifying Multiple Outliers in Multivariate Data*. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 54, N° 3, pp. 761-771. Wiley.

Hayes, D., Finney, A. (2014). *Dominant patterns of expenditure among older people in the United Kingdom: Segmenting the older consumer using the Living Costs and Food Survey*. Journal of Population Ageing, Vol. 7, pp. 97–113.

- Heller, A. (1978). *La teoría de las necesidades en Marx*. Madrid: Península.
- Herrera, S. (2006). Sobre las formas de clasificación en Durkheim y Bourdieu. Voces y contextos. Otoño, Vol. 2, pp. 1 – 18.
- Instituto Nacional de Estadística y Geografía de México [INEGI] (2010). *Nota Técnica: Estratificación Multivariada*. México: INEGI.
- Instituto Nacional de Estadísticas [INE Chile] (2018a). *Metodología VIII Encuesta de Presupuestos Familiares*. Instituto Nacional de Estadísticas de Chile. Obtenido de <http://www.ine.cl/docs/default-source/ingresos-y-gastos/epf/viii-epf/documentacion/metodolog%C3%ADa-viii-epf.pdf?sfvrsn=4>
- _____ (2018b). *Informe de Principales Resultados VIII Encuesta de Presupuestos Familiares (EPF)*. Obtenido de <http://www.ine.cl/docs/default-source/ingresos-y-gastos/epf/viii-epf/principales-resultados/informe-de-principales-resultados-viii-epf.pdf?sfvrsn=8>
- Izquierdo, A. I. (1997). *Necesidades, utopía y revolución en Ágnes Heller*. La Ventana (5), 204-220.
- Keister, L., Benton, R., Moody, J. (2016). *Lifestyles through expenditures: A case-based approach to saving*. Sociological Science 3: 650-684.
- Larraín, J. (2001). *Identidad chilena*. Santiago de Chile, LOM.
- Lê, S., Josse, J., Husson, F. (2008). *FactoMineR: An R Package for Multivariate Analysis*. Journal of Statistics Software, Vol. 25, pp. 1 – 18.
- Lipovetsky, G. (2003). *La era del vacío*. Barcelona: Anagrama.
- Madero-Cabib, I. Fasang, A- (2016). *Gendered work–family life courses and financial well-being in retirement*. Advances in Life Course Research, Vol. 27, pp. 43–60.
- Madero-Cabib, I., Gauthier, J. A., Le Goff, J. M. (2016). *The Influence of Interlocked Employment–Family Trajectories on Retirement Timing*. Work, Aging and Retirement, Vol. 2, N° 1, pp. 38–53.b
- Marinas, J. M. (2001). *La fábula del bazar: Orígenes de la cultura del consumo*. Madrid: Antonio Machado Libros.
- Matus, J.M., Silva, N., Marinovic, A. y Flores, K (2010). *Una visión global de la deuda financiera de los hogares chilenos en la última década*. Estudios Económicos Estadísticos N° 81, Banco Central de Chile.
- Max-Neef, Manfred (1998). *Desarrollo a Escala Humana. Conceptos, aplicaciones y algunas reflexiones*. Barcelona: Icaria.
- Mayol, A. Azocar, C. y Azocar, C. (2013). *El Chile Profundo. Modelos culturales de la desigualdad y sus resistencias*. Santiago: Liberalia Ediciones.

Ministerio de Desarrollo Social de Chile [MDS] (2018a). *Ingresos de los hogares: síntesis de resultados*. Obtenido de: http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Resultados_ingresos_Casen_2017.pdf

_____ (2018b). *Canasta básica de alimentos. Informes históricos*. Obtenido de: http://observatorio.ministeriodesarrollosocial.gob.cl/ipc_pob_informe.php

_____ (2002). *Chile Actual: Anatomía de un mito*. Santiago: LOM Ediciones.

Organización de Naciones Unidas [ONU] (2001). *Clasificación de Gastos por Finalidades*. Informes Estadísticos, Naciones Unidas, Departamento de Asuntos Económicos y Sociales, División de Estadística, Nueva York. Obtenido de <https://unstats.un.org/unsd/cr/registry/regdnld.asp?Lg=3>

Organización de Naciones Unidas [ONU] (2019). *Objetivos de Desarrollo Sostenible*. 24-04-2019, de United Nations Sitio web: <https://www.un.org/sustainabledevelopment/>

Organización para la Cooperación y el Desarrollo Económico [OCDE] (2018). *OECD Social and Welfare Statistics: Income distribution*. Obtenido de: <https://data.oecd.org/inequality/income-inequality.htm>

Programa de Naciones Unidas para el Desarrollo [PNUD] (1998). *Desarrollo humano en Chile. Las paradojas de la modernización*. Santiago: PNUD.

Programa de Naciones Unidas para el Desarrollo [PNUD] (2002). *Desarrollo humano en Chile. Nosotros los chilenos: un desafío cultural*. Santiago: PNUD.

Roark , H. (2018). *Unsupervised Learning in R*. Curso E-Learning disponible en la plataforma de aprendizaje: www.datacamp.com

Simmel, G. (2002). *Sobre la individualidad y las formas sociales: escritos escogidos*. Buenos Aires: Prometeo Libros.

Superintendencia de Bancos e Instituciones Financieras [SBIF] (2017). *Informe de Endeudamiento 2017*. Santiago: SBIF.

_____ (2018). *Informe de Endeudamiento 2018*. Santiago: SBIF.

Veblen, T. (1963). *Teoría de la clase ociosa*. México: Fondo de Cultura Económica.

Weber, M (2002). *Economía y Sociedad*. Fondo de Cultura Económica, México.

8. Anexos

Anexo 1: división y grupo de la Clasificación del Consumo Individual por Finalidades

Se expone a continuación la Clasificación del Consumo Individual por Finalidades (CCIF), a nivel de división y grupo (omitiendo clase, subclase y producto), con cada código y su glosa asociada.

CCIF	GLOSA
01.0.0.00.00	ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS
01.1.0.00.00	ALIMENTOS
01.2.0.00.00	BEBIDAS NO ALCOHÓLICAS
01.3.0.00.00	GASTOS NO DESGLOSADOS EN ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS
02.0.0.00.00	BEBIDAS ALCOHÓLICAS, TABACO Y ESTUPEFACIENTES
02.1.0.00.00	BEBIDAS ALCOHÓLICAS
02.2.0.00.00	TABACO
02.3.0.00.00	ESTUPEFACIENTES
02.4.0.00.00	GASTOS NO DESGLOSADOS EN BEBIDAS ALCOHÓLICAS, TABACO Y ESTUPEFACIENTES
03.0.0.00.00	PRENDAS DE VESTIR Y CALZADO
03.1.0.00.00	PRENDAS DE VESTIR

03.2.0.00.00	CALZADO
03.3.0.00.00	GASTOS NO DESGLOSADOS EN PRENDAS DE VESTIR Y CALZADO
04.0.0.00.00	ALOJAMIENTO, AGUA, ELECTRICIDAD, GAS Y OTROS COMBUSTIBLES
04.1.0.00.00	ARRIENDOS EFECTIVOS DEL ALOJAMIENTO
04.2.0.00.00	ARRIENDOS IMPUTADOS DEL ALOJAMIENTO
04.3.0.00.00	CONSERVACIÓN Y REPARACIÓN DE LA VIVIENDA
04.4.0.00.00	SUMINISTRO DE AGUA Y SERVICIOS RELACIONADOS CON LA VIVIENDA
04.5.0.00.00	ELECTRICIDAD, GAS Y OTROS COMBUSTIBLES
04.6.0.00.00	GASTOS NO DESGLOSADOS EN ALOJAMIENTO, AGUA, ELECTRICIDAD, GAS Y OTROS COMBUSTIBLES
05.0.0.00.00	MUEBLES, ARTÍCULOS PARA EL HOGAR Y PARA LA CONSERVACIÓN ORDINARIA DEL HOGAR
05.1.0.00.00	MUEBLES Y ACCESORIOS, ALFOMBRAS Y OTROS MATERIALES PARA PISOS
05.2.0.00.00	PRODUCTOS TEXTILES PARA EL HOGAR
05.3.0.00.00	ARTEFACTOS PARA EL HOGAR
05.4.0.00.00	ARTÍCULOS DE VIDRIO Y CRISTAL, VAJILLA Y UTENSILIOS PARA EL HOGAR
05.5.0.00.00	HERRAMIENTAS Y EQUIPOS PARA EL HOGAR Y EL JARDÍN
05.6.0.00.00	BIENES Y SERVICIOS PARA LA CONSERVACIÓN ORDINARIA DEL HOGAR
05.7.0.00.00	GASTOS NO DESGLOSADOS EN MUEBLES, ARTÍCULOS PARA EL HOGAR Y PARA LA CONSERVACIÓN ORDINARIA DEL HOGAR
06.0.0.00.00	SALUD
06.1.0.00.00	PRODUCTOS, ARTEFACTOS Y EQUIPOS MÉDICOS
06.2.0.00.00	SERVICIOS PARA PACIENTES EXTERNOS
06.3.0.00.00	SERVICIOS DE HOSPITAL
06.4.0.00.00	GASTOS NO DESGLOSADOS EN SALUD
07.0.0.00.00	TRANSPORTE
07.1.0.00.00	ADQUISICIÓN DE VEHÍCULOS
07.2.0.00.00	FUNCIONAMIENTO DE EQUIPO DE TRANSPORTE PERSONAL
07.3.0.00.00	SERVICIOS DE TRANSPORTE
07.4.0.00.00	GASTOS NO DESGLOSADOS EN TRANSPORTE
08.0.0.00.00	COMUNICACIONES
08.1.0.00.00	SERVICIOS POSTALES
08.2.0.00.00	EQUIPO TELEFÓNICO Y DE FACSIMILE
08.3.0.00.00	SERVICIOS TELEFÓNICOS Y DE FACSIMILE
08.4.0.00.00	GASTOS NO DESGLOSADOS EN COMUNICACIONES
09.0.0.00.00	RECREACIÓN Y CULTURA
09.1.0.00.00	EQUIPO AUDIOVISUAL, FOTOGRÁFICO Y DE PROCESAMIENTO DE INFORMACIÓN
09.2.0.00.00	OTROS BIENES DURABLES PARA EL OCIO Y LA CULTURA
09.3.0.00.00	OTROS ARTÍCULOS Y EQUIPO PARA RECREACIÓN, JARDINES Y ANIMALES DOMÉSTICOS
09.4.0.00.00	SERVICIOS DE RECREACIÓN Y CULTURALES
09.5.0.00.00	PERIÓDICOS, LIBROS Y PAPELES, Y ÚTILES DE OFICINA
09.6.0.00.00	PAQUETES TURÍSTICOS
09.7.0.00.00	GASTOS NO DESGLOSADOS EN RECREACIÓN Y CULTURA

10.0.0.00.00 EDUCACIÓN
10.1.0.00.00 ENSEÑANZA PREESCOLAR Y BÁSICA
10.2.0.00.00 ENSEÑANZA SECUNDARIA
10.3.0.00.00 ENSEÑANZA POSTSECUNDARIA NO TERCARIA
10.4.0.00.00 ENSEÑANZA TERCARIA
10.5.0.00.00 ENSEÑANZA NO ATRIBUIBLE A NINGÚN NIVEL
10.6.0.00.00 GASTOS NO DESGLOSADOS EN EDUCACIÓN
11.0.0.00.00 RESTAURANTES Y HOTELES
11.1.0.00.00 SERVICIOS DE SUMINISTRO DE COMIDAS POR CONTRATO
11.2.0.00.00 SERVICIOS DE ALOJAMIENTO
11.3.0.00.00 GASTOS NO DESGLOSADOS EN RESTAURANTES Y HOTELES
12.0.0.00.00 BIENES Y SERVICIOS DIVERSOS
12.1.0.00.00 CUIDADO PERSONAL
12.2.0.00.00 PROSTITUCIÓN
12.3.0.00.00 EFECTOS PERSONALES N.C.P.
12.4.0.00.00 PROTECCIÓN SOCIAL
12.5.0.00.00 SEGUROS
12.6.0.00.00 SERVICIOS FINANCIEROS N.C.P.
12.7.0.00.00 OTROS SERVICIOS N.C.P.
12.8.0.00.00 GASTOS NO DESGLOSADOS EN BIENES Y SERVICIOS DIVERSOS

Fuente: Elaboración propia en base a VIII EPF-INE.

Anexo 2: detección y depuración de *outliers* o casos atípicos

Detección de *outliers*

Para la detección de *outliers* se utilizó una mezcla de dos técnicas. Para todas las variables de gasto, salvo la división 1 (“Alimentos y bebidas no alcohólicas”), se implementó el método Hadi, mientras que para tratar la división 1 se utilizó la metodología de Rango intercuartíl.

El método de Hadi ofrece la posibilidad de realizar detección de casos atípicos en más de una variable, aunque en este caso sólo se aplicaron cálculos univariados. Hadi opera de la siguiente manera (Hadi, 1992): en primer lugar, toma una porción de un vector numérico (cada variable de gasto), y calcula distancias con un algoritmo similar a la distancia de Mahalanobis, utilizando alguna medida de tendencia central (por defecto la media), y en base a estas distancias ordena los casos; luego, a través de un proceso iterativo, va agregando las observaciones una a una y repitiendo el primer paso hasta alcanzar la última observación del vector. Finalmente, una vez encontrado el punto de detención, identifica como *outliers* a los casos con distancias más altas, que superan cierto umbral.

Dado que Hadi ordena de forma ascendente, mediante estimadores robustos de posición y escala, se vuelve altamente probable que *outliers* aparezcan hacia el final de los datos y es altamente improbable encontrarlos al principio (Hadi, 1992).

El otro método utilizado fue el del rango intercuartílico, solamente sobre la división 1, de alimentación y bebidas no alcohólicas. Hadi es altamente sensible a la cantidad de casos que presentan un valor distinto de cero. Cuando hay una gran cantidad de valores iguales a cero, generalmente Hadi no puede operar y arroja error. En el otro sentido, cuando los valores se distribuyen por el continuo sin tener una moda importante, Hadi tiende a no identificar *outliers*, incluso habiendo observaciones que declararon dedicar más del 95% de su gasto total en esta división.

El rango intercuartílico es la diferencia entre el tercer y el primer cuartil, es decir, entre el percentil 75 y 25 de una distribución. Utilizando esta medida como unidad es posible establecer el límite de lo que será considerado atípico rigidizando o relajando la cantidad de rangos intercuartílicos utilizados, lo que dependerá del objetivo del estudio y del criterio de quien realiza

el análisis. En este caso, fue necesario relajar bastante la medida, dado que importaba detectar y tratar a los casos realmente extremos en el ala superior, para que no afectaran la clasificación. Se comenzó la revisión con un rango intercuartílico de 1,5 y finalmente se determinó utilizar uno de 2,5.

Depuración de *outliers*

Antes de tratar los *outliers* detectados, se realizó una serie de ajustes en el sentido de corregir algunas situaciones que podrían generar ruido en la clasificación.

En primer lugar, se eliminaron 2 hogares (de los 15.239 presentes en la muestra), que declaraban un gasto total igual a cero. En segundo lugar, se asumió un criterio normativo para eliminar hogares que declararon un gasto total por debajo del valor de la canasta básica alimentaria (CBA) deflactada a julio de 2016, que es el mes inicial del trabajo de campo de la VIII EPF.

Los métodos de detección de valores atípicos, no solo Hadi, tienen problemas para captar outliers en la cola inferior de la distribución. La desviación estándar del gasto mensual total por hogar es de \$1.090.841, para una media de \$1.122.399, por lo que la desviación casi alcanza el cero. Ante esto, fue necesario asumir un criterio normativo para depurar casos que presentaban un gasto muy bajo, asumiendo que este tipo de encuestas sufren de subdeclaración y lo más probable es que dicho gasto sea más fruto de un reporte de mala calidad que de un gasto efectivamente muy bajo.

Para este procedimiento se asumió un umbral conservador, eliminándose de la muestra a los hogares que presentaban un gasto total per cápita menor a la canasta básica alimentaria deflactada a julio de 2016, avaluada en \$40.118 (MDS, 2018a). Se define el umbral como conservador al considerar que actualmente la CBA ya no es equivalente siquiera a la pobreza extrema. Para el mismo mes de referencia la línea de pobreza por persona equivalente era de \$153.952 y la de pobreza extrema por persona equivalente de \$102.635, por lo que los hogares eliminados no deberían haber estado en condiciones de asegurar su alimentación de ser certera su declaración. Con esta depuración se eliminan 157 hogares.

Se evaluó también eliminar a los hogares con gastos totales muy elevados. Fueron revisados y en general se trataba de hogares con un gasto muy elevado en salud, por hospitalizaciones. Se decidió mantenerlos, dado que son gastos probablemente bien declarados, y porque serán tratados a nivel de gasto por división, como se explica a continuación.

El tercer paso en la depuración de casos atípicos se basó en el análisis de *outliers* realizado a las variables de gasto a utilizar para la clasificación. En la tabla 6 se expone la cantidad de observaciones marcadas como *outliers*, el método utilizado para el análisis y la cantidad de observaciones efectivamente depuradas.

La decisión en relación al tratamiento de estos *outliers* fue no eliminarlos, sino que truncarlos, para disminuir las distancias entre los hogares con gastos elevados en cada variable, permitiendo una agrupación más óptima. Para lograr un suavizamiento más sutil (similar a los que se han realizado en versiones de Casen con sus factores de expansión), se truncaron solo los *outliers* que estaban por sobre el percentil 90 de la distribución, considerando solo estos casos atípicos, truncándose justamente al valor del gasto porcentual que se encontraba en esa posición.

En promedio se detectaron 90 observaciones como atípicas, que fluctúan entre un mínimo de 8, en los gastos en alimentación y bebidas no alcohólicas, y un máximo de 204, para los gastos en servicios financieros y seguros. Como resultado del tratamiento, fueron depurados en promedio 8 observaciones por tipo de gasto, con un mínimo de 3 en alojamiento y en educación, y un máximo de 21 en servicios financieros y seguros. Como se mencionó anteriormente, no se depuró el gasto total, además porque este gasto no participa como variable en la clasificación.

Tabla 6: Cantidad de *outliers* detectados y depurados y método de detección

Gasto	Variable	OL detectados	OL depurados	Método
Gasto total	Gasto_tot	185	0	Hadi
Alimentos y bebidas no alcohólicas	D01	8	6	RI
Alcohol	G02_1	113	12	Hadi
Tabaco	G02_2	67	7	Hadi
Vestuario y calzado	D03	92	9	Hadi
Alojamiento	D04	34	3	Hadi
Muebles	D05	112	12	Hadi
Salud	D06	135	14	Hadi
Adquisición y funcionamiento de vehículos	G07_1_2	20	2	Hadi
Servicios de transporte	G07_3	78	7	Hadi
Comunicaciones	D08	70	7	Hadi
Recreación y cultura	D09	100	10	Hadi
Educación	D10	25	3	Hadi
Restaurantes y hoteles	D11	82	9	Hadi
Cuidado personal	G12_1_3	117	12	Hadi
Gastos en servicios financieros y seguros	G12_5_6	204	21	Hadi

Fuente: Elaboración propia en base a VIII EPF-INE.

Como resultado del proceso de depuración, se eliminaron 159 hogares en total (1,04% de la muestra), resultando una base de datos con 15.080 hogares, con los cuales se realizaron los análisis.

Anexo 3: Descripción de variables suplementarias

Tabla 7: Variables suplementarias			
Nombre	Descripción	Tipo de variable	Categorías
Zona	Corresponde a la zona geográfica a la que pertenece cada hogar según la representatividad de la VIII EPF.	Categórica	1. Gran Santiago 2. Resto de capitales regionales
Gasto mensual total	Corresponde al gasto en consumo final mensualizado de los hogares en todos los ítems de gasto considerados por la VIII EPF.	Numérica - continua	
Ingreso disponible mensual total	Corresponde al ingreso disponible total mensualizado de los hogares, de acuerdo a la definición de ingreso utilizada por la VIII EPF.	Numérica - continua	
Quintil de ingreso disponible mensual per cápita	Para construir esta variable se ordena a los hogares expandidos según su ingreso disponible mensual per cápita, de forma ascendente y luego se establecen cortes para dividir la muestra en 5 porciones, cada una de alrededor de un 20% del total. Es un proxy de estratificación socioeconómica.	Ordinal	1. Quintil 1 2. Quintil 2 3. Quintil 3 4. Quintil 4 5. Quintil 5
Tenencia de la vivienda principal	Esta variable indica cuál es la situación de los hogares encuestados en torno a la propiedad o tipo de ocupación de la vivienda. Esta variable se utiliza en otros contextos como parte de los componentes del indicador de "tenencia segura de la vivienda", que permite la identificación de hogares vulnerables por pertenecer a barrios marginales. Se unieron algunas categorías de la variable original para ajustarla a los fines de este estudio.	Categórica	1. Propia, totalmente pagada 2. Propia pagándose 3. Arrendada 4. Cedida, herencia, sucesión y otros
Edad de la jefatura de hogar	Es una variable a nivel hogar que indica la edad de quien se declaró jefe o jefa en el hogar.	Numérica - discreta	
Tramo de edad de la jefatura de hogar	Corresponde a la variable de edad de quien se declaró jefe o jefa de hogar, ordenado en tramos. Se utiliza como una medida más distribucional, complementando la información del promedio de edad de la jefatura de hogar.	Ordinal	1. 15 - 24 años 2. 25 - 34 años 3. 35 - 44 años 4. 45 - 54 años 5. 55 - 64 años 6. 65 o más

Hogar adulto mayor	Esta variable identifica a los hogares en los cuales todos sus miembros son personas de 60 años o más. La construcción de esta variable fue tomada del IPC de adulto mayor que publica Clapes UC.	Dicotómica	0. Hogar no adulto mayor 1. Hogar adulto mayor
Estructura del hogar	Esta variable corresponde a una tipología de hogares utilizada para estudios de Cepal, que fue publicada en la investigación Tipos de hogares urbanos y etapas del ciclo de vida familiar (Arriagada, 1997). Esta variable considera la cantidad de personas, núcleos al interior del hogar y las relaciones de parentesco.	Categórica	1. Hogares unipersonales 2. Hogares monoparentales 3. Hogares nucleares 4. Hogares extensos 5. Hogares compuestos 6. Hogares sin núcleo
Ciclo familiar	Esta variable, al igual que la estructura del hogar, también es una tipología utilizada por Cepal (Arriagada, 1997). Esta variable, además de la cantidad de personas, núcleos al interior del hogar y sus relaciones de parentesco, utiliza las edades de los hijos y de sus madres para su construcción.	Categórica	1. Hogar no familiar 2. Pareja joven sin hijos 3. Inicio de la familia 4. Expansión 5. Consolidación 6. Desmembramiento 7. Nido vacío
Escolaridad de el/la sustentador/a principal en tramos	Es una variable a nivel hogar que indica el mayor nivel educativo logrado por la persona que aporta más al presupuesto del hogar.	Ordinal	1. Menos que media completa 2. Media completa 3. Superior incompleta 4. Superior completa 5. Postgrado
Tipo de estudiante	Esta variable indica, para todos los miembros del hogar, si es que asisten a la educación formal y, en caso de asistir, a cuál, en términos agregados.	Categórica	1. No estudia 2. Escolar 3. Superior
Dependencia administrativa	Esta variable indica la dependencia administrativa del establecimiento educacional al cual asisten las personas que declararon asistir a la educación superior.	Categórica	1. Universidad privada 2. Universidad del consejo de rectores 3. Centro de formación técnica (CFT) o Instituto profesional (IP)
Sistema de salud	Corresponde al sistema de salud en el cual se encuentra afiliado el jefe o la jefa de hogar, y en el caso de que esta persona no tenga un sistema de salud, el de su pareja o cónyuge.	Categórica	1. No tiene 2. FONASA 3. FFAA 4. ISAPRE 5. Otro

Deuda total mensual	Es la suma de los compromisos financieros declarados por todos los miembros de un hogar en todos los ítems consultados a pagarse dentro de un mes. Incluye el valor de la cuota, intereses, gastos de administración, comisiones y seguros asociados al pasivo financiero.	Numérica - continua
Carga financiera total	Es la proporción que alcanza la deuda total mensual respecto al Ingreso disponible mensual total del mismo hogar. Puede ser mayor que 100% cuando la deuda total mensual es mayor que el ingreso disponible mensual total.	Numérica - continua
Deuda hipotecaria mensual	Es la suma de las deudas hipotecarias declaradas por todos los miembros de un hogar a pagarse dentro de un mes. Incluye el valor de la cuota, intereses, gastos de administración, comisiones y seguros asociados a los créditos hipotecarios.	Numérica - continua
Carga hipotecaria	Es la proporción que alcanza la deuda hipotecaria mensual respecto al Ingreso disponible mensual total del mismo hogar.	Numérica - continua
Deuda de consumo mensual	Es la suma de las deudas en créditos de consumo declaradas por todos los miembros de un hogar a pagarse dentro de un mes. Incluye el valor de la cuota, intereses, gastos de administración, comisiones y seguros asociados a los créditos hipotecarios.	Numérica - continua
Carga de consumo	Es la proporción que alcanza la deuda hipotecaria mensual respecto al Ingreso disponible mensual total del mismo hogar.	Numérica - continua
Deuda automotriz mensual	Corresponde a la suma de las deudas declaradas por todos los miembros de un hogar en créditos automotrices. Incluye el valor de la cuota, intereses, gastos de administración, comisiones y seguros asociados a los créditos automotrices.	Numérica - continua
Carga automotriz	Es la proporción que alcanza la deuda automotriz mensual respecto al Ingreso disponible mensual total del mismo hogar.	Numérica - continua

Deuda educacional mensual	Corresponde a la suma de las deudas declaradas por todos los miembros de un hogar en créditos educacionales. Incluye el valor de la cuota, intereses, gastos de administración, comisiones y seguros asociados a los créditos educacionales.	Numérica - continua
Carga educacional	Es la proporción que alcanza la deuda educacional mensual respecto al Ingreso disponible mensual total del mismo hogar.	Numérica - continua
Deuda cuentas corrientes mensual	Corresponde a la suma de las deudas declaradas por todos los miembros de un hogar por el uso de líneas de crédito de cuentas corrientes. Incluye el valor del monto adeudado, intereses, gastos de administración, comisiones y seguros asociados a cuentas corrientes.	Numérica - continua
Carga cuentas corrientes	Es la proporción que alcanza la deuda líneas de crédito de cuentas corrientes respecto al Ingreso disponible mensual total del mismo hogar.	Numérica - continua
Deuda tarjetas de crédito bancarias mensual	Corresponde a la suma de las deudas declaradas por todos los miembros de un hogar por el uso de tarjetas de crédito de entidades bancarias. Incluye el valor de la cuota, intereses, gastos de administración, comisiones y seguros asociados a tarjetas de crédito.	Numérica - continua
Carga tarjetas de crédito bancarias	Es la proporción que alcanza la deuda en tarjetas de crédito bancarias respecto al Ingreso disponible mensual total del mismo hogar.	Numérica - continua
Deuda tarjetas de casas comerciales mensual	Corresponde a la suma de las deudas declaradas por todos los miembros de un hogar por el uso de tarjetas de casas comerciales. Incluye el valor de la cuota, intereses, gastos de administración, comisiones y seguros asociados a tarjetas de casa comercial.	Numérica - continua
Carga tarjetas de casas comerciales	Es la proporción que alcanza la deuda en tarjetas de crédito bancarias respecto al Ingreso disponible mensual total del mismo hogar.	Numérica - continua

Sobre endeudamiento	Es una variable construida para señalar a los hogares que tienen una carga financiera total mayor al 40% de su ingreso disponible total mensual. Este umbral se ha considerado crítico por parte del <i>Bank for International Settlements</i> (BIS, 2014)	Dicotómica	0. Hogar no sobre endeudado 1. Hogar sobreendeudado
---------------------	--	------------	--
