



PROYECTO DE GRADO

Presentado ante la ilustre UNIVERSIDAD DE LOS ANDES como requisito parcial para
obtener el Título de INGENIERO DE SISTEMAS

IDENTIFICACIÓN DE LOS PATRONES DE GASTOS FAMILIARES
EN VENEZUELA Y EN LA CIUDAD DE MÉRIDA, PERIODO
2.004-2.005

Por

Br: José Vladimir Nava Lacruz

Tutor: Prof. Georges Jabbour

Cotutor: Prof. Gerardo Colmenares

Octubre 2.009

©2.009 Universidad de Los Andes Mérida, Venezuela

Identificación de los patrones de gastos familiares en Venezuela y en la ciudad de Mérida, periodo 2.004-2.005

Br. José Vladimir Nava Lacruz

Proyecto de Grado — Investigación de Operaciones, 163 páginas

Resumen: El presente estudio está dirigido a la identificación de patrones de gastos familiares, tanto a nivel nacional como en la ciudad de Mérida, a partir de los datos obtenidos en la III Encuesta Nacional de Presupuestos Familiares (III ENPF). La importancia de este estudio radica en la extracción de un nuevo conocimiento al analizar los datos antes mencionados, estableciendo a su vez comparaciones entre los patrones de gastos familiares de la ciudad de Mérida con los extraídos a nivel nacional, analizando a su vez las características socio-económicas de las familias clasificadas bajo los mismos modos de comportamiento de gasto familiar. En vista de que los patrones de gastos constituyen una clasificación otorgada a los distintos modos de comportamiento del gasto familiar, en este estudio se construyeron modelos de clasificación basados en Máquinas de Vectores Soporte, ya que esta técnica de aprendizaje automático ha demostrado una gran capacidad para la generalización y el pronóstico en problemas de clasificación.

Palabras claves: Patrones de gastos familiares, Encuesta Nacional de Presupuestos Familiares, Análisis de conglomerados, Máquinas de vectores soporte.

*A mis padres y mis hermanos,
por ser la fuente inagotable de motivación
necesaria para la culminación de
este proyecto.*

Índice

Índice.....	iv
Índice de Tablas.....	vii
Índice de Figuras.....	x
Agradecimientos.....	xi
Capítulo 1 Introducción.....	1
1.1 Antecedentes.....	3
1.2 Planteamiento del problema.....	7
1.3 Objetivos.....	8
1.4 Justificación.....	9
1.5 Organización del documento.....	10
Capítulo 2 Patrones de gastos familiares.....	12
2.1 Definición de patrón de gastos familiares.....	13
2.2 Importancia del estudio de los patrones de gastos familiares.....	14
2.3 Encuesta Nacional de Presupuestos Familiares.....	15
Capítulo 3 Técnicas de análisis multivariante de datos.....	21
3.1 Análisis factorial.....	21
3.2 Análisis de conglomerados.....	28
Capítulo 4 Máquinas de vectores soporte.....	35
4.1 Máquinas de vectores soporte para clasificación binaria.....	35
4.2 Máquinas de vectores soporte para la multclasificación.....	40
4.3 Procedimiento recomendado para la clasificación con MVS.....	41
Capítulo 5 Preprocesamiento de los datos.....	45

5.1	Estructura original de los datos de la III Encuesta Nacional de Presupuestos Familiares	46
5.2	Definición de grupos de gastos e ingresos	48
5.3	Creación de la matriz de datos	50
5.4	Integración de la información en una matriz de datos única	55
5.5	Manejo de datos faltantes.....	57
5.6	Manejo de valores atípicos.....	58
Capítulo 6	Identificación de patrones de gastos familiares.....	60
6.1	Análisis exploratorio de las variables de gasto.....	61
6.2	Metodología aplicada en la extracción de patrones de gastos en las familias venezolanas.....	71
6.3	Descripción de los patrones de gastos familiares a nivel nacional	77
6.4	Descripción de los patrones de gastos existentes en las familias de la ciudad de Mérida	81
6.5	Análisis comparativo entre los patrones de gastos familiares existentes en la ciudad de Mérida y los extraídos a nivel nacional.....	85
6.6	Características socio-económicas de las familias clasificadas bajo los distintos patrones de gastos familiares.....	91
Capítulo 7	Experimentación con máquinas de vectores de soporte	101
7.1	Metodología para la experimentación con MVS.....	102
7.2	Resultados obtenidos en las pruebas con MVS: patrones de gastos familiares a nivel nacional.....	107
7.3	Resultados obtenidos en la experimentación con MVS: patrones de gastos familiares obtenidos en la ciudad de Mérida.....	122
Capítulo 8	Conclusiones y recomendaciones	124
8.1	Conclusiones	124
8.2	Recomendaciones	127

Referencias bibliográficas	128
Anexos	132

Índice de Tablas

Tabla 4.1: Principales funciones núcleo.....	40
Tabla 6.1: Test de Kolmogorov-Smirnov para las variables de gasto.....	62
Tabla 6.2: Matriz de correlaciones de las variables de gasto.....	64
Tabla 6.3: Medidas de adecuación del modelo factorial.....	65
Tabla 6.4: Comunalidades explicadas por los primeros 8 factores bajo distintos métodos de obtención de factores.	67
Tabla 6.5: Varianza explicada por los primeros 8 factores, método de los componentes principales.	68
Tabla 6.6: Matriz de carga factorial, método ACP.....	69
Tabla 6.7: SCDG para particiones impuestas por el algoritmo K medias con y sin variables estandarizadas.....	72
Tabla 6.8: SCDG de distintos números de particiones, para distintos ordenes de las observaciones dentro de la muestra.....	73
Tabla 6.9: SCDG para las divisiones del conglomerado con más observaciones en las particiones originales de 3, 4, 5 y 6 grupos.....	74
Tabla 6.10: SCDG para las divisiones del segundo conglomerado con más observaciones en las particiones originales de 3, 4, 5 y 6 grupos.	74
Tabla 6.11: SCDG de los modelos finales para la representación de patrones de gasto a nivel nacional.....	75
Tabla 6.12: Patrones de gastos a nivel nacional y su presencia en la muestra de estudio.	76
Tabla 6.13: Estructura del patrón de gastos familiar a nivel nacional tipo A.....	78
Tabla 6.14: Descripciones resumidas de los patrones de gastos nacionales.	79

Tabla 6.15: Patrones de gastos en las familias merideñas y su presencia en la muestra de estudio.....	82
Tabla 6.16: Descripciones resumidas de los patrones de gasto de las familias de la ciudad de Mérida, municipio Libertador.	83
Tabla 6.17: Tabla de equivalencias entre los patrones de gastos locales y nacionales.	86
Tabla 7.1: Tamaños muestrales para distintos niveles de error máximo permitido. Método de Tortora.	108
Tabla 7.2: Modelos MVS con kernel lineal. Patrones de gastos nacionales como clase de salida.....	109
Tabla 7.3: Desempeño de los modelos MVS propuestos con kernel lineal. Patrones de gastos nacionales como clase de salida.	110
Tabla 7.4: Modelos MVS con kernel RBF. Patrones de gastos nacionales como clase de salida.....	111
Tabla 7.5: Desempeño de los modelos MVS propuestos con kernel RBF. Patrones de gastos nacionales como clase de salida.....	112
Tabla 7.6: Modelos MVS con kernel polinomial. Patrones de gastos nacionales como clase de salida.	113
Tabla 7.7: Desempeño de los modelos MVS propuestos con kernel polinomial. Patrones de gastos nacionales como clase de salida.	113
Tabla 7.8: Desempeño del mejor modelo MVS con kernel lineal para distintos tamaños muestrales.	116
Tabla 7.9: Matriz de clasificación obtenida por el modelo MVS con kernel lineal. .	117
Tabla 7.10: Desempeño del mejor modelo MVS con kernel RBF para distintos tamaños muestrales.	118
Tabla 7.11: Matriz de clasificación obtenida por el modelo MVS con kernel RBF...	119

Tabla 7.12: Desempeño del mejor modelo MVS con kernel polinomial para distintos tamaños muestrales.	120
Tabla 7.13: Matriz de clasificación obtenida por el mejor modelo MVS con kernel polinomial.	121
Tabla 7.14: Mejores modelos MVS para la clasificación de los patrones de gastos familiares en la ciudad de Mérida.	122
Tabla 7.15: Matriz de clasificación obtenida por el mejor modelo MVS con los patrones de gasto de la ciudad de Mérida como clase de salida.	123

Índice de Figuras

Figura 4.1: Hiperplanos separadores e hiperplano de separación óptimo.....	36
Figura 4.2: MVS con margen blando.....	37
Figura 7.1: Ejemplo de gráfico de contornos. Búsqueda de rejilla en el espacio dado por los intervalos $[2^0, 2^{20}]$ para C y $[2^{-15}, 2^5]$ para γ .	105
Figura 7.2: Esquema del procedimiento general para la construcción	106
del modelo de clasificación de MVS.....	106

Agradecimientos

Mis sinceros agradecimientos a:

A Dios, por ser quien me dio la fuerza y salud necesaria para avanzar hasta donde me he propuesto. Gracias por acompañarme en todo momento.

A Mi familia: Pilar fundamental de mi vida.

A Los Profesores Georges Jabbour y Gerardo Colmenares, por guiarme en la realización de este proyecto de grado y brindarme sus oportunos consejos.

A la Universidad de Los Andes y a la escuela de Ingeniería de Sistemas, por brindarme los conocimientos necesarios para alcanzar esta meta.

A todos mis compañeros de estudio, por compartir los momentos gratos y menos gratos de esta experiencia: convertirnos en Ingenieros de Sistemas.

Capítulo 1

Introducción

La era actual se caracteriza por factores como la transculturización, la globalización y los continuos avances científicos y tecnológicos, que aunado con los rasgos inherentes a las sociedades de cada país, entre ellos la cultura, los sistemas de valores existentes, las características geográficas del país y la situación económica de sus habitantes; influyen en el comportamiento de consumo de bienes y servicios tanto de manera individual como en el comportamiento de familias enteras. Debido a que estos comportamientos de consumo se encuentran en constante cambio, es importante diagnosticar las distintas variaciones en los modos de consumo que presentan los habitantes de un país, pues ello podría reflejar variaciones en las condiciones de vida, el bienestar social y demás indicadores de desarrollo humano de una nación.

Dada la importancia del estudio del comportamiento de consumo, tanto a nivel individual como a nivel de familias y grupos sociales enteros, en muchos países son conducidas numerosas encuestas de carácter socio-económico, destacando en nuestro país la Encuesta Nacional de Presupuestos Familiares (ENPF). Desde su inicio, las diversas ENPF se han centrado en el estudio de la familia y la manera en que ésta maneja su presupuesto, obteniendo así información

sobre los ingresos, gastos, características de las viviendas que habitan las familias, la composición del hogar y otras variables económicas y sociales de sus miembros [2].

La gran cantidad de datos recabados en la III Encuesta Nacional de Presupuestos Familiares, aplicada en el período que abarca desde Noviembre de 2.004 hasta Noviembre del año 2.005, representa una fuente invaluable de información socio-económica para numerosas investigaciones sociales basadas en dicho período, incluyendo el presente estudio, cuyo enfoque está en la extracción y estudio de modos de comportamiento generales existentes en las familias venezolanas respecto a los gastos incurridos en los distintos tipos de rubros, bienes y servicios. Estos modos de comportamiento familiar respecto al gasto son los llamados patrones de gastos familiares, los cuales representan esquemas de prioridades y magnitudes desembolsadas en los distintos bienes y servicios (es decir, son modalidades en que se distribuye el presupuesto familiar), presentes de manera similar en familias que usualmente comparten características socio-económicas semejantes.

A diferencia de estudios previos sobre presupuestos familiares, entre los que destacan las investigaciones de Anido (1.998), Márquez (2.004), Sosa (2.006) y Aranguren (2.008), cuya característica común radica en el uso de técnicas de reducción de dimensionalidad para la definición de grupos y patrones de gastos, entre ellas el análisis de componentes principales, el análisis factorial y el análisis factorial de correspondencias; en el presente estudio se aborda la identificación de patrones de gastos familiares, tanto a nivel nacional como para las familias en la ciudad de Mérida, mediante técnicas del análisis de conglomerados, concretamente el algoritmo de las K-medias.

En vista de que los patrones de gastos a extraer en el presente estudio representan en su conjunto un sistema de clasificación del gasto familiar, es importante construir modelos de clasificación en base a dichos patrones, de modo que sea posible asignar o predecir el modo de gasto de una familia conociendo los valores que toman sus variables de gasto, es decir los desembolsos totales en cada rubro considerado. Para ello, en el presente estudio se utilizó el método de las Máquinas de Vectores Soporte (MVS), que según la opinión de diversos investigadores (Gunn, 1.997), ha demostrado poseer un gran desempeño tanto en problemas de clasificación como de regresión, debido a la incorporación del principio de Minimización de Riesgo Estructural, que en el caso de la clasificación dota a las MVS de mayor capacidad para generalizar, es decir, mayor capacidad de pronóstico de nuevas observaciones.

La importancia de este estudio radica en la extracción de nuevo conocimiento al analizar los datos de la III ENPF, pues aún no se han conducido suficientes investigaciones a los datos obtenidos en esta encuesta, la cual cuenta con un factor de expansión para las familias entrevistadas en la ciudad de Mérida, municipio Libertador, lo que permite establecer comparaciones entre los patrones de gastos familiares de la ciudad de Mérida con los obtenidos a nivel nacional.

1.1 Antecedentes

Hasta ahora se han desarrollado y culminado tres Encuestas Nacionales de Presupuestos Familiares (ENPF). La primera durante el período 1.987-1.988 (I ENPF), la segunda durante 1.997-1.998 (II ENPF) y la tercera durante el período 2.004-2.005 (III ENPF). Además se han realizado encuestas de presupuestos familiares a nivel regional (EPF), entre ellas la EPF de 1.967 y la EPF de 1.986, ambas

en la ciudad de Mérida y llevadas a cabo por el IIES – ULA – CORPOANDES (Anido *et al.*, 2.007). A partir de estas encuestas se han estudiado diversos aspectos económicos, entre ellos el análisis de la distribución del gasto familiar, la concentración de ingreso, etc. Estas investigaciones (la mayoría de ámbito regional) servirán de referencia para el presente estudio, pues están basadas en los datos recabados por la I y II ENPF. Algunas de las investigaciones basadas en las encuestas de presupuestos familiares se citan a continuación:

Anido (1.998) realizó un estudio con el objeto de calcular los coeficientes de elasticidad precio, elasticidad ingreso y elasticidad cruzada de la demanda para la ciudad de Mérida, a partir de los datos de la EPF del 86. Especificó y estimó un sistema lineal de gasto, utilizando como métodos de estimación mínimos cuadrados ordinarios y regresiones aparentemente no relacionadas. Los resultados obtenidos revelaron que:

- La demanda promedio de las familias de la ciudad de Mérida, en las categorías del gasto correspondientes a alimentación, vivienda y sus servicios, indican que solo éstos son bienes de carácter normales y necesarios.
- Existe una escasa respuesta del consumidor en la demanda de todos los bienes cuando los precios de éstos varían.
- En promedio, las familias urbanas de Mérida ponderan prioritariamente sus gastos en alimentación y vivienda, y de esta última le importan principalmente sus características de equipamiento y comodidad, antes que otros rasgos relativos a su estructura.

Márquez (2.004) diseñó e implementó una base de datos, la cual permitió realizar de forma eficiente un conjunto de consultas a los datos de la II Encuesta de Presupuestos Familiares. Aprovechando las bondades de la base de datos, realizó una serie de análisis estadísticos, entre ellos los análisis de correspondencias simple y múltiple, los cuales permitieron describir de forma clara la relación existente entre distintas variables socio-económicas de carácter cualitativo. Por otro lado, con la idea de relacionar las variables cualitativas antes mencionadas con variables cuantitativas de gastos del hogar, realizó una serie de análisis de componentes principales sobre estas últimas variables, categorizando a los hogares con las variables socio-económicas del tipo categórico o cualitativo. Entre las conclusiones obtenidas se destaca:

- Carabobo, Distrito Federal, Miranda, Anzoátegui, Zulia y Bolívar, son los estados en los cuales los hogares tienen mayores desembolsos.
- Los hogares cuyo jefe del hogar tiene alto nivel educativo tienden a tener altos niveles de consumo.
- Aquellos hogares con bajas características cualitativas (viviendas en condiciones poco favorables, bajo nivel educativo, desempleo del jefe del hogar, etc.) presentan bajos consumos en bienes de primera necesidad.

Sosa (2.006) desarrolló una investigación para conocer la estructura del gasto en familias de ocho regiones de Venezuela, usando los datos de la I Encuesta Nacional de Presupuestos Familiares. Como primer paso, realizó un análisis de componentes principales, observando una relación existente entre treinta y cuatro variables que identifican once grupos de gasto. En segundo lugar, ejecutó un análisis de correspondencias múltiples, que permitió identificar las relaciones entre variables

categorías en viviendas; esto con la finalidad de conocer las características que presentan las viviendas de las familias venezolanas. Por último, calculó la distribución del ingreso, el índice de Gini y la Curva de Lorentz para cada una de las ocho regiones consideradas, para determinar si el ingreso estaba equitativamente distribuido entre las familias de cada región. A través de esta investigación se pudo determinar lo siguiente:

- La región Zuliana, opuesta a la región de los Andes y Guayana, presenta gran desigualdad en la distribución del ingreso.
- Los gastos que prevalecieron de manera general, para las ocho regiones consideradas, fueron los gastos en alquiler de vivienda y sus servicios, gastos en alimentos y en transporte.

Anido *et al.* (2007) realizaron un análisis comparativo de las encuestas de presupuestos familiares hechas desde 1967 hasta 1998, con el fin de estudiar la estructura del gasto y del ingreso familiar en la ciudad de Mérida, Venezuela. Analizaron las dos primeras ENPF y las EPF hechas en la ciudad de Mérida (EPF67 y EPF86). El énfasis del estudio radica en el análisis del gasto según su finalidad (alimentación, vestido y calzado, servicios de la vivienda, transporte, salud, educación, etc.) y el estudio de la concentración del ingreso y del gasto familiar, mediante el coeficiente de Gini. Esta investigación arrojó los siguientes resultados:

- El gasto alimentario es la mayor fracción del presupuesto total en los diversos períodos.
- El poder adquisitivo de las familias merideñas sufría de un mercado deterioro, incidiendo en el nivel de bienestar del hogar.

- Existía un aumento progresivo en los niveles de pobreza de la ciudad y un incremento progresivo en la concentración del ingreso en la ciudad de Mérida.

Aranguren (2.008) llevó a cabo una investigación dirigida a la identificación de patrones de consumo de las familias venezolanas, a partir de los datos obtenidos en la II Encuesta Nacional de Presupuestos Familiares. En este estudio se utilizó, en primer lugar, la técnica multivariante Análisis de Componentes Principales para definir variables latentes con mayor poder discriminante; y en segundo lugar, utilizó Máquinas de Vectores de Soporte como herramienta de clasificación. Dicho estudio arrojó cuatro modos prevaletentes de consumo en el país para los años estudiados: 1) Gastos básicos; 2) Gastos de educación y servicios diversos, 3) Gastos de recreación; y 4) Gastos financieros, tributarios y legales. Además se demostró la alta capacidad de generalización que presentan las máquinas de vectores de soporte como método de clasificación no lineal.

1.2 Planteamiento del problema

La mayoría de los estudios hechos por el Instituto de Investigaciones Económicas y Sociales (IIES), perteneciente a la Universidad de Los Andes, han sido de carácter regional y han utilizado métodos estadísticos lineales tales como las técnicas multivariantes. El inconveniente del uso de muchos de estos métodos lineales radica en que los datos de estudio deben cumplir con ciertos supuestos para que sus resultados sean válidos. Por ello se iniciaron investigaciones de estos patrones de gastos usando técnicas que requieren de mínima verificación de supuestos iniciales, entre ellos técnicas no lineales como las máquinas de vectores soporte (Aranguren, 2.008). Sin embargo, aún faltan investigaciones que apliquen técnicas alternativas en

la extracción de patrones de gastos familiares, de modo que permitan reconocer cuáles son las técnicas más apropiadas en la identificación de patrones de gastos provenientes de los datos de Encuestas Nacionales de Presupuestos Familiares. A su vez, la mayoría de investigaciones sobre patrones de gastos hechas por el IIES carecen de comparaciones entre los resultados obtenidos en la ciudad de Mérida con los resultados a nivel nacional.

1.3 Objetivos

1.3.1 Objetivo general

Identificar los patrones de gasto, tanto de las familias venezolanas como de las merideñas, a partir de los datos obtenidos en la III Encuesta Nacional de Presupuestos Familiares (III ENPF), aplicada en el período 2.004-2.005.

1.3.2 Objetivos específicos

- Realizar una revisión bibliográfica concerniente a las Encuestas de Presupuestos Familiares y la teoría económica relacionada con el estudio de los patrones de gastos familiares.
- Identificar la estructura de la III Encuesta Nacional de Presupuestos Familiares.
- Obtener una muestra de estudio o vista minable a partir de los datos de la encuesta.

- Seleccionar y usar un método estadístico multivariante, que permita una reducción inicial de las variables originales consideradas en el estudio, en variables latentes.
- Utilizar un método proveniente del análisis de conglomerados para identificar los patrones de gastos tanto en las familias venezolanas como en las de la ciudad de Mérida.
- Comparar los resultados obtenidos de los patrones de gasto en la ciudad de Mérida con los obtenidos a nivel nacional.
- Analizar las características socio-económicas de las familias que comparten los mismos patrones de gastos, verificando de esta manera la validez de los distintos patrones extraídos.
- Construir modelos de clasificación basados en Máquinas de Vectores Soporte para el pronóstico de los patrones de gastos familiares extraídos.

1.4 Justificación

Actualmente existe la necesidad de desarrollar una investigación de los patrones de gastos familiares, tanto en Venezuela como en la ciudad de Mérida, usando los datos obtenidos en la III Encuesta Nacional de Presupuestos Familiares (III ENPF). Como la adquisición de los datos recopilados en esta encuesta es relativamente reciente (2.006), aún no se han llevado a cabo suficientes investigaciones estadísticas, por ello, su estudio permitirá conocer los distintos tipos de comportamientos respecto al gasto, presentes en las familias en el período de estudio (2.004 - 2.005). Además, la mayoría de los estudios sobre patrones de gasto hechos hasta ahora han utilizado

métodos lineales (que requieren en su mayoría el cumplimiento de supuestos) y no han establecido comparaciones entre los resultados obtenidos en la ciudad de Mérida y los extraídos a nivel nacional. En el presente estudio se pretende hacer uso de métodos alternativos para la extracción de los patrones de gasto familiares. Por último, se hará un estudio comparativo entre los resultados de Mérida con los obtenidos a nivel nacional, identificando a su vez las características socio-económicas de las familias clasificadas bajo los mismos patrones de gastos.

1.5 Organización del documento

En este documento se presentan siete capítulos en total, incluyendo este capítulo introductorio. En el capítulo 2 se define el concepto de patrones de gastos familiares, su importancia y se describe, a manera general, las características de la III Encuesta Nacional de Presupuestos Familiares, incluyendo la definición de los objetivos, ámbito geográfico, las variables investigadas e instrumentos básicos de muestreo de esta investigación socio-económica. El capítulo 3 explica los fundamentos teóricos de las técnicas de análisis multivariante utilizadas, tanto para fines exploratorios: análisis factorial, como para la extracción y descripción de patrones de gasto familiar: análisis de conglomerados. El capítulo 4 explica las nociones básicas sobre las Máquinas de Vectores Soporte. El capítulo 5 se centra en la parte experimental inicial del estudio, el preprocesamiento de los datos. El capítulo 6 se centra en la identificación de patrones de gastos familiares y las comparaciones generales entre los resultados obtenidos en la ciudad de Mérida respecto a los extraídos a nivel nacional, para posteriormente describir las características socio-económicas de las familias clasificadas bajo los mismos patrones de gasto. A su vez, se presentan los resultados obtenidos por modelos de clasificación basados en Máquinas de Vectores

Soporte. Por último, el capítulo 7 presenta las conclusiones y recomendaciones del estudio realizado.

Capítulo 2

Patrones de gastos familiares

Venezuela no escapa de la realidad mundial, cuya situación actual está caracterizada por la introducción de nuevas tecnologías, la globalización, los cambios en las economías y la industrialización, entre otros factores. Esta realidad ha condicionado nuestra economía y nuestro comportamiento como sociedad. Cada día son diseñados e introducidos al mercado nuevos productos, bienes y servicios que, aunado con los previamente existentes, buscan satisfacer nuestras necesidades a través de su consumo. Dada la situación económica de nuestro país, el consumidor venezolano ha priorizado los gastos en que incurre a aquellos que maximizan su satisfacción, administrando de manera creativa los ingresos monetarios. Se busca, en la mayoría de las veces, la eficiencia en las adquisiciones hechas no sólo para el consumo personal sino también a nivel familiar.

Toda familia tiene prioridades a satisfacer: alimento, vestido, vivienda, educación, transporte, recreación, salud, etc.; que varían en orden y en magnitud de familia en familia, sin embargo esto no implica la inexistencia de comportamientos comunes. De hecho, las familias venezolanas comparten más similitudes que diferencias en lo que respecta a la cultura y su sistema de valores, factores que ejercen una influencia en el consumo. Surgen entonces comportamientos de consumo basados en las valoraciones, prioridades, percepciones e intereses de los

miembros de las familias, que se refleja en la secuencia de elecciones y acciones del hogar, incluyendo la selección, compra, uso, mantenimiento y reparación de los productos y servicios. Estos comportamientos se mantienen en su mayoría en cada familia y se presentan de manera similar en muchas familias de características comparables, definiendo de esta manera patrones de gastos familiares.

2.1 Definición de patrón de gastos familiares

En el presente estudio se define a un patrón de gastos familiares como un comportamiento, tendencia o predisposición de un grupo familiar en la adquisición y consumo de bienes y servicios pertenecientes a distintos rubros, que se repite o se presenta de un modo similar en familias con características similares. Bajo este enfoque, un patrón de gastos familiares describe una estructura o esquema de prioridades y magnitudes desembolsadas en los distintos bienes y servicios, por lo que representa una modalidad en la que una familia distribuye su presupuesto.

Como existen diversos factores que influyen en el comportamiento de gasto de una familia (siendo tal vez el nivel económico el más importante de ellos), las prioridades otorgadas a los distintos rubros varían dentro de los estratos de nuestra sociedad, por lo que existen diferentes patrones de gastos dentro de las familias venezolanas. A los efectos del presente estudio, los patrones de gastos familiares extraídos guardan diferencias significativas entre sí (tanto en las prioridades otorgadas a las distintas variables de gasto como en sus montos desembolsados), representando cada uno de ellos un modo de comportamiento general al que tienden un grupo representativo de familias que fueron entrevistadas en la III ENPF.

2.2 Importancia del estudio de los patrones de gastos familiares

La identificación y entendimiento de los patrones de gastos en las familias venezolanas facilita la toma de decisiones en materia política social. Nuestro país ha registrado en los años anteriores fluctuaciones a nivel económico, hecho apoyado principalmente por las ganancias de la industria petrolera, por lo que es conveniente estudiar los posibles cambios en el comportamiento de consumo de los hogares venezolanos y determinar si dicho comportamiento refleja un desarrollo en cuanto a bienestar social se refiere.

Tal como menciona González (2.002), “un sistema económico debe ser juzgado por su capacidad de asegurar la satisfacción de las necesidades de consumo de las personas, lo que equivale a asegurar un nivel de vida adecuado, entendido como el acceso en grado suficiente al conjunto de satisfactores requeridos”. Entre este conjunto de “satisfactores” se encuentran los bienes y servicios, que pueden ser vistos como requisitos materiales para una determinada calidad de vida. El bienestar de la sociedad debe reflejarse en el consumo satisfactorio para todos los integrantes de la misma, es por esa razón que los estudios de patrones de gastos familiares deben tener un carácter periódico. Estos estudios permiten describir los umbrales de confort material a los que aspiran las personas y que, en la mayoría de los casos, varían y están desigualmente incorporados entre los distintos grupos sociales. La descripción del comportamiento de consumo por parte de la familia venezolana pudiera expresar males sociales o un excesivo costo de la vida en nuestra sociedad.

Por estas razones son conducidas numerosas investigaciones de carácter socio-económico en nuestro país, entre ellas destaca la Encuesta Nacional de

Presupuestos Familiares, la cual representa una gran fuente de información necesaria para el estudio de los patrones de gastos familiares (Aranguren, 2.008). En la siguiente sección se describe las características generales de este importante estudio socio-económico.

2.3 Encuesta Nacional de Presupuestos Familiares

La Encuesta Nacional de Presupuestos Familiares (ENPF) es una de las investigaciones estadísticas más importantes que se realizan en el país. Se trata de un estudio de carácter cooperativo entre instituciones como el Banco Central de Venezuela (BCV), el Instituto Nacional de Estadística (INE), la Corporación Venezolana de Guayana (CVG) y la Universidad de Los Andes (ULA); cuyo propósito es identificar las prioridades económicas y sociales de los venezolanos [2].

Desde sus inicios, la Encuesta Nacional de Presupuestos Familiares (ENPF) se ha enfocado en el estudio de tres principales variables: los gastos del hogar, los ingresos del hogar y diversas variables socio-demográficas [2]. Hasta ahora han sido completadas en nuestro país tres ENPF: I Encuesta Nacional de Presupuestos Familiares (I ENPF) en el periodo 1.988 – 1.989, la II Encuesta Nacional de Presupuestos Familiares (II ENPF) en el periodo de 1.997 – 1.998 y la III Encuesta Nacional de Presupuestos Familiares en el período de 2.004 – 2.005 (Anido *et al*, 2.007). En el presente estudio, los patrones de gastos familiares a extraer tanto a nivel nacional como en la ciudad de Mérida se basan en los datos recabados por la III Encuesta Nacional de Presupuestos Familiares.

2.3.1 Objetivos de la Encuesta Nacional de Presupuestos Familiares

El objetivo fundamental de esta investigación socio-económica es recabar una gran cantidad de datos acerca de la familia venezolana, incluyendo los ingresos, egresos, características de las viviendas que habitan, su composición y otras variables económicas y sociales de sus miembros (Anido *et al*, 2.007). A nivel específico, la ejecución de este tipo de encuestas en nuestro país busca satisfacer los siguientes objetivos:

1. Obtener información que permita actualizar la estructura de la canasta de bienes y servicios familiar y sus diversas ponderaciones, de modo que permita el cálculo del Índice de Precios al Consumidor (IPC) del Área Metropolitana de Caracas.
2. Ampliar, a nivel nacional, la cobertura geográfica del Índice de Precios al Consumidor.
3. Actualizar las cuentas del sector hogares, componente necesario para el Sistema de Cuentas Nacionales (SCN). El SCN es un instrumento estadístico que permite orientar y armonizar el registro de las transacciones económicas del país.
4. Proporcionar información necesaria para realizar análisis econométricos.
5. Disponer de un conjunto de indicadores sociales para el estudio de las condiciones de vida en los hogares.

2.3.2 Características generales de la III Encuesta Nacional de Presupuestos Familiares

La III Encuesta Nacional de Presupuestos Familiares (III ENPF) fue una investigación que comprendió un periodo de 56 semanas, desde Noviembre del 2.004 hasta Noviembre del 2.005. Esta encuesta tuvo una cobertura geográfica de todo el territorio nacional, considerando distintos tipos de dominios o ciudades con cantidades de población dentro de diversos rangos preestablecidos: Dominio 1, que representa el Área Metropolitana de Caracas y sus zonas adyacentes; Dominio 2, donde se incluyen ciudades importantes y con gran cantidad de población; Dominio 3, representado por las ciudades medianas o con población entre 50.000 y 250.000 habitantes; el Dominio 4 que comprende a las ciudades pequeñas o con población entre los 5.000 y 50.000 habitantes y el Dominio 5 que incluye a las localidades con menos de 5.000 habitantes [2].

La muestra extraída por la III ENPF, representativa de los hogares individuales residentes en el país, se obtuvo mediante un muestreo probabilístico estratificado bietápico, con un tamaño de 9.120 viviendas [2]. Para esta muestra de estudio fueron investigadas tres tipos de variables: los gastos del hogar, los ingresos del hogar y variables socio-demográficas. Dentro de los gastos se incluyen los bienes y servicios consumidos por cualquier miembro del hogar, sin importar quien lo haya pagado e independientemente del periodo en que haya sido consumido; y los gastos no imputables al consumo, por ejemplo las transferencias de dinero, pago de multas, impuestos, etc. Los ingresos del hogar están referidos a todos los ingresos que han sido recibidos por sus miembros sin importar su origen ni su naturaleza; mientras que las variables socio-demográficas fueron las características de las viviendas (tipo, apariencia, disponibilidad de los servicios básicos, etc.), las características de los

miembros del hogar (nombre, sexo, edad, nivel de educación alcanzado, etc.) y el equipamiento del hogar (Aranguren, 2008).

2.3.3 Períodos de referencia usados en la III Encuesta Nacional de Presupuestos Familiares

Dado que el estudio del presupuesto familiar considera una gran cantidad de bienes, servicios, diversas formas de gastos e ingresos, deben definirse varios periodos de referencia. Esto se debe a que no todos los bienes son adquiridos en la misma frecuencia, pues el consumo de ellos depende de factores como el tipo de artículo o bien (de primera necesidad, de lujo, etc.) y su tipo de duración (percederos, no percederos, etc.), entre otros. Los periodos de referencia son “los períodos de tiempo dentro de los cuales va a ser considerada la realización de determinados pagos o la percepción de ciertos ingresos” (Meza, 2.001). Trabajar con ciertos períodos de referencia fue necesario debido a que tanto los gastos como los ingresos son variables que no se efectúan o perciben regularmente en un cierto período.

La III ENPF registra los diversos gastos e ingresos en periodos de referencia que son definidos por los criterios de importancia de los mismos, la frecuencia con que se realizan dichos gastos e ingresos y la capacidad de las familias encuestadas en recordar dichos montos. Los gastos que son más frecuentes y/o los que son más difíciles de recordar debido a sus bajos montos se clasificaron con los períodos de menor referencia, mientras que aquellos gastos más costosos y por lo tanto, más fáciles de recordar, se clasificaron dentro de períodos de referencia más altos.

Los periodos de referencia definidos en la III ENPF para los gastos fueron los periodos diario, mensual, trimestral y anual. A continuación se describe detalladamente cada uno de los tipos de gasto según su período de referencia:

- **Gastos diarios:** incluye a los gastos que se realizan con alta frecuencia de consumo y/o con bajos precios. Los datos sobre gastos diarios registrados por la III ENPF corresponden a aquellos efectuados y registrados por las familias en la semana de la entrevista.
- **Gastos mensuales:** dentro de los gastos mensuales se incluyen pagos en alquileres de viviendas, mensualidades escolares o de estacionamientos, servicios del hogar como luz, agua, teléfono, y demás gastos efectuados por la familia en el mes anterior al de la entrevista.
- **Gastos trimestrales:** son aquellos gastos cuya periodicidad es más frecuente que la semestral pero no más que la mensual. Dentro de este periodo de referencia destacan los gastos en muebles, equipos del hogar, servicios de salud y mantenimiento de la vivienda. Dentro de los gastos trimestrales se incluyeron los gastos efectuados (y recordados) por la familia en los últimos tres meses con base a la semana de entrevista.
- **Gastos anuales:** son los gastos menos comunes y/o con un valor elevado que facilita al encuestado recordarlos, entre ellos destacan la compra de vehículos, seguros, pagos de impuestos y remodelaciones de la vivienda.

Respecto a los ingresos, los periodos de referencia definidos son: diario, semanal, mensual y anual. Las fuentes de ingresos cuyos periodos de referencia son el diario o el semanal corresponden generalmente a personas que se dedican al trabajo informal o al trabajo independiente, entre ellos destacan vendedores de

comida, taxistas, dueños de bodega, criadores de animales, etc. Los ingresos que fueron etiquetados como mensuales fueron aquellos que se reciben de fuentes de trabajo formal: sueldos, salarios, honorarios profesionales, becas, etc. Los ingresos anuales incluyen beneficios que se tienden a obtener con poca frecuencia a través de utilidades, herencias en dinero, premios de loterías, etc.

2.3.4 Instrumentos para la recolección de la información

Para la recolección de la información de las variables de interés en la III ENPF, se utilizaron nueve cuestionarios o EPF, cada uno con preguntas específicas, los cuales debían ser completados por los miembros del hogar entrevistado. Cada cuestionario tiene su propósito específico, sea contabilizar los gastos en un periodo de referencia dado, los ingresos del hogar o las características de la familia y de la vivienda, entre otros. Los nueve cuestionarios fueron: EPF1, características de las unidades básicas de muestreo; EPF2, características de la vivienda; EPF3, características generales del hogar; EPF4, gastos diarios del hogar; EPF5, gastos diarios personales; EPF6, gastos mensuales; EPF7, gastos trimestrales y anuales; EPF8, Ingresos y EPF9, producción por trabajo independiente, micro negocio o finca agrícola [2].

Capítulo 3

Técnicas de análisis multivariante de datos

En este capítulo se introducen las técnicas multivariantes utilizadas para la extracción de patrones de gastos familiares. Dado que estos patrones no son evidentes a primera vista, es necesario aplicar un conjunto de técnicas que permitan una mejor descripción de los distintos comportamientos de gastos familiares en cuanto a su estructura y su frecuencia observada en las familias encuestadas. La extracción de los patrones de gastos familiares, que representan una clasificación impuesta a las familias según su comportamiento de gasto (que es la variable de interés), se obtuvo a través del uso de un método de agrupación por conglomerados, luego de estudiar los datos originales mediante el análisis factorial.

A continuación se introducen los fundamentos teóricos de las técnicas multivariantes a utilizadas en este estudio.

3.1 Análisis factorial

El análisis factorial es un método multivariante que busca la reducción de las dimensiones de un conjunto de datos, identificando variables ocultas o factores que expliquen la mayor parte de la variabilidad de estos datos originales (Pérez, 2.004). Esta técnica es útil cuando las variables originales guardan correlaciones

significativas entre sí, con lo que las variables ejercen influencias similares en el fenómeno de estudio.

Cuando en un conjunto de p variables medidas para un fenómeno de estudio existen correlaciones significativas, el análisis factorial trata de hallar k factores que expliquen suficientemente a este conjunto de variables observadas, de modo que éstos factores sean fácilmente interpretables y que se necesite la menor cantidad de ellos (con lo que $k < p$). El conjunto de factores no solo se usa para facilitar la comprensión del fenómeno en estudio, sino también para sustituir al conjunto de variables originales en técnicas multivariantes de uso posterior, entre ellos los modelos de regresión y las técnicas de agrupamiento (clustering), pues estos métodos requieren de variables incorrelacionadas para la obtención de resultados válidos (Pérez, 2.004).

3.1.1 Modelo matemático

El modelo matemático del análisis factorial supone que para p variables estandarizadas X_1, X_2, \dots, X_p , existe un número de factores comunes F_1, F_2, \dots, F_k para todas las variables y un factor único (o específico) para cada variable. Al expresar las variables originales en función de estos factores, se tiene la expresión 3.1:

$$\begin{aligned}
 X_1 &= l_{11}F_1 + l_{12}F_2 + \dots + l_{1k}F_k + e_1 \\
 X_2 &= l_{21}F_1 + l_{22}F_2 + \dots + l_{2k}F_k + e_2 \\
 &\dots \dots \dots \\
 X_p &= l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pk}F_k + e_p
 \end{aligned}
 \tag{3.1}$$

La expresión anterior se puede representar bajo la forma matricial $\mathbf{X} = \mathbf{L}\mathbf{F} + \mathbf{e}$, donde \mathbf{X} es el vector de las variables originales, \mathbf{L} es la matriz de carga factorial, \mathbf{F} es el vector de factores comunes y \mathbf{e} es el vector de factores únicos. En la expresión 3.1, los valores l_{ih} son las cargas factoriales y representan la correlación entre la variable i y el factor h . En un modelo factorial, los factores comunes y específicos deben cumplir, en la mayor medida posible, con los siguientes supuestos (Gondar, 2.000):

1. Los factores comunes son variables con media cero y varianza 1. Además deben estar incorrelacionados entre sí.
2. Los factores únicos son variables con media cero. Sus varianzas pueden ser distintas. Se supone que están incorrelacionados entre sí, de lo contrario la información contenida en ellos estaría en los factores comunes.
3. Los factores comunes y los factores únicos están incorrelacionados entre sí.

3.1.2 Comunalidades y especificidades

Cuando se representa a un conjunto de variables estandarizadas X_i por un modelo factorial, la varianza de cada una de estas variables está dada por:

$$Var(X_i) = 1 = l_{i1}^2 + l_{i2}^2 + \dots + l_{ip}^2 + \omega_i^2 \quad (3.2)$$

En la expresión 3.2, cada l_{ih}^2 representa la proporción de varianza total de la variable X_i explicada por el factor h , mientras que la suma de las proporciones l_{ih}^2 , es decir $h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{ip}^2$, se le conoce como comunalidad. La comunalidad representa la proporción de varianza de la variable X_i que explican los distintos

factores en su conjunto, con lo que su valor varía entre 0 (los factores no explican nada de la variable) y 1 (los factores explican el 100% de la variable). Respecto al término ω_i^2 , este es denominado especificidad y representa la contribución del factor único en la variabilidad total de la variable X_i . Un modelo factorial es adecuado cuando presenta altas comunales para las variables originales.

Por otro lado, la suma de las cantidades $l_{1h}^2 + l_{2h}^2 + \dots + l_{ph}^2 = g_h$ constituye la capacidad del factor h para explicar la varianza total de las variables, es por eso que la contribución del factor h en la explicación de la varianza total es g_h/p , cuando se trata de p variables estandarizadas.

3.1.3 Métodos para la obtención de los factores

La construcción de un modelo factorial se centra en la estimación de los coeficientes l_{ih} de la matriz de carga factorial L en el modelo $X = L.F + e$. Estas cargas factoriales estimadas indican los pesos de los distintos factores en la comunalidad de cada variable original. Una vez estimada la comunalidad h_i^2 a partir de las estimaciones de los l_{ih} (dado que $h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{ip}^2$), la especificidad ω_i^2 de la variable se estima mediante la relación $\omega_i^2 = 1 - h_i^2$.

Existen diversos métodos para la obtención de la matriz de carga factorial, entre ellos destacan el método de los componentes principales, el método de máxima verosimilitud, el método del centroide, el método del factor principal, el método alpha y el método de los ejes principales, entre otros. Cada uno de estos métodos difiere tanto en el algoritmo de cálculo como en los supuestos considerados (normalidad), por lo que gozan de características distintivas que le otorgan ventajas

respecto a los demás métodos (por ejemplo, el método de los componentes principales siempre garantiza una única solución). Una descripción detallada de cada uno de estos métodos de obtención de factores, incluyendo las diversas formulaciones matemáticas y los supuestos considerados, se puede hallar en Pérez (2.004).

3.1.4 Bondad de ajuste del modelo factorial

Como se explicó anteriormente, un modelo factorial resulta útil cuando las variables originales están correlacionadas entre sí de un modo lo suficientemente alto como para poder ser resumidas por un conjunto de k factores, incorrelacionados entre sí, que explican la mayor parte del comportamiento de dichas variables. El primer paso para determinar la adecuación del modelo factorial es analizar la matriz de correlación muestral, verificando la existencia de correlaciones altas entre las variables (mayores a 0.5) [1]. Adicionalmente, existen pruebas estadísticas que permiten evaluar la adecuación del análisis factorial, entre ellos los más comunes son el test de esfericidad de Barlett y la medida de adecuación muestral de Kaiser, Meyer y Olkin (Pérez, 2.004).

- **Test de esfericidad de Barlett:** En este test se contrasta las hipótesis nula de que la matriz de correlaciones R_p es una matriz identidad, en cuyo caso no existirían correlaciones significativas entre las p variables y el análisis factorial no sería adecuado. Se plantea el contraste: $H_0: |R_p| = 1$ vs $H_1: |R_p| \neq 1$. La hipótesis se contrasta mediante el uso de la matriz de correlación muestral R y el estadístico introducido por Barlett $B = -[n - 1 - (2p + 5)] \ln |R|$, que

bajo la hipótesis nula tiene una distribución chi-cuadrado con $p(p - 1)/2$ grados de libertad.

- **Medidas de adecuación de Kaiser, Meyer y Olkin (KMO):** Esta medida busca contrastar el efecto de los coeficientes de correlación parcial en la correlación entre las variables. El estadístico KMO varía entre 0 y 1, siendo recomendables valores mayores a 0.7. Valores bajos de este estadístico implicarían que las correlaciones entre algún par de variables no pueden ser explicadas por el resto de variables.

3.1.5 Rotación de los factores

En ciertos casos, los factores extraídos no tienen una interpretación razonable, dado que la configuración de las cargas de las variables originales dentro de cada factor impide la descripción del factor como tal (los valores de carga dentro de los factores son similares para varias de las variables originales). Como el propósito de estos factores es facilitar la comprensión del fenómeno de estudio y las interrelaciones existentes entre las variables, se pueden introducir transformaciones a la matriz de carga L para facilitar la descripción de los factores (Gondar, 2.000). A estas transformaciones se le conoce como rotaciones factoriales y consisten en multiplicar a la matriz de carga factorial por una matriz de transformación, manteniendo constante las comunalidades para cada variable original. Existen dos tipos de rotaciones factoriales:

- **Rotaciones ortogonales:** En este tipo de rotaciones los factores resultantes siguen siendo incorrelacionados entre sí, es decir, los ejes de los factores son ortogonales. Las principales rotaciones ortogonales son Varimax, Quartimax y Equamax.

- **Rotaciones oblicuas:** En las rotaciones oblicuas los factores no son incorrelacionados, con los que los ejes de los factores ya no son ortogonales y por ende dos o más factores explicarán un comportamiento similar. Estas rotaciones se utilizan cuando los factores no son interpretables mediante rotaciones ortogonales. Entre las principales rotaciones oblicuas destacan: Oblimax, Oblimin, Quartimin y Promax.

Mayor información sobre los fundamentos matemáticos de las rotaciones ortogonales y oblicuas puede verse en Pérez (2.004).

3.1.6 Medición de los factores

Los factores extraídos por el análisis factorial no sólo se usan para entender las correlaciones de un grupo de variables, sino también para sustituir a dichas variables originales (correlacionadas), pues estos factores constituyen un grupo de variables independientes entre sí, útiles para análisis posteriores como la regresión múltiple o el análisis de conglomerados.

Para utilizar a los factores extraídos en análisis posteriores, es necesario calcular los valores que toma cada factor F_j en cada observación dados los valores de sus variables originales X_i , para lo cual existen diversos métodos basados en estimaciones de dichos factores. Los métodos más usados para el cálculo de las puntuaciones o valores de los factores para cada observación son cuatro: método de los componentes principales, método de regresión por mínimos cuadrados, método de Bartlett y el método de Anderson-Rubin. La formulación matemática de estos métodos también puede apreciarse en Pérez (2.004).

3.2 Análisis de conglomerados

El análisis de conglomerados o análisis de clúster comprende un conjunto de técnicas de clasificación automática o no supervisada, es decir, de reconocimiento de patrones sin supervisión (Peña, 2.002). Dada una muestra de observaciones o individuos, cada uno de los cuales posee un número de variables o atributos, el análisis de conglomerados define grupos de observaciones lo más homogéneos posible en base a las variables observadas. La formación de estos grupos homogéneos busca la maximización de la similitud de las observaciones que pertenecen a un mismo grupo, maximizando a su vez las diferencias entre las observaciones pertenecientes a grupos distintos. El propósito de este tipo de métodos multivariantes es clasificar a los datos de un fenómeno de estudio en grupos o conglomerados, previamente desconocidos pero que se pueden extraer en función de la disposición de los datos.

Este tipo de análisis se puede utilizar tanto para la agrupación de observaciones (individuos) como en la agrupación de variables, por lo que existen tres tipos de problemas que aborda el análisis de conglomerados: 1) La partición de datos o agrupamiento no jerárquico, 2) La construcción de jerarquías o agrupamiento jerárquico y 3) La clasificación de variables.

1. **Partición de datos o agrupamiento no jerárquico:** Bajo este problema se tiene un conjunto de datos que se sospechan son heterogéneos y se desea dividirlos en un número de grupos prefijado, de modo que cada individuo pertenezca a uno y sólo uno los grupos, que todos los elementos queden clasificados y que cada grupo sea internamente homogéneo, es decir que los elementos que lo constituyen estén muy cercanos.

2. **Construcción de jerarquías o agrupamiento jerárquico:** El propósito del agrupamiento jerárquico es construir una estructura arborescente, donde los elementos forman jerarquías de acuerdo a sus similitudes, con lo que grupos de niveles más bajos van siendo englobados en otros niveles superiores (Pérez, 2.004). Cada nivel de jerarquía permite construir un tipo de partición de los datos en grupos. Este tipo de clasificación es muy frecuente en Biología al clasificar animales, plantas, etc.
3. **Clasificación de variables:** Cuando se consideran muchas variables resulta útil el análisis exploratorio que divida a dichas variables en grupos. Con ello se busca reducir la dimensión del problema en estudio. Para este problema se pueden usar tanto métodos de agrupamiento jerárquico como no jerárquico.

Para llevar a cabo un análisis de conglomerados es necesario tomar tres decisiones fundamentales (Hair *et al*, 1.999): Primero, establecer cómo se mide la proximidad entre dos observaciones, la cual puede definirse por medidas de correlación, medidas de distancia y medidas de similitud, entre otras. Segundo, elegir la manera en que se formarán los conglomerados, de modo que las observaciones más similares pertenezcan al mismo conglomerado; y en tercer lugar, elegir el número de conglomerados que se deben formar. En el caso del presente estudio, el método de agrupación utilizado fue el de K-medias, que pertenece a los métodos de agrupamiento no jerárquico. En la próxima sección se describen los principales criterios de proximidad entre observaciones: distancias y similitudes.

3.2.1 Medidas de distancia y similaridad

La formación de conglomerados está condicionada por la manera en cómo se mide la proximidad o similaridad entre dos observaciones. Generalmente se definen los grupos usando medidas de distancia cuando las variables de cada observación son numéricas, mientras que para las variables categóricas y binarias se recomienda el uso de medidas de similaridad (Pérez, 2.004).

3.2.1.1 Medidas de distancia

Matemáticamente, la distancia entre dos puntos A y B es aquella medida que satisface los siguientes axiomas:

1. $d(A, B) \geq 0$ y $d(A, A) = 0$
2. $d(A, B) = d(B, A)$
3. $d(A, B) \leq d(A, C) + d(C, B)$

Existen varios tipos de medidas de distancia, cada una de ellos producirá resultados distintos para el mismo método de agrupamiento y el mismo conjunto de datos. Las medidas tradicionales de distancias son aquellas que se aplican a dos observaciones o instancias. Suponiendo que las observaciones A y B poseen una dimensión n , las distancias más comúnmente utilizadas son la distancia Euclídea, la distancia de Manhattan, la distancia de Chebychev, la distancia del coseno y la distancia de Mahalanobis (Hernández *et al.*, 2.004).

Las medidas de distancias anteriores son susceptibles a las magnitudes de los valores de las variables. Si una o varias variables tienen mayor dispersión o magnitudes que el resto, éstas tendrán mayor impacto en los valores de distancia y por ende, en la formación de los grupos. Modificar la escala de las variables puede

alterar completamente el resultado del análisis. Es recomendable estandarizar los datos cuando se desee asignar un peso semejante a las variables, sin importar su variabilidad original, sin embargo, en ocasiones esto no suele ser adecuado (Peña, 2.002).

3.2.1.2 Medidas de similaridad

El coeficiente de similaridad según la variable j entre dos observaciones A y B ($j = 1, \dots, p$), es una función no negativa y simétrica S_{jAB} que cumple con las condiciones:

1. $S_{jAA} = 1$
2. $0 \leq S_{jAb} \leq 1$
3. $S_{jAB} = S_{jBA}$

Para variables numéricas, las principales medidas de similaridad son los coeficientes de correlación de Pearson, los coeficientes entre variables como lambda, tau, etc.; o la expresión:

$$S_{jAB} = 1 - \frac{|X_{Aj} - X_{Bj}|}{\text{rango}(X_j)} \quad (3.3)$$

Respecto a las variables cualitativas, los coeficientes de similaridad pueden construirse individualmente o por bloques. En el caso de una variable binaria, la similaridad para dos observaciones será 1 si ambos poseen el atributo y 0 en caso contrario.

3.2.2 Métodos de agrupamiento no jerárquico. Algoritmo de las K-medias

Los métodos no jerárquicos o métodos partitivos tienen por objetivo realizar una partición de los individuos en K grupos. Los K grupos o clases no están relacionadas entre sí y tampoco están organizadas jerárquicamente. La asignación de observaciones a estos grupos se hace mediante un proceso que optimice algún criterio de selección. Uno de los métodos de agrupamiento no jerárquico más común es el algoritmo de las K-medias, el cual divide al conjunto de datos a clasificar de acuerdo a un número de grupos prefijado K , reasignando las observaciones a medida que itera el proceso (Peña, 2.002). Este algoritmo tiende a reducir progresivamente la varianza dentro de los grupos hasta que se alcance algún criterio de parada. Dada una muestra de n observaciones con p variables, las etapas de este algoritmo comprenden:

1. Seleccionar K puntos como centros iniciales. Esto puede hacerse seleccionándolos de manera aleatoria del conjunto de datos n , tomando los K puntos más alejados entre sí o seleccionar los centros con información previa.
2. Calcular las distancias euclídeas de cada observación a los centros de los K grupos y asignarlas al grupo cuyo centroide esté más cercano. Esta asignación es secuencial, de modo que al introducir una observación a un grupo deben recalcularse las coordenadas del nuevo centro del grupo.
3. Definir un criterio de optimalidad y comprobar si reasignando alguna de las observaciones mejora el criterio. Cuando no sea posible mejorar más el criterio de optimalidad, finaliza el proceso de partición.

3.2.2.1 Criterio de optimalidad

Reconociendo a la distancia euclídea como medida de proximidad en este método de agrupación, el criterio de optimización es la minimización de la *suma de cuadrados dentro de los grupos* (SCDG), cuya expresión es:

$$\text{SCDG} = \sum_{k=1}^K \sum_{j=1}^p \sum_{i=1}^{n_k} (X_{ijk} - \bar{X}_{jk})^2 \quad (3.4)$$

donde X_{ijk} es el valor de la variable j en la observación i del grupo k , \bar{X}_{jk} es la media de la variable j en el grupo k y n_k es el número de observaciones en el grupo k . Alternativamente se puede minimizar la traza de la matriz W , que se obtiene mediante la expresión:

$$W = \sum_{k=1}^K \sum_{i=1}^{n_k} (X_{ik} - \bar{X}_k)(X_{ik} - \bar{X}_k)^T \quad (3.5)$$

En este caso, al minimizar la traza de W se minimiza la SCDG. A este criterio se le llama criterio de la traza y fue propuesto por Ward en 1.963.

En vista de que las observaciones se asignan secuencialmente a los grupos, el cálculo de los centroides de cada grupo puede verse afectado por el orden de las observaciones en el conjunto total de datos, por lo que conviene repetir el algoritmo permutando el orden de los datos dentro de la muestra para descartar algún efecto significativo (Peña, 2.002).

3.2.2.2 Elección del número de grupos

Uno de los parámetros que queda a criterio del analista a la hora de utilizar el método de las k -medias, es la elección del número de grupos o particiones que permitan clasificar correctamente al conjunto de datos. La minimización de la SCDG no sirve como criterio de la elección del número K , puesto que impondría el valor de K igual al número de observaciones. Un criterio comúnmente utilizado es el cálculo de la magnitud en la reducción de la SCDG para un agrupamiento de K a uno de $K+1$ grupos. Para ello se construye el estadístico F de reducción de variabilidad:

$$F = \frac{SCDG(K) - SCDG(K + 1)}{SCDG(K + 1)/(n - K - 1)} \quad (3.6)$$

El valor de este estadístico se compara con una F de p grados de numerador y $p(n - K - 1)$ grados de denominador, considerando más conveniente una clasificación con $K+1$ particiones si se rechaza la hipótesis de que la reducción de SCDG de K a $K+1$ grupos no es significativa (es decir, el valor del estadístico es mayor al valor tabulado para la F). Sin embargo, este criterio no es del todo razonable pues los supuestos necesarios para asumir una distribución F no necesariamente son satisfechos por los datos. Una regla empírica, sugerida por Hartigan en 1.975, establece la conveniencia de considerar un grupo más si el valor del estadístico es superior a 10 (Peña, 2.002).

A pesar de los criterios matemáticos expuestos, uno de los criterios más importantes en la elección de K lo constituye la interpretabilidad de los grupos obtenidos y su utilidad en la descripción del conjunto total de datos.

Capítulo 4

Máquinas de vectores soporte

Las Máquinas de vectores soporte (MVS), conocidas también como máquinas de soporte vectorial, constituyen una técnica de aprendizaje automático que se basa en el principio de Minimización del Riesgo estructural (Hernández *et al*, 2.004). En contraste con los métodos clásicos para la clasificación y predicción (i.e redes neuronales), que buscan la solución del problema mediante la reducción del espacio de datos y el uso de funciones no lineales para la clasificación en dicho espacio de dimensión menor; las MVS buscan la solución del problema en un espacio de dimensión mayor, en donde es posible hallar un separador lineal o hiperplano (Peña, 2.002). La búsqueda en este espacio de mayor dimensión se lleva a cabo mediante el uso de funciones núcleo (*kernel functions*), que trasladan los datos o ejemplos de aprendizaje desde el espacio de entrada hacia un espacio de características.

4.1 Máquinas de vectores soporte para clasificación binaria

En el problema de clasificación para dos clases, el objetivo consiste en separar las observaciones de dichas clases mediante una función (hiperplano), inducida por los patrones (ejemplos) disponibles. Para el caso en que estos dos conjuntos de datos

son linealmente separables, existen infinitos hiperplanos separadores, pero sólo uno maximiza el margen de separación entre las clases, tal como lo muestra la figura 4.1:

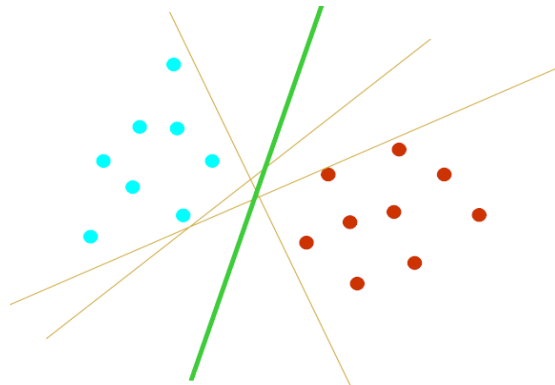


Figura 4.1: Hiperplanos separadores e hiperplano de separación óptimo.

La idea de las máquinas de vectores soporte es encontrar el hiperplano óptimo en términos de generalización, es decir, aquel que permita clasificar de manera correcta en la mayoría de los casos observaciones nuevas (Gunn, 1.997). Esto se consigue al seleccionar un hiperplano separador que esté a la misma distancia de los ejemplos más cercanos de cada clase, maximizando la distancia mínima entre el conjunto de datos y el hiperplano (principio de la maximización del margen). Cuando los datos no son linealmente separables en el espacio de entrada, son trasladados a un espacio de características de mayor dimensión mediante funciones núcleo, donde puede encontrarse un separador lineal con un margen máximo tal como en el caso anterior. En ocasiones, los ejemplos de aprendizaje no son linealmente separables inclusive en el espacio de características. Para estos casos se introduce el modelo de MVS con margen blando, el cual busca admitir errores en la clasificación de los ejemplos de manera controlada para obtener un clasificador más general, mediante el uso de variables de holgura (Betancourt, 2.005). En la figura 4.2 se puede apreciar el propósito de la introducción de las variables de holgura ξ como

cuantificadores del error en la clasificación por parte de los modelos MVS con margen blando:

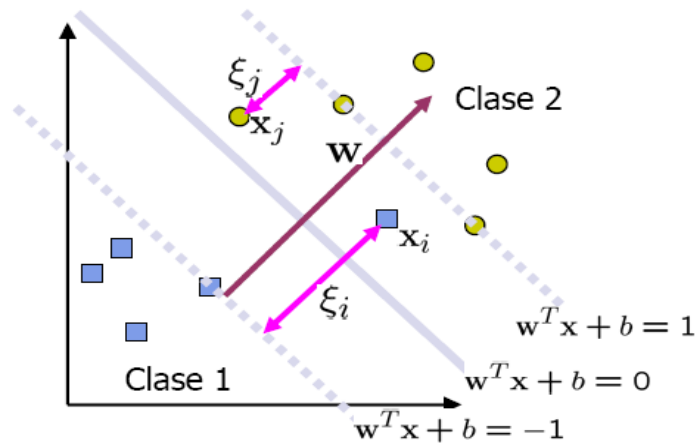


Figura 4.2: MVS con margen blando.

4.1.1 Formulación matemática del modelo MVS con margen blando

Dado un conjunto de vectores de entrenamiento x_i (con $i = 1, \dots, N$), pertenecientes a dos clases, y un vector Y que almacena en sus elementos y_i la clase correspondiente para el vector x_i (1,-1), el hiperplano separador del modelo MVS con margen blando es aquel que satisface las condiciones del siguiente problema de optimización:

$$\text{Minimizar } \frac{1}{2} \langle w, w \rangle + C \sum_{i=1}^N \xi_i$$

$$\text{sujeto a: } y_i [\langle w, x_i \rangle + b] \geq 1 - \xi_i, i = 1, \dots, N$$

$$\xi_i \geq 0 \quad 1 \leq i \leq N \quad (4.1)$$

El parámetro C en la formulación anterior es el término de regularización de los errores de clasificación, el cual establece un compromiso entre la maximización del margen entre las clases y las magnitudes que puedan tomar las variables de holgura ξ_i (y por ende la magnitud del error en la clasificación). Valores grandes de C obligan al método de aprendizaje a tener pocas variables de holgura diferentes de 0, por lo que esta constante constituye un parámetro de aprendizaje a fijar. Alternativamente, el modelo MVS con margen blando se puede representar mediante su formulación dual:

$$\begin{aligned} \text{Maximizar} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N y_i y_j \alpha_i \alpha_j K \langle x_i, x_j \rangle \\ \text{sujeto a:} \quad & \sum_{i=1}^N y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C \quad 1 \leq i \leq N \end{aligned} \quad (4.2)$$

Sea α_i^* la solución óptima de este problema dual, el hiperplano separador óptimo en el espacio de características está dado por la expresión:

$$h(x) = \sum_{i=1}^N y_i \alpha_i^* K \langle x_i, x \rangle + b^* \quad (4.3)$$

donde b^* es tal que $y_i h(x_i) = 1$ para todo i , cumpliendo con $0 \leq \alpha_i^* \leq C$. El clasificador asociado a este modelo MVS es $f(x) = \text{signo}(h(x))$. Bajo este modelo, las variables de holgura ξ_i son diferentes de cero cuando $\alpha_i^* = C$. Los ejemplos con $\alpha_i^* > 0$ constituyen los vectores de soporte, pero en el caso del margen blando, existen dos tipos de vectores de soporte:

- Cuando $0 < \alpha_i^* < C$, el ejemplo es un vector de soporte.
- Cuando $\alpha_i^* = C$, el ejemplo es un vector de soporte acotado que corresponde a un error en la clasificación ($\xi_i > 0$).

4.1.2 Funciones núcleo

El término $K(x_i, x_j)$ en la expresión 4.2 es la función núcleo, que también se conoce como función kernel. Con el uso de la función núcleo, los datos de un conjunto de aprendizaje son trasladados a un espacio de mayor dimensión, sin necesidad de conocer la función de transformación implicada (Gunn, 1997). Dado que las MVS solo dependen de la existencia de un producto escalar en el espacio de características, la función núcleo es la encargada de calcular este producto escalar de la función de transformación Φ en el espacio de características, es decir, calcula $\langle \Phi(x), \Phi(y) \rangle$. Para que una función $K(x, y) = \langle \Phi(x), \Phi(y) \rangle$ sea una función núcleo, debe cumplir con las siguientes condiciones:

1. Ser simétrica: $K(x, y) = K(y, x)$
2. La matriz $\mathbf{K} = \left(K(x_i, x_j) \right)_{i,j=1}^N$ debe ser semi definida positiva, es decir, carecer de autovalores negativos.
3. Cumplir con la desigualdad de Cauchy-Schwarz: $K(x, y)^2 \leq K(x, x)K(y, y)$
4. La función se puede expandir en una serie uniformemente convergente (Teorema de Mercer).

Las funciones núcleo más comúnmente utilizadas son las siguientes:

Función Kernel	Expresión	Parámetros
Lineal	$\langle x, y \rangle$	Ninguno
Polinómica	$(\gamma \langle x, y \rangle + r)^d$	$r \in R, d: \text{grado}, \gamma: \text{Gamma}$
Función de Base Radial (RBF)	$\exp(-\gamma \ x - y\ ^2)$	$\gamma: \text{Gamma}$
Sigmoidal	$\tanh(s(x, y) + r)$	$s, r \in R$
Multicuadrática inversa	$\frac{1}{\sqrt{\ x - y\ ^2 + coef}}$	$coef \geq 0$

Tabla 4.1: Principales funciones núcleo.

4.2 Máquinas de vectores soporte para la multclasificación

Para el problema de clasificación multiclase han sido propuesto varios métodos, la mayoría de ellos se basan en un esquema de binarización, convirtiendo un problema multiclase en varios binarios (Hernández *et al.*, 2004). Bajo el esquema de binarización existen dos enfoques principales para la clasificación de un conjunto de datos con k clases distintas:

1. **Enfoque Uno vs. Resto:** Se entrenan k clasificadores, donde una clase es positiva y las restantes son negativas. Posteriormente se predice el valor de un ejemplo usando todos los clasificadores. La clase predicha será aquella con la que se consiguió un margen mayor (en el caso de que se clasifique como positiva en más de un clasificador) (Moro y Hurtado, 2006).
2. **Enfoque Uno vs. Uno:** Se construyen y entrenan $K(K - 1)/2$ clasificadores, cada uno para dos clases diferentes. Para los conjuntos de datos de entrenamiento pertenecientes a las clases i y j , se soluciona el siguiente problema de clasificación binaria:

$$\begin{aligned}
 \text{Minimizar} \quad & \frac{1}{2} \langle w^{i,j}, w^{i,j} \rangle + C \sum_{i=1}^N \xi_t^{i,j} \\
 \text{sujeto a:} \quad & [\langle w^{i,j}, \Phi(x_t) \rangle + b^{i,j}] \geq 1 - \xi_t^{i,j} \quad \text{Si } x_t \text{ pertenece a clase } i \\
 & [\langle w^{i,j}, \Phi(x_t) \rangle + b^{i,j}] \leq -1 + \xi_t^{i,j} \quad \text{Si } x_t \text{ pertenece a clase } j \\
 & \xi_t^{i,j} > 0
 \end{aligned}$$

Para predecir la clase de un ejemplo de prueba, el método utiliza una estrategia de votación: la clase predicha por cada clasificador binario se considera como un voto para dicha clase, eligiéndose al final la clase con mayor cantidad de votos (C. Chang y C. Lin, 2.009).

Además de estas estrategias de binarización, se han formulado variantes en donde al modificar la función objetivo se obtiene un clasificador multiclase simultáneo. El problema con este enfoque es que puede resultar computacionalmente costoso (Hernández, 2.004).

4.3 Procedimiento recomendado para la clasificación con MVS

A pesar de que el entrenamiento y la validación de un clasificador basado en MVS resulta ser más sencillo que en otros modelos de clasificación, es necesario tener en consideración varios aspectos para construir modelos con el mejor desempeño posible. C. Hsu, C. Chang y C. Lin, pertenecientes a la National University of

Taiwan, proponen una metodología estándar para la clasificación con MVS, cuyos pasos se explican a continuación en las siguientes secciones.

4.3.1 Preprocesamiento de los datos

En la mayoría de los casos, entrenar una MVS con los datos originales dificulta la obtención de modelos con un buen desempeño. C. Hsu, C. Chang y C. Lin recomiendan dos tratamientos previos al conjunto de datos antes del entrenamiento de un clasificador MVS: numerización de variables categóricas y cambio en la escala de los datos (normalización, estandarización, etc.).

- **Numerización de variables categóricas:** Las MVS requieren que los atributos de los ejemplos para el aprendizaje sean números reales. Se recomienda usar m variables numéricas para representar una variable categórica de m valores. Solo una de estas variables tendrá valor uno y el resto cero.
- **Cambio en la escala de los datos:** Cambiar la escala de los datos tiene la ventaja de evitar que variables con grandes rangos numéricos dominen sobre los que tienen rangos numéricos pequeños. Por otro lado, se evitan las dificultades en el cálculo numérico, en especial cuando el conjunto de datos es grande. Se recomiendan la normalización $[0,1]$ o la estandarización $[-1,1]$ de las variables originales, tanto para los datos de entrenamiento como los de validación.

4.3.2 Elección del modelo

En la formulación de las MVS, los parámetros a asignar al modelo de clasificación son la constante de penalización C y los parámetros específicos de cada función

núcleo. Es por ello que el procedimiento para la elección del mejor modelo considera dos aspectos: las funciones núcleo a probar y la búsqueda de los parámetros ideales para la función núcleo correspondiente.

Función núcleo

C. Hsu, C. Chang y C. Lin recomiendan a la función de Base Radial (RBF) como primera elección, ya que mapea de manera no lineal un conjunto de datos de entrada a un espacio de mayor dimensión. También se ha demostrado que la función núcleo lineal es un caso especial de la RBF, así como también se ha planteado que el comportamiento de la función sigmoideal es similar a la de RBF para ciertos parámetros. La segunda razón para preferir a RBF es que tiene menor cantidad de hiperplanos que el kernel polinomial, por lo que el modelo goza de una menor complejidad.

A pesar de las ventajas de RBF, en algunas circunstancias es preferible usar otras funciones núcleo, por ejemplo, cuando el número de ejemplos es muy grande es conveniente usar el kernel lineal. Por esta razón, en el presente estudio se construyeron modelos MVS basados en las funciones kernel lineal, polinomial y función de base radial.

Búsqueda de parámetros del modelo: Validación cruzada y búsqueda de rejilla

Dependiendo de la función núcleo seleccionada para el entrenamiento, la elección de los parámetros consistirá en fijar el parámetro de regularización del error en la clasificación, C , y los k parámetros de la fórmula matemática de la función núcleo.

Se recomienda la búsqueda exhaustiva de los parámetros del modelo en un espacio de dimensión $K+1$, que también se le conoce como búsqueda de rejilla o mallado.

En cuanto a la medida del desempeño de las configuraciones de los parámetros en el modelo de MVS, C. Hsu, C. Chang y C. Lin recomiendan el procedimiento de validación cruzada de v pliegues, ya que permite prevenir el problema del sobreajuste. En la validación cruzada de v pliegues se divide el conjunto de entrenamiento en v subconjuntos de igual tamaño y luego se prueban secuencialmente cada subconjunto con el clasificador obtenido al entrenar con los $v-1$ subconjuntos restantes. De este modo cada uno de los ejemplos de entrenamiento es predicho, por lo que la precisión de la validación cruzada es el porcentaje de datos clasificados correctamente.

4.3.3 Entrenamiento

Después de la obtención de la configuración ideal para un kernel dado, el paso siguiente es entrenar la MVS con el conjunto total de datos para el aprendizaje. Al evaluar el desempeño de dicho entrenamiento, los resultados no deberían diferir mucho a los obtenidos en la fase de elección de parámetros y validación cruzada.

4.3.4 Validación

Una vez entrenado el modelo con los datos de entrenamiento, se debe comprobar la validez del mismo al tratar de predecir clases de un conjunto distinto al de entrenamiento, pero que corresponda al mismo fenómeno en estudio.

Capítulo 5

Preprocesamiento de los datos

La identificación de patrones de gastos familiares a efectuar en este estudio se centra en la minería de datos, que constituye el “proceso de extracción de conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos” (Hernández *et. al*, 2.004). La minería de datos presenta dos principales retos: 1) trabajar con grandes volúmenes de datos, con los problemas que ello conlleva: ruido, datos ausentes, datos atípicos, etc.; y 2) usar técnicas adecuadas para el análisis de los datos y la extracción de conocimiento útil y novedoso.

El preprocesamiento de los datos constituye la etapa donde se aborda el primer reto de la minería de datos. En dicha etapa se llevan a cabo la integración, manipulación y transformación del conjunto de datos de estudio, para obtener un conjunto de datos conciso, válido y que permita la extracción de información comprensible. Este conjunto de datos es comúnmente conocido como vista minable. Para el caso de la identificación de patrones de gasto, tanto en Venezuela como en la ciudad de Mérida, fue necesario procesar un gran volumen de datos provenientes de la III Encuesta Nacional de Presupuestos Familiares, lo cual requirió un largo período de tiempo, incluyendo el reconocimiento de la naturaleza de los datos, las estructuras de las fuentes de datos y las variables incluidas.

El procedimiento para la obtención de la vista minable, estructura de datos a partir de la cual se identificarán los patrones de gastos familiares del período 2.004 -2.005, constó de la ejecución de los siguientes pasos: 1) Reconocimiento de la estructura original de los datos de la III ENPF y selección de las fuentes de datos y variables útiles para el estudio. 2) Definición de los grupos de gastos e ingresos. 3) Creación de la matriz de datos, mediante la integración de los montos de gastos e ingresos registrados por las familias. 4) Manejo de valores faltantes y 5) Manejo de valores atípicos. En las siguientes secciones se detalla el procedimiento realizado paso a paso.

5.1 Estructura original de los datos de la III Encuesta Nacional de Presupuestos Familiares

Los datos de estudio recabados por la III Encuesta Nacional de Presupuestos Familiares fueron otorgados por el Instituto de Investigaciones Económicas y Sociales (IIES). Se trata de una base de datos conformada por cuarenta tablas, todas en el formato dBase (cuya extensión es .dbf). Para conocer y entender el contenido de cada una de las tablas fue necesario utilizar el catálogo de las variables existentes en la encuesta, realizado por el Banco Central de Venezuela.

No todas las tablas de esta base de datos contenían información relevante para el estudio. Después de la revisión minuciosa de cada tabla, su propósito y las variables que contenían, se llegó a la conclusión de que las tablas que constituyen la fuente primaria de datos para el presente estudio son las siguientes [2]:

1. Viviendas: Características de las viviendas censadas.
2. Hogares1: Características generales de los hogares encuestados y modo de propiedad de vivienda.
3. Hogares2: Registros de los vehículos pertenecientes a los hogares y sus características.
4. Personas_gd: Gastos diarios del hogar.
5. Personas_gdc: Gastos diarios del hogar en comidas.
6. Personas_gm: Gastos mensuales del hogar.
7. Personas_gt: Gastos trimestrales del hogar.
8. Personas_ga: Gastos anuales del hogar.
9. Personas_I: Ingresos monetarios de los miembros del hogar.
10. Personas_ti3: Sueldos, salarios y bonificaciones generadas por trabajo independiente, desempeñados por miembros del hogar.
11. Personas_ti4: Transacciones hechas por los miembros de la familia entorno al establecimiento independiente de trabajo de su propiedad.
12. Personas_ti5: Gastos en insumos utilizados por las personas con establecimientos de trabajo informal.
13. Personas_ti6: Gastos en materiales necesarios para el trabajo diario del establecimiento informal de los miembros de los hogares encuestados.
14. Personas_ti8: Ganancias de productores independientes del área de la agricultura, la ganadería y otras actividades agropecuarias.

15. Personas_ti19: Gastos e ingresos registrados por miembros de los hogares entrevistados con cultivos o producción agrícola.
16. Personas_ti20: Gastos en bienes y servicios utilizados en el mantenimiento de los establecimientos de agricultura, ganadería, cría de animales y otros referentes a actividades económicas rurales.
17. Personas_ti21: Registros correspondientes a activos (maquinarias, equipos, etc.) pertenecientes a los establecimientos de trabajo independientes de miembros de los hogares entrevistados.
18. Personas_am2: Ingresos por becas de las distintas misiones de carácter social.

Las variables seleccionadas en cada una de las tablas anteriores fueron aquellas en donde se especificaban montos de gastos incurridos en cualquier rubro por cualquier miembro de la familia, montos de ingresos monetarios o en especie registrados por cualquier miembro de la familia e información característica de la familia: ubicación geográfica, número de miembros del hogar, número de vehículos del hogar, tipo de vivienda y modo de propiedad de la misma.

5.2 Definición de grupos de gastos e ingresos

Las encuestas de presupuestos familiares se caracterizan por considerar un vasto número de bienes y servicios de los distintos tipos de rubros ofrecidos al consumidor. Estos productos y servicios conforman grupos definidos por características, naturaleza y uso similares. Toda encuesta de presupuestos familiares usa un sistema de clasificación donde el nivel más específico lo constituyen los bienes y servicios como tal, mientras que los niveles superiores lo constituyen bienes de un mismo tipo, por ejemplo: la leche es un producto como tal que forma parte del

nivel de abstracción más específico, mientras que el grupo de los productos lácteos constituye un grupo más general.

La clasificación de los gastos propuesta en la III Encuesta Nacional de Presupuestos Familiares consta de nueve grupos generales: 1) Alimentos, Bebidas y Tabaco, 2) Vestido y calzado, 3) Gastos de la vivienda y sus servicios, 4) Mobiliario, equipos del hogar y mantenimiento, 5) Salud, 6) Transporte y comunicaciones, 7) Educación, cultura y esparcimiento, 8) Artículos, efectos personales y servicios diversos y 9) Otros gastos. Aunque esta clasificación resulta suficientemente lógica, no es del todo conveniente adoptarla en el desarrollo de este estudio. Al revisar el conjunto de rubros, productos y servicios que comprende a cada grupo original, se pudo apreciar que éstos resultaban ser muy extensos y no permitirían una buena identificación del comportamiento de gasto de las familias, razón por la cual se decidió desglosar y reagrupar los diversos productos y servicios en nuevos grupos de gasto, formando una nueva clasificación.

Los grupos de gastos considerados por la nueva clasificación son 16, todos ellos definidos luego de considerar una serie de recomendaciones hechas por profesores con amplia experiencia en encuestas de presupuestos familiares, pertenecientes al IIES (Aranguren, 2.008). Estos nuevos grupos permiten la distribución adecuada de los montos de gastos desembolsados entre los distintos tipos de rubros, sin considerar una excesiva cantidad de grupos de gastos que dificultarían la extracción de patrones. Los 16 grupos son los siguientes: 1) Alimentos y bebidas no alcohólicas, 2) Bebidas alcohólicas y tabaco, 3) Comidas fuera del hogar, 4) Vestido y calzado, 5) Vivienda y sus servicios, 6) Mobiliario y equipos del hogar, 7) Mantenimiento de la vivienda, 8) Salud, 9) Transporte, 10) Equipos y servicios de comunicaciones, 11) Educación y cultura, 12) Recreación

y esparcimiento, 13) Artículos, efectos personales y servicios diversos, 14) Servicios financieros, tributarios y legales, 15) Viajes y 16) Otros gastos. La composición de cada uno de estos grupos de gasto se muestra en el [anexo I](#).

Respecto a los distintos tipos de ingresos considerados para las familias, se estudió la clasificación propuesta en la III Encuesta Nacional de Presupuestos Familiares, la cual se caracteriza por siete grupos principales de fuentes de ingreso: 1) Sueldos y salarios nominales, 2) Transferencias recibidas en dinero, 3) Ingresos mensuales por renta de propiedades en el país, 4) Ingresos ocasionales, 5) Ingresos provenientes del exterior, 6) Ingresos por trabajo independiente y 7) Ingresos por misiones sociales.

5.3 Creación de la matriz de datos

La estructura final de la vista minable a utilizar para la extracción de patrones y conocimiento sobre el gasto familiar, se definió como una matriz cuyas columnas contemplan las siguientes variables: variables características de la familia como el código único del hogar, el dominio geográfico donde está ubicado el hogar, la entidad, tipo de vivienda, modo de propiedad de la vivienda, número de miembros del hogar y número de vehículos del hogar; las variables de gasto (16 en total) y las variables de ingreso familiar (7 en total). Cada fila de esta matriz representa a una familia diferente, y cada variable de gasto e ingreso contabilizan los montos totales por la familia en cuestión.

La obtención de esta matriz única de presupuesto familiar requirió un procedimiento específico para cada fuente primaria de datos considerada, dado que los datos se encontraban en ciertos casos dispuestos de modo diferente. Para

procesar cada fuente original de datos fue necesario crear rutinas y funciones (macros escritas en el lenguaje Visual Basic, en hojas de cálculo de Excel) para contabilizar de manera automática los montos totales para cada variable de gasto e ingreso para cada familia. En las siguientes secciones se describirá en detalle los procedimientos hechos para la obtención de esta matriz de presupuestos familiares.

5.3.1 Totalización de los montos de los gastos por familia entrevistada

De las fuentes primarias de datos se pudieron extraer tres tipos de gastos: gastos en bienes y servicios para el hogar, gastos necesarios para el establecimiento y administración de pequeñas y medianas empresas (Pymes) y trabajadores independientes y los gastos incurridos en establecimientos de producción agropecuaria, propiedades de los hogares entrevistados. Estos gastos estaban distribuidos en distintas tablas, por lo que el procedimiento para la totalización de dichos gastos fue hecho tabla por tabla.

5.3.1.1 Gastos en bienes y servicios del hogar

Los gastos en bienes y servicios para el hogar provienen de las tablas “gastos diarios”, “gastos diarios en comida”, “gastos mensuales”, “gastos trimestrales” y “gastos anuales”. En vista de que las tablas anteriores están bajo períodos de referencia distintos, fue necesario elegir un periodo de referencia en el cual se expresen todos los montos de gastos e ingresos. El periodo de referencia elegido como base para este estudio es el mensual, pues es muy representativo para muchos ingresos y gastos en el hogar.

Todos los gastos expresados en diferentes períodos de referencia fueron llevados a su equivalente mensual de la siguiente manera: los gastos con período de referencia diario fueron multiplicados por cuatro, pues estos gastos fueron registrados en la semana de la entrevista por los encuestados, de modo que todos constituyen un gasto semanal; los gastos mensuales no ameritaron cambio de escala; los gastos cuyo período de referencia es el trimestral fueron divididos entre tres para llevarlos a su equivalente mensual, mientras que los gastos anuales fueron divididos entre doce.

5.3.1.2 Gastos en establecimientos de trabajo independiente y Pymes

Muchos hogares entrevistados en la III Encuesta Nacional de Presupuestos Familiares poseen un establecimiento propio (abasto, cafetín, tienda de ropa, consultorio, pequeña empresa, etc.), del cual devengan gran parte de sus ingresos monetarios. Sin embargo, el establecimiento y la gestión diaria de estas unidades comerciales requieren de desembolsos monetarios periódicos, entre ellos pagos en alquileres de locales, equipos, pagos de impuestos (I.S.L.R), materiales e insumos necesarios para las actividades diarias de producción, venta y comercialización de productos. Las tablas que registran los gastos necesarios para el mantenimiento de establecimientos de trabajo independiente, propiedad de los hogares, fueron: "personas_ti4", "personas_ti5", "personas_ti6" y "personas_ti21".

Cabe destacar que los montos de gastos en cada una de las tablas pertenecían a periodos de referencia distintos, razón por la cual se diseñó una función en Visual Basic que permitiera identificar el periodo de referencia de cada bien adquirido (dependiendo de su código y el periodo asignado por la clasificación inicial del

BCV), para posteriormente llevar dichos montos a su equivalente mensual, mediante la multiplicación por el coeficiente indicado (si el bien era de referencia anual, el coeficiente valdría 1/12). Todos estos montos fueron distribuidos en los 16 grupos de gastos y posteriormente fueron totalizados por cada familia.

5.3.1.3 Gastos en establecimientos de producción agropecuaria

Entre los hogares entrevistados en la III ENPF se encontraban algunos que eran productores agropecuarios o se beneficiaban del cultivo de tierras, de la cría de animales, la pesca, etc. Todas estas unidades de producción hacen gastos en insumos necesarios para la producción, por ejemplo alimento para animales, abonos, fertilizantes, maquinarias, equipos, etc. Las tablas que contenían estos tipos de gastos eran “personas_ti19” y “personas_ti20”. Para la totalización de los gastos de estas tablas se siguió un procedimiento similar al de las tablas de gastos para establecimientos independientes, pues también había gastos de distintos períodos de referencia que debían ser expresados en su equivalente mensual.

Después de la totalización de gastos en cada uno de los 16 rubros por concepto de gastos en bienes y servicios para el hogar, gastos en establecimientos de trabajo independiente y gastos por producción agropecuaria, se procedió a totalizar los montos de cada uno de estos gastos para cada familia, creando una matriz de gastos familiares, donde cada fila representa a una familia diferente y cada variable de gasto representa el total desembolsado por la familia en dicho rubro. Este procedimiento también fue ejecutado con la ayuda de una rutina escrita en el lenguaje de programación Visual Basic.

5.3.2 Totalización de los ingresos por familia entrevistada

Del mismo modo que los gastos, los ingresos fueron expresados y totalizados bajo el período de referencia mensual. A partir de las fuentes primarias de datos se pudo recabar información acerca de ingresos por trabajo formal, ingresos por trabajo independiente, ingresos por trabajo en actividades agropecuarias e ingresos por becas otorgados por misiones de carácter social. Los datos de ingresos por trabajo formal se encontraban en la tabla “personas_i”, en la que se especifican montos de ingresos por sueldos y salarios, pensiones, ganancias por renta de propiedades, beneficios otorgados por empresas, etc. Para procesar estos datos fue necesario identificar los períodos de referencia de cada variable dentro de esta tabla, mediante el uso de las definiciones en el cuestionario EPF8: “Ingresos”, luego llevar los montos a sus equivalentes mensuales y totalizar todos estos datos para cada hogar, con la ayuda de una rutina dedicada para ello, escrita en una macro en el lenguaje Visual Basic.

Por otro lado, las fuentes de información sobre los ingresos por trabajo independiente fueron recabadas de las tablas “personas_ti3” y “personas_ti4”, mientras que los datos sobre ingresos provenientes del trabajo agropecuario provenían de la tabla “personas_ti8”. El procesamiento de los ingresos registrados en estas tablas requirió mucho tiempo debido a que las variables no tenían asignados períodos de referencia, situación que ameritó el tratamiento individual de cada registro de ingreso. Los periodos de frecuencia de los ingresos por trabajo independiente se especificaron en base a los siguientes criterios: ocupación desempeñada por la persona que registra el ingreso, número de horas de trabajo y el monto total de gastos en insumos, materiales y herramientas utilizadas en el establecimiento; mientras que para los ingresos por trabajo en el sector agropecuario

se consideraron los criterios: tipo de actividad agropecuaria, cantidad de área dedicada para la producción, número de animales en la producción y horas de trabajo diarias. Después de esto, los montos fueron llevados a su equivalente mensual y totalizados para cada familia. Respecto a los ingresos por becas de misiones de carácter social, estos montos estaban registrados en la tabla "personas_am2". El tratamiento de estos datos fue el mismo que el hecho con los ingresos por actividades agropecuarias, con la particularidad de que los períodos de referencia fueron asignados dependiendo del tipo de misión social.

Una vez contabilizados todos los montos totales de ingresos por trabajo formal, ingresos por trabajo independiente, ingresos por trabajo en actividades agropecuarias e ingresos por becas otorgados por misiones de carácter social, se totalizaron todos los montos de ingresos de cada familia entre las distintas variables de ingreso mencionados anteriormente, obteniendo una matriz de ingresos familiares, donde cada fila representaba los ingresos de cada hogar.

5.4 Integración de la información en una matriz de datos única

Una vez que se totalizaron todos los gastos e ingresos para cada hogar, el paso siguiente fue la creación de una matriz de datos única con información referente a las familias. Primero se contabilizaron el número de familias que poseían información sobre el dominio geográfico donde está ubicado el hogar, entidad, tipo de vivienda, modo de propiedad de la vivienda, número de miembros del hogar y número de vehículos del hogar; contabilizando 8.948 hogares. Posteriormente fueron contabilizadas las familias con valores registrados de gastos e ingresos, resultando

en 8.451 y 8.372, respectivamente. Toda esta información se unificó usando el campo clave CUH, que constituye el identificador único de cada hogar, variable creada al concatenar el código único de la vivienda con el número de hogar que tenía asignado la familia en cuestión. Para unificar los datos de los hogares sólo se consideraron las familias que tuvieran información sobre sus gastos, ingresos y sus características generales, obteniendo de esta manera una matriz con 8.372 filas (familias) y 32 columnas. Las variables que componen esta matriz única de presupuestos son:

- **Variables para la caracterización del hogar:** 1) Código único del hogar, 2) Dominio, 3) Estrato social, 4) Entidad, 5) Tipo de vivienda, 6) Modo de propiedad de la vivienda, 7) Número de miembros del hogar, 8) Número de vehículos del hogar.
- **Gastos del hogar:** 9) Alimentos y bebidas no alcohólicas, 10) Bebidas alcohólicas y tabaco, 11) Comidas fuera del hogar, 12) Vestido y calzado, 13) Vivienda y sus servicios, 14) Mobiliario y equipos del hogar, 15) Mantenimiento de la vivienda, 16) Salud, 17) Transporte, 18) Equipos y servicios de comunicaciones, 19) Educación y cultura, 20) Recreación y esparcimiento, 21) Artículos, efectos personales y servicios diversos, 22) Servicios financieros, tributarios y legales, 23) Viajes, 24) Otros gastos.
- **Ingresos del hogar:** 25) Sueldos y salarios nominales, 26) Transferencias recibidas en dinero, 27) Ingresos mensuales por renta de propiedades en el país, 28) Ingresos ocasionales y otros ingresos, 29) Ingresos provenientes del exterior, 30) Ingreso por trabajo informal, 31) Ingreso por trabajo independiente en el sector agropecuario y 32) Ingreso por misiones sociales.

Adicionalmente fue creada una variable llamada Ingresos/Gastos, cuyo propósito fue el de cuantificar el estado del presupuesto de la familia en el periodo de estudio: si el valor es menor a 1, significa que el presupuesto de la familia fue deficitario.

5.5 Manejo de datos faltantes

En la sección 5.3 se mencionó la acción tomada para el manejo de registros sin información de ingresos monetarios en siquiera una de las variables consideradas: descartar al hogar en cuestión (y a los valores de gastos registrados por dicha familia). Esto se justifica debido a que un hogar debe obtener, de cualquier manera, un ingreso (aunque sea pequeño) en alguno de los distintos períodos de referencia (diario, semanal, mensual o anual) para poder incurrir en gastos. En cuanto a los valores faltantes en las variables de gastos, estos fueron sustituidos por ceros pues se supuso que la falta del valor en alguna(s) de las variables de gasto se debe a que la familia no registró desembolsos en bienes y servicios pertenecientes a dichos rubros.

Por último, el manejo de los datos faltantes en las variables para la caracterización del hogar consistió en descartar del estudio a aquellos registros donde existiera ausencia en la información de al menos una de estas variables. Todas estas acciones hechas a la matriz de datos garantizan obtener una vista minable sin datos faltantes, no obstante, todavía falta validar una condición necesaria que debe cumplir esta matriz de datos: la validez de sus datos. En la próxima sección se abordará este aspecto.

5.6 Manejo de valores atípicos

Para identificar los valores atípicos presentes en las distintas variables de gasto e ingresos, se utilizaron gráficos de control univariante, obtenidos mediante el uso del paquete estadístico Minitab. Los registros de familias con valores atípicos detectados por los gráficos de control fueron descartados del estudio, esto motivado a que, a pesar de tener la certidumbre de que la mayoría se trataban de datos verídicos, éstos representaban casos absolutamente fuera de lo común: gastos en más de 2.500.000 Bs. mensuales por concepto de comidas fuera del hogar, gastos superiores a 16.500.000 Bs. mensuales por gastos en salud y 147.000.000 de Bs. mensuales por ingresos debido al trabajo independiente, entre otros casos. No resultó lógico sustituir estos valores atípicos con la media poblacional de la variable implicada, pues sería necesario modificar los valores de otras variables (sobre todo las de ingreso) para mantener las proporciones definidas por el comportamiento de gasto de la familia, con lo que irremediabilmente se modificaría el comportamiento de gasto de dicha familia, pues se modifican las magnitudes de gastos e ingresos.

Los gráficos de control permitieron detectar la existencia de 29 valores atípicos: 2 pertenecientes a la variable de gastos Alimentos y bebidas no alcohólicas, 1 en la variable Comidas fuera del hogar, 2 en la variable Vestido y calzado, 2 en Vivienda y sus servicios, 1 en Salud, 1 en la variable Transporte, 2 en la variable Equipos y servicios de comunicaciones, 1 en la variable de gastos Recreación y esparcimiento, 1 por Artículos personales, 1 por Servicios financieros, tributarios y legales y 1 por la variable Viajes. Respecto a los ingresos, fueron detectados 3 registros con datos atípicos en la variable de sueldos y salarios, 1 en la variable de transferencias recibidas en dinero, 3 en ingresos por la renta de propiedades, 2 en

Ingresos ocasionales, 2 de Ingresos del exterior, 1 en Ingreso por trabajo independiente, 1 en ingreso por trabajo del agro y 1 valor de ingreso por misiones.

Luego de descartar a los registros pertenecientes a las familias con los valores atípicos anteriores, sólo resta validar que los registros de las familias reflejen valores de presupuesto útiles para el presente estudio. Para ello se usó la variable de estatus del presupuesto (la relación Ingreso / Gasto familiar). Muchos de los hogares presentaban un presupuesto deficitario, pues sus gastos superaban ampliamente a su ingreso total; mientras que en otras familias los ingresos superaban en decenas de veces a los gastos totales en que incurrían. Las familias con una relación de ingresos/gastos extremadamente grande o extremadamente pequeña no constituyen buenos datos, ya que por un lado no representan comportamientos comunes dentro de la familia venezolana, y por el otro, pueden influir significativamente sobre la estimación de los parámetros de los modelos.

Para tratar con esto, se definieron umbrales de la variable Ingreso/Gasto, que especificarían el rango válido de esta variable y, por consiguiente, las familias que serían incluidas y descartadas de la vista minable. El rango considerado para la variable Ingreso/Gasto es $[0.20, 32]$, con lo que se incluyen en el estudio a aquellas familias cuyo gasto es 5 veces mayor a su ingreso hasta las familias cuyo ingreso total supera en 32 veces sus gastos. Bajo esta consideración, la vista minable final quedó conformada por 8.292 hogares.

De esta manera se concluyó la fase del preprocesamiento, la cual fue una etapa de suma importancia en el desarrollo del presente estudio sobre identificación de patrones de gastos familiares.

Capítulo 6

Identificación de patrones de gastos familiares

Este capítulo explica detalladamente los distintos análisis hechos en la primera etapa de experimentación: identificación y estudio de los patrones de gastos tanto en las familias venezolanas como en las familias de la ciudad de Mérida (municipio Libertador). En primer lugar se efectúa un análisis a las distintas variables de gasto, para posteriormente extraer grupos familiares que comparten valores de desembolsos similares, cuyos comportamientos definen a los distintos patrones de gastos familiares. Una vez definidos los patrones de gastos, se establecen comparaciones entre los modos de comportamiento del gasto existentes en las familias de la ciudad de Mérida respecto a los de la familia venezolana, identificando a su vez las características socio-económicas de las familias clasificadas bajo cada uno de los patrones de gastos. A continuación se explican detalladamente cada uno de los puntos mencionados anteriormente.

6.1 Análisis exploratorio de las variables de gasto

Una vez obtenida la vista minable o matriz de datos con los valores de las distintas variables socio-demográficas, variables de gasto y variables de ingreso para 8.292 familias entrevistadas en la III ENPF, se procedió a examinar el comportamiento general de las variables de gasto con el fin de verificar su adecuación a distintos métodos para la extracción de patrones de gastos. Los análisis realizados fueron hechos usando el paquete estadístico SPSS 17, los cuales se enfocaron en la verificación del supuesto de normalidad de las variables de gasto, el análisis de las interdependencias entre las variables reflejadas en la matriz de correlaciones y en verificar la adecuación de modelos factoriales en la representación de las variables de gasto. En las siguientes secciones se presentan los resultados obtenidos en los distintos análisis realizados.

6.1.1 Verificación del supuesto de normalidad en las variables de gasto

El cumplimiento del supuesto de normalidad en las variables de estudio posibilita el uso de una amplia gama de métodos estadísticos y la obtención de resultados de mayor validez. En el caso de las variables de gasto consideradas en este estudio, el test de normalidad de Kolmogorov-Smirnov arroja el rechazo de la hipótesis acerca de la normalidad de estas variables, hecho que se refleja en la siguiente tabla:

Rubros	Parámetros dist. Normal		Máximas diferencias			Estadístico	Sig.
	Media	Desv. Std.	Absoluta	Positiva	Negativa	K-S	
Alimentos	377.846,47	488.424,666	0,22	0,192	-0,22	19,995	0,00
Bebidas y tabaco	3.3320,2	127.142,037	0,397	0,29	-0,397	36,118	0,00
Restaurantes	70.221,63	131.895,084	0,297	0,219	-0,297	27,065	0,00
Vestido y calzado	151.638,14	325.371,969	0,321	0,242	-0,321	29,193	0,00
Vivienda y serv.	15.8798,2	231.350,626	0,246	0,197	-0,246	22,422	0,00
Mobiliario y equip.	59.948,34	198.375,882	0,381	0,32	-0,381	34,717	0,00
Mant_vivienda	50.709,37	145.007,095	0,363	0,283	-0,363	33,081	0,00
Salud	13.9625,3	267.380,181	0,301	0,21	-0,301	27,388	0,00
Transporte	302.097,76	79.0813,15	0,351	0,321	-0,351	31,983	0,00
Equip_comunic.	108.297,71	139.094,195	0,218	0,148	-0,218	19,861	0,00
Educación-cultura	96.976,01	222.437,757	0,331	0,257	-0,331	30,18	0,00
Recreación	47.864,86	174.614,581	0,392	0,285	-0,392	35,695	0,00
Artículos pers.	66.450,43	148.328,125	0,327	0,232	-0,327	29,784	0,00
Servicios legales	5.457,86	50.239,388	0,457	0,375	-0,457	41,591	0,00
Viajes	3.263,18	39.316,011	0,515	0,515	-0,467	46,928	0,00
Otros gastos	22.713,35	122.383,508	0,426	0,361	-0,426	38,827	0,00

Tabla 6.1: Test de Kolmogorov-Smirnov para las variables de gasto.

Los valores presentes en las columnas “Máximas diferencias” de la tabla anterior se refieren a las mayores diferencias absolutas, positivas y negativas entre la función de distribución de los datos de la muestra y la función de distribución normal con los parámetros especificados en la columnas “Parámetros dist. Normal”. La mayor diferencia absoluta se utiliza para calcular el estadístico Z de Kolmogorov-Smirnov, multiplicando dicha diferencia por la raíz cuadrada del tamaño muestral, para posteriormente comparar este valor Z con una tabla de valores críticos de Kolmogorov-Smirnov dado el tamaño muestral [3].

Como el valor p para estas pruebas resulta ser cero, se rechaza la hipótesis nula acerca de la igualdad de la función de distribución de la muestra con una función de distribución normal. Este resultado constituye una razón por la que

resultaría inapropiado utilizar métodos estadísticos de reducción de variables, entre ellos el análisis de componentes principales y el análisis factorial, pues estos métodos se basan en gran medida en el supuesto de normalidad.

6.1.2 Análisis de la matriz de correlaciones

En el capítulo 3 se mencionó la importancia de la presencia de altas correlaciones entre las distintas variables medidas para la aplicación de un modelo factorial. Una de las primeras formas para medir estas dependencias es mediante la matriz de correlación, en la cual el elemento (i, j) corresponde al coeficiente de correlación lineal entre las variables i y j .

En el caso de las 16 variables de gasto estudiadas en esta investigación, la matriz de correlación posee los siguientes valores:

	VG1	VG2	VG3	VG4	VG5	VG6	VG7	VG8	VG9	VG10	VG11	VG12	VG13	VG14	VG15	VG16
VG1	1	0,153	0,126	0,106	0,112	0,075	0,206	0,118	0,082	0,145	0,133	0,08	0,218	0,029	0,001	0,095
VG2	0,153	1	0,114	0,037	0,019	0,024	0,044	0,015	0,031	0,051	0,023	0,167	0,041	0,047	0,009	0,011
VG3	0,126	0,114	1	0,18	0,307	0,09	0,102	0,235	0,222	0,33	0,224	0,216	0,17	0,101	0,117	0,059
VG4	0,106	0,037	0,18	1	0,135	0,177	0,107	0,1	0,106	0,178	0,166	0,133	0,216	0,041	0,044	0,058
VG5	0,112	0,019	0,307	0,135	1	0,118	0,137	0,368	0,328	0,449	0,346	0,154	0,166	0,143	0,074	0,1
VG6	0,075	0,024	0,09	0,177	0,118	1	0,119	0,105	0,092	0,135	0,112	0,118	0,137	0,023	0,067	0,126
VG7	0,206	0,044	0,102	0,107	0,137	0,119	1	0,141	0,114	0,131	0,126	0,064	0,187	0,059	0,079	0,182
VG8	0,118	0,015	0,235	0,1	0,368	0,105	0,141	1	0,228	0,35	0,252	0,128	0,142	0,105	0,058	0,058
VG9	0,082	0,031	0,222	0,106	0,328	0,092	0,114	0,228	1	0,293	0,23	0,12	0,123	0,086	0,1	0,098
VG10	0,145	0,051	0,33	0,178	0,449	0,135	0,131	0,35	0,293	1	0,355	0,192	0,191	0,113	0,103	0,073
VG11	0,133	0,023	0,224	0,166	0,346	0,112	0,126	0,252	0,23	0,355	1	0,149	0,165	0,075	0,079	0,077
VG12	0,08	0,167	0,216	0,133	0,154	0,118	0,064	0,128	0,12	0,192	0,149	1	0,107	0,055	0,048	0,05
VG13	0,218	0,041	0,17	0,216	0,166	0,137	0,187	0,142	0,123	0,191	0,165	0,107	1	0,029	0,12	0,066
VG14	0,029	0,047	0,101	0,041	0,143	0,023	0,059	0,105	0,086	0,113	0,075	0,055	0,029	1	0,019	0,041
VG15	0,001	0,009	0,117	0,044	0,074	0,067	0,079	0,058	0,1	0,103	0,079	0,048	0,12	0,019	1	0,015
VG16	0,095	0,011	0,059	0,058	0,1	0,126	0,182	0,058	0,098	0,073	0,077	0,05	0,066	0,041	0,015	1

Tabla 6.2: Matriz de correlaciones de las variables de gasto.

La matriz de correlaciones muestra como las variables de gasto no presentan dependencias significativas en la mayoría de sus celdas, pues los coeficientes de correlación no sobrepasan un valor mínimo recomendado de 0,6. Las variables con correlaciones moderadas entre sí son: **comidas fuera del hogar, vivienda y sus servicios, salud, equipos y servicios de comunicaciones y la variable educación y cultura**. Correlaciones bajas entre las variables de gasto representan un claro indicio de la independencia entre éstas, por lo que no es recomendable representar a estas variables mediante un modelo factorial. Sin embargo, es necesario examinar los estadísticos de adecuación del modelo factorial a los datos y si la interpretación de los factores extraídos guarda algún grado de coherencia con el estudio de patrones de gasto.

6.1.3 Medidas de adecuación de los datos para el análisis factorial

El propósito de aplicar el análisis factorial radica en descubrir variables latentes, formadas a partir de las variables de gasto originales, que expliquen la mayor parte de la variabilidad de los datos y describan un conjunto de variables independientes entre sí (Pérez, 2.004). Si las variables de gasto originales gozan de un alto grado de adecuación para el modelo factorial, los factores extraídos pasarían a ser las variables de estudio en técnicas posteriores, basando así las interpretaciones y conclusiones subyacentes en estas nuevas variables. En el caso del análisis factorial, las medidas de adecuación generalmente se basan en contrastar la independencia entre las distintas variables, siendo el estadístico KMO y el test de esfericidad de Barlett las medidas más comúnmente utilizadas. En la tabla 6.3 se muestran los valores que toman estas dos medidas de adecuación del modelo factorial para los datos de las variables de gasto de las 8.292 familias incluidas en el presente estudio:

Medidas de adecuación del modelo factorial		
Medida de adecuación de Kaiser-Meyer-Olkin.		0,837
Test de esfericidad de Bartlett	Approx. Chi-Square	1.3142,151
	g.l	120
	Sig.	0,000

Tabla 6.3: Medidas de adecuación del modelo factorial.

A pesar de que la matriz de correlación de las variables de gasto arroja valores bajos de dependencia, la medida KMO y el test de Barlett indican un buen ajuste del modelo factorial a los datos, por lo que las variables de gasto pueden ser resumidas por un conjunto de factores comunes. El valor 0,837 del KMO es superior al mínimo recomendado (0,7), por lo que las correlaciones entre los factores únicos

de las variables de gasto son despreciables, mientras que el valor p del test de esfericidad de Barlett rechaza la hipótesis acerca de la igualdad del determinante de la matriz de correlación a la unidad, en cuyo caso las variables serían independientes.

6.1.4 Análisis de modelos factoriales para las variables de gasto

Para decidir la conveniencia de representar a las variables de gasto mediante un modelo factorial, se utilizaron diversos métodos de extracción de factores. Tal vez la consideración más importante a tomar en cuenta es la facilidad en la interpretación de los factores, de modo que ellos describan comportamientos no observados y relevantes para el estudio de patrones de gastos familiares. Uno de los primeros análisis a ejecutar es el de las comunalidades de las variables explicadas por el conjunto de factores extraídos. Como la comunalidad de las variables varía con el método de extracción, en la tabla 6.4 se comparan las comunalidades de las distintas variables dependiendo del método de extracción de factores: método de los componentes principales, método de la máxima verosimilitud, método de los mínimos cuadrados generalizados, método de los mínimos cuadrados no ponderados, método de factorización de imagen, método de factorización alfa y el método de factorización de ejes principales; para los primeros ocho factores.

	Componentes principales	Max. versos.	Min. Cuadr. grl.	Min. Cuadr. no pond.	Fact. Imagen	Fact. Alfa	Fact. Ejes princ.
Alimentos	0,658	0,478	0,478	0,31	0,09	0,476	0,431
Bebidas y tabaco	0,717	0,174	0,174	0,282	0,038	0,177	0,189
Restaurantes	0,474	0,298	0,301	0,328	0,186	0,291	0,289
Vestido y calzado	0,754	0,999	0,999	0,244	0,091	0,748	0,452
Vivienda y serv.	0,579	0,493	0,495	0,486	0,298	0,504	0,493
Mobiliario y equip.	0,923	0,12	0,12	0,23	0,063	0,124	0,136
Mant_vivienda	0,54	0,223	0,223	0,221	0,093	0,223	0,234
Salud	0,522	0,285	0,287	0,309	0,193	0,291	0,303
Transporte	0,405	0,216	0,218	0,214	0,155	0,217	0,221
Equip_comunic.	0,546	0,455	0,457	0,447	0,301	0,456	0,453
Educación-cultura	0,401	0,274	0,276	0,314	0,193	0,265	0,274
Recreación	0,568	0,334	0,336	0,198	0,088	0,328	0,312
Artículos pers.	0,573	0,228	0,23	0,271	0,12	0,236	0,239
Servicios legales	0,976	0,076	0,075	0,05	0,029	0,078	0,076
Viajes	0,914	0,178	0,176	0,998	0,026	0,193	0,252
Otros gastos	0,851	0,999	0,999	0,304	0,042	0,489	0,317
% Explicado	65,01%	36,44%	36,53%	32,54%	12,54%	31,85%	29,19%

Tabla 6.4: Comunidades explicadas por los primeros 8 factores bajo distintos métodos de obtención de factores.

En la tabla anterior se puede apreciar que el método de extracción por componentes principales explica la mayor proporción de las variables originales, es decir aquel donde los factores extraídos explican mayor comunalidad de las variables originales. Dado que las variables fueron tipificadas, la varianza de cada variable original equivale a la unidad, por lo que la variabilidad total es igual a 16 (el número de variables) y el porcentaje explicado es la división de la suma de las comunalidades de las variables entre 16. Cabe destacar que proporciones menores al 40% de variabilidad total, explicada por un conjunto de 8 factores, indican una baja adecuación del modelo, mientras que la calidad del modelo obtenido por la técnica de componentes principales es cuestionable. Es necesario evaluar los coeficientes de carga factorial presentes en cada factor para determinar si tienen alguna

interpretación lógica y coherente. A continuación se explican los resultados obtenidos por éste método de extracción de factores.

Modelo factorial: método de los componentes principales

Bajo este método, la cantidad de factores asociados a autovalores mayores a 1 son cuatro. Los factores extraídos presentan las proporciones dadas en la siguiente tabla:

Factor	Total	% de varianza	% acumulado
1	3,156	19,727	19,727
2	1,300	8,126	27,853
3	1,140	7,126	34,979
4	1,063	6,647	41,626
5	0,995	6,217	47,843
6	0,984	6,149	53,992
7	0,935	5,846	59,838
8	0,827	5,166	65,004

Tabla 6.5: Varianza explicada por los primeros 8 factores, método de los componentes principales.

Estos factores presentan los siguientes coeficientes o cargas factoriales:

Rubros	Factor							
	1	2	3	4	5	6	7	8
Alimentos	0,347	0,477	0,111	0,300	-0,409	-0,157	0,024	0,120
Bebidas y tabaco	0,156	0,305	0,706	0,191	0,008	0,161	-0,151	0,125
Restaurantes	0,566	-0,105	0,253	-0,105	-0,011	0,077	-0,024	-0,249
Vestido y calzado	0,387	0,274	0,000	-0,408	0,167	-0,295	0,344	-0,359
Vivienda y serv.	0,667	-0,331	-0,078	0,108	-0,012	-0,066	-0,042	0,039
Mobiliario y equip.	0,319	0,310	-0,197	-0,284	0,504	-0,084	-0,003	0,586
Mant_vivienda	0,364	0,434	-0,299	0,279	-0,137	0,174	-0,009	0,048
Salud	0,560	-0,266	-0,080	0,138	-0,095	-0,101	-0,024	0,304
Transporte	0,516	-0,238	-0,095	0,063	0,019	0,112	-0,195	-0,137
Equip_comunic.	0,684	-0,252	0,014	-0,011	-0,051	-0,087	-0,044	0,043
Educación-cultura	0,573	-0,173	-0,068	-0,024	-0,070	-0,168	-0,057	-0,021
Recreación	0,381	0,098	0,489	-0,146	0,329	0,016	-0,198	-0,070
Artículos pers.	0,436	0,376	-0,102	-0,241	-0,330	-0,048	0,234	-0,082
Servicios legales	0,225	-0,153	0,105	0,376	0,272	0,349	0,744	0,026
Viajes	0,214	0,021	-0,117	-0,456	-0,199	0,772	-0,086	0,053
Otros gastos	0,232	0,323	-0,370	0,369	0,423	0,141	-0,303	-0,360

Tabla 6.6: Matriz de carga factorial, método ACP.

En la tabla 6.6 se puede apreciar que los coeficientes de carga factorial no resultan ser del todo satisfactorios, pues las magnitudes en los factores 1, 2 y 4 no son lo suficientemente próximos a 1 o a -1; mientras que los factores 3, 5, 6, 7 y 8 resaltan solamente a una variable de gasto. En el primer factor, las cargas para las variables comidas fuera del hogar (restaurantes), vivienda y sus servicios, salud, transporte, equipos de comunicaciones y la variable educación y cultura tienen magnitudes moderadas, por lo que se podría interpretar a este factor como gastos en bienes y servicios básicos; los factores 2 y 4 presentan coeficientes con magnitudes bajas que dificultan su interpretación; el factor 3 sólo destaca a la variable bebidas alcohólicas y tabaco; el factor 6 destaca a la variable viajes; el factor 7 a la variable servicios financieros, tributarios y legales y los factores 5 y 8 resaltan ambos a la

variable mobiliario y equipos del hogar. Este modelo factorial no es apropiado para representar a las variables de gasto, debido a que sus factores son complejos de interpretar y el modelo no explica un porcentaje de variabilidad total superior al 70% para factores extraídos.

Los distintos resultados obtenidos en esta sección indican que no es recomendable usar un método de reducción de dimensionalidad (el análisis factorial) en las variables de gasto, debido a tres razones principales:

1. Las variables de gasto no están lo suficientemente correlacionadas entre sí, hecho que se muestra en los coeficientes de la matriz de correlación, por lo que cada una de las variables es representativa para el estudio y merece un análisis detallado. Como no existen grandes dependencias entre las variables, las comunalidades de las variables extraídas por los factores y la variabilidad total explicada por el modelo factorial no alcanza el porcentaje mínimo recomendable (70 %), y de hecho, está lejos de alcanzarlo.
2. Las variables de gasto no cumplen con el supuesto de normalidad, por lo que las interpretaciones hechas a los factores pueden ser inválidas.
3. De los ocho factores extraídos, cinco de ellos destacan únicamente a una sola variable de gasto, mientras que otro factor engloba a seis rubros de gasto distintos. Para el primer caso sería más apropiado estudiar cada variable de gasto que al factor como tal; mientras que en el segundo caso, se estaría perdiendo información valiosa si se sustituyen a los seis rubros de gasto por el factor correspondiente.

En esta sección se concluye que no es adecuado representar a las variables de gasto mediante un modelo factorial, por lo que la extracción y descripción de patrones de gasto presentes en las familias venezolanas estará basado en las 16 variables de gasto originales.

6.2 Metodología aplicada en la extracción de patrones de gastos en las familias venezolanas

Para generalizar los distintos comportamientos de gasto presentes en las familias venezolanas, se utilizaron los valores de desembolsos mensuales en 16 variables de gasto de una muestra de 8.292 familias entrevistadas en la III ENPF, agrupando a las observaciones que presentaban valores de desembolsos similares entre los distintos rubros a través del método de agrupamiento de las K-medias, disponible en el paquete estadístico Minitab 15. El método de K-medias permitió extraer los patrones de gastos en las familias venezolanas, los cuales están descritos por el comportamiento promedio del gasto en las familias pertenecientes a un mismo conglomerado, es decir, los valores que toma el centroide del conglomerado. En vista de que la agrupación hecha por el método de K-medias depende del orden de las observaciones dentro de la muestra, de la escala que presentan las variables de gasto y de la cantidad de particiones o conglomerados a extraer, las pruebas hechas para extraer los patrones finales de gasto fueron numerosas.

El método general para la extracción de patrones de gastos en las familias venezolanas consistió primero en evaluar el efecto de la estandarización de las variables de gasto en la formación de las particiones, verificando si este cambio de escala permitía obtener grupos más homogéneos de observaciones. Después de

varias pruebas para distintos números de particiones, se concluye que no es recomendable estandarizar los datos pues las variables de gasto tienen diferentes relevancias que condicionan los modos de gasto en las familias venezolanas. Al comparar la suma de cuadrados dentro de los grupos (SCDG) para las particiones obtenidas al estandarizar respecto a no estandarizar los datos, se puede notar que al estandarizar los datos los grupos son menos homogéneos, hecho que se puede apreciar en la siguiente tabla, donde se comparan las SCDG para 3, 4, 5 y 6 particiones:

N° particiones	SCDG de clasificación obtenida por datos sin estandarizar	SCDG de clasificación obtenida por datos estandarizados
3	6.298.112.151.777.940	8.447.531.874.571.670
4	5.483.786.844.397.700	7.723.025.125.446.920
5	4.814.207.984.746.820	6.990.924.020.302.490
6	4.502.338.478.009.610	6.372.872.614.430.280

Tabla 6.7: SCDG para particiones impuestas por el algoritmo K medias con y sin variables estandarizadas.

Trabajando con los datos sin estandarizar (pues se supone que la SCDG es menor que la obtenida al estandarizar los datos), el paso siguiente fue probar diversos números de particiones para distintos órdenes de aparición de las observaciones dentro de la muestra total, determinando el orden ideal dado un número de partición dado, es decir, aquel que presente la menor SCDG. Cabe destacar que el número de particiones consideradas para las pruebas tuvieron un límite de 6, pues para un mayor número de particiones no se obtenían grupos que tuvieran un porcentaje representativo de observaciones respecto a la muestra total. El número de particiones a considerar tampoco se basó en criterios experimentales como el sugerido por Hartigan, sino en encontrar aquella partición donde los

distintos conglomerados tuviesen un número considerable de observaciones, el comportamiento de gasto descrito por el centroide de un conglomerado fuera lo suficientemente distinto al de los demás conglomerados y la SCDG fuese mínima. A continuación se presenta una tabla que resume las primeras pruebas hechas y las SCDG obtenidas en cada prueba.

SCDG	N° de grupos			
	3	4	5	6
Normal	6.298.110 x10 ⁹	5.483.788 x10 ⁹	4.814.217 x10 ⁹	4.502.343 x10 ⁹
Inverso	6.298.100 x10⁹	5.483.778 x10 ⁹	4.814.207 x10 ⁹	4.456.129 x10 ⁹
Aleatorio 1	6.298.110 x10 ⁹	5.626.402 x10 ⁹	4.813.867 x10⁹	4.456.111 x10 ⁹
Aleatorio 2	6.298.110 x10 ⁹	5.483.688 x10 ⁹	4.814.217 x10 ⁹	4.456.116 x10 ⁹
Aleatorio 3	6.298.110 x10 ⁹	5.483.778 x10 ⁹	4.814.217 x10 ⁹	4.456.116 x10 ⁹
Aleatorio 4	6.298.110 x10 ⁹	5.483.600 x10 ⁹	4.814.217 x10 ⁹	4.456.107 x10⁹
Aleatorio 5	6.298.110 x10 ⁹	5.483.678 x10 ⁹	4.814.217 x10 ⁹	4.502.393 x10 ⁹
Aleatorio 6	6.298.110 x10 ⁹	5.483.688 x10 ⁹	5.776.264 x10 ⁹	4.456.129 x10 ⁹
Aleatorio 7	6.298.110 x10 ⁹	5.483.599 x10⁹	5.118.168 x10 ⁹	4.456.129 x10 ⁹
Aleatorio 8	6.298.110 x10 ⁹	5.908.681 x10 ⁹	4.814.217 x10 ⁹	4.456.115 x10 ⁹

Tabla 6.8: SCDG de distintos números de particiones, para distintos ordenes de las observaciones dentro de la muestra.

Las particiones de 3, 4, 5 y 6 conglomerados comparten una característica en común: poseen 2 conglomerados que incluyen a una gran parte de las observaciones de la muestra de estudio. A fin de descubrir comportamientos no observados en esta primera “división” de los datos, se aplicó nuevamente el método de agrupación de K-medias en cada uno de los dos conglomerados con mayor cantidad de observaciones, lo que permitió descubrir nuevas categorías representativas para el estudio de patrones de gastos. En las tablas 6.9 y 6.10 se resumen las SCDG obtenidas en esta segunda “división”, hechas a las observaciones del primer y segundo grupo más numerosos de las particiones originales de 3, 4, 5 y 6 conglomerados.

Partición original de pertenencia	N° de grupos para 2 ^{da} división		
	2	3	4
Partición de 3	1.318.255 x10 ⁹	1.191.920 x10 ⁹	1.130.917 x10 ⁹
Partición de 4	1.307.967 x10 ⁹	1.208.123 x10 ⁹	1.087.646 x10 ⁹
Partición de 5	992.559 x10 ⁹	894.286 x10 ⁹	822.196 x10 ⁹
Partición de 6	859.519 x10 ⁹	776.741 x10 ⁹	710.758 x10 ⁹

Tabla 6.9: SCDG para las divisiones del conglomerado con más observaciones en las particiones originales de 3, 4, 5 y 6 grupos.

Partición original de pertenencia	N° de grupos para 2 ^{da} división	
	2	3
Partición de 3	2.897.207 x10 ⁹	2.567.794 x10 ⁹
Partición de 4	2.169.134 x10 ⁹	1.834.078 x10 ⁹
Partición de 5	2.280.591 x10 ⁹	1.988.326 x10 ⁹
Partición de 6	1.729.372 x10 ⁹	1.601.515 x10 ⁹

Tabla 6.10: SCDG para las divisiones del segundo conglomerado con más observaciones en las particiones originales de 3, 4, 5 y 6 grupos.

Las tablas anteriores indican a primera vista que la SCDG al dividir los dos conglomerados con más observaciones es menor para las particiones iniciales de 5 y 6 conglomerados, por lo que la elección de un modelo final que generalice a los patrones de gasto debe basarse en estos dos tipos de particiones iniciales y a las posibilidades de segmentación de los dos conglomerados que agrupan la mayor cantidad de observaciones, de modo que los grupos finales sean representativos y la SCDG sea la mínima posible para el número final de grupos a fijar.

El porcentaje de reducción de la SCDG cuando se segmenta en 3 grupos y no en 2 al primer conglomerado con más observaciones se aproxima al 10%, mientras que la reducción debida a la división en 4 grupos y no en 3 alcanza el 8 %, razón suficiente para dividir a este conglomerado en más de 2 grupos. Respecto a la división del segundo conglomerado con más observaciones, al dividir en 3 grupos, uno de ellos no es representativo ya que incluye pocas observaciones. Los distintos

modelos formados por las particiones iniciales de 5 y 6 conglomerados se muestran en la tabla 6.11, en donde la suma de cuadrados dentro de los grupos equivale a la totalización de las SCDG resultante de la segmentación del primer y segundo conglomerado con más observaciones más las SCDG de los conglomerados iniciales restantes sin dividir. El número de conglomerados finales es igual a la suma de las particiones hechas para los dos conglomerados iniciales con mayor número de observaciones más el número de conglomerados restantes de la primera división. La nomenclatura de cada modelo que se muestra en la tabla 6.11 es de la forma $n,1:a,2:b$, donde n es el número de conglomerados iniciales en la primera partición, a es el número de particiones hechas al primer conglomerado con mayor número de observaciones y b es el número de particiones hechas al segundo conglomerado con más observaciones; de manera que 5,1:3,2:2 se refiere al modelo formado por dividir inicialmente en 5 conglomerados a la muestra total de observaciones, para posteriormente dividir en 3 grupos al conglomerado inicial con más observaciones y en 2 grupos al segundo conglomerado inicial con más observaciones.

Modelo	Nomenclatura	Nº total de grupos	SCDG
1	5,1:3,2:1	7	4.575.073 x10 ⁹
2	5,1:3,2:2	8	4.297.814 x10 ⁹
3	5,1:4,2:1	8	4.502.983 x10 ⁹
4	5,1:4,2:2	9	4.225.724 x10 ⁹
5	6,1:3,2:1	8	4.255.898 x10 ⁹
6	6,1:3,2:2	9	4.002.840 x10 ⁹
7	6,1:4,2:1	9	4.189.915 x10 ⁹
8	6,1:4,2:2	10	3.936.857 x10 ⁹

Tabla 6.11: SCDG de los modelos finales para la representación de patrones de gasto a nivel nacional.

Debido al compromiso entre la sencillez del modelo final y la relevancia de cada conglomerado para representar a un patrón de gastos familiar distinto al de los demás conglomerados, el modelo elegido para generalizar a los patrones de gastos en las familias venezolanas es el número 2, pues la reducción en la SCDG respecto al modelo 1 alcanza un 6,06 %, mientras que la diferencia en porcentaje de SCDG al elegir los modelos 3, 4 o 5 no sobrepasa el 1,68 %. A pesar de que los modelos 6, 7 y 8 presentan SCDG menores a la del modelo 2, se prefiere al modelo 2 pues todos estos modelos presentan un mayor número de conglomerados con pocas observaciones, varios de los cuales siguen algún comportamiento similar al de otros conglomerados. Siendo el modelo 2 la representación final de los distintos comportamientos de gastos familiares, los patrones de gastos fueron etiquetados alfabéticamente desde el conglomerado con más observaciones hasta aquel que posee la menor cantidad de observaciones, tal como se muestra en la tabla 6.12.

Patrón de gastos	N° total de observaciones	Porcentaje del total
A	3.410	41,12%
B	1.626	19,61%
C	1.271	15,33%
D	1.045	12,60%
E	527	6,36%
F	302	3,64%
G	71	0,86%
H	40	0,48%
Total	8.292	100,00%

Tabla 6.12: Patrones de gastos a nivel nacional y su presencia en la muestra de estudio.

6.3 Descripción de los patrones de gastos familiares a nivel nacional

El uso del algoritmo k-medias como método de agrupamiento permitió extraer ocho grupos de familias con comportamientos de gasto diferentes entre sí, tanto en las prioridades asignadas a los distintos rubros considerados (que se definen como las proporciones del gasto total), como en los montos mensuales promedio desembolsados por las familias en cada uno de ellos. A continuación se explica, a modo de ejemplo, la estructura del patrón de gastos tipo A en las familias venezolanas, que se caracteriza por ser el más común.

Patrón de gasto A: gastos de baja escala, enfocados principalmente en bienes y servicios prioritarios.

La estructura del gasto presente en las familias que presentan este patrón tiende a tener el siguiente orden de prioridades y magnitudes:

RUBRO	Promedio mensual (Bs.)	% del gasto
Alimentos y bebidas no alcohólicas	180.748	31,67%
Transporte	57.639	10,10%
Vestido y calzado	55.171	9,67%
Vivienda y sus servicios	51.158	8,96%
Equipos y servicios de comunicaciones	39.006	6,83%
Salud	28.757	5,04%
Artículos, efectos personales y servicios diversos	26.025	4,56%
Comidas fuera del hogar	25.117	4,40%
Mobiliario y equipos del hogar	22.769	3,99%
Educación y cultura	21.151	3,71%
Mantenimiento de la vivienda	18.968	3,32%
Bebidas alcohólicas y tabaco	18.323	3,21%
Recreación y esparcimiento	16.556	2,90%
Otros	8.129	1,42%
Servicios financieros, tributarios y legales	757	0,13%
Viajes	511	0,09%
TOTAL	570.786	100%

Tabla 6.13: Estructura del patrón de gastos familiar a nivel nacional tipo A.

Esta estructura indica que el comportamiento de las familias que presentan este patrón de gastos se caracteriza por la preferencia al consumo de bienes y servicios básicos que garanticen las condiciones mínimas requeridas para la vida en nuestra sociedad: Alimento, vestido, transporte y vivienda; siendo la prioridad fundamental el rubro de alimentos, que alcanza el 31,6% del monto total mensual desembolsado. Los desembolsos en cualquier rubro considerado son los menores en comparación con los demás patrones de gastos, sin embargo este comportamiento describe al 41,12% del total de familias encuestadas (lo que equivale a 3.410 familias). Los gastos en los rubros salud, equipos y servicios de comunicaciones y artículos de uso personal pasan a un segundo nivel de prioridad, superando en relevancia a los gastos en educación, comidas fuera del hogar, mantenimiento de la vivienda y mobiliario. Las familias que describen estructuras de gasto similares a la

de la tabla anterior destinan menos del 8% del presupuesto familiar a rubros no esenciales como bebidas alcohólicas y tabaco, recreación y esparcimiento, servicios financieros, tributarios, legales y viajes.

Analizando de manera similar las estructuras de los siete modos restantes de comportamiento familiar del gasto, los distintos patrones de gastos obtenidos en las familias venezolanas se caracterizan por las descripciones dadas en la siguiente tabla:

Patrón de gastos	Descripción	Gasto total mensual promedio (Bs.)
A	Desembolsos de baja magnitud, enfocados principalmente en bienes y servicios de primera necesidad.	570.786
B	Desembolsos moderados, con prioridad al rubro de alimentos, bienes y servicios básicos.	1.328.054
C	Desembolsos moderados con alta prioridad en la vivienda, la salud y servicios básicos.	1.554.728
D	Altos desembolsos en bienes y servicios básicos, la salud y la educación, con desembolsos moderados para los rubros restantes.	3.390.906
E	Gastos altos con prioridad a los rubros alimentos, bebidas no alcohólicas, vestido y calzado.	3.568.728
F	Gastos prioritarios y a gran escala en vehículos y servicios de transporte, con desembolsos moderados en los rubros restantes.	5.132.165
G	Gastos excesivamente altos en el rubro de transporte y desembolsos altos en los rubros restantes.	9.911.661
H	Gastos excesivamente altos en el rubro de alimentos y bebidas no alcohólicas, con desembolsos moderados para los rubros restantes.	7.428.135

Tabla 6.14: Descripciones resumidas de los patrones de gastos nacionales.

La tabla anterior indica, de modo general, las diferencias existentes entre los patrones de gastos existentes en las familias venezolanas. Al comparar las

proporciones del gasto total invertidas en cada una de las variables de gasto y determinar cuáles son los patrones con mayores proporciones por cada uno de los rubros considerados, se pueden identificar comportamientos distintivos para cada patrón de gastos. Al examinar la estructura de cada uno de los patrones de gastos en las familias venezolanas y comparar sus proporciones con las de los demás patrones, se pueden extraer varios veredictos:

1. Las familias clasificadas con el patrón de gastos tipo B son las que destinan mayor proporción de sus presupuestos a rubros de baja relevancia e incluso innecesarios: artículos y servicios de uso personal, bebidas alcohólicas, tabaco y otros. También son las familias cuya proporción de gasto dedicada al mantenimiento de la vivienda es el superior.
2. Las familias con mayor proporción de gasto en los rubros salud, equipos y servicios de comunicaciones y la vivienda y sus servicios son las que presentan el patrón de gastos tipo C.
3. Las familias cuyo patrón de gastos es D son las que otorgan mayor prioridad a los gastos en educación y cultura entre los distintos patrones de gastos. También son los que destinan mayor proporción de su presupuesto al consumo de comidas fuera del hogar (restaurantes, cafetines, puestos de comida rápida, etc.) y las que tienen mayor proporción de gastos en servicios financieros y en viajes.
4. Las familias con patrón de gasto E son las que destinan la mayor proporción del presupuesto hacia la adquisición de bienes para el hogar, vestido y calzado y equipos para la recreación y el esparcimiento.

5. Las familias con patrón de gastos tipo G son las que destinan la mayor proporción del gasto total al rubro de transporte.
6. Las familias con patrón de gastos H son las que destinan la mayor proporción del gasto total al rubro de alimentos y bebidas no alcohólicas.

Una comprensión detallada de cada una de las estructuras de gasto de estos modos de comportamiento familiar hacia el gasto se puede hallar en el anexo II.

6.4 Descripción de los patrones de gastos existentes en las familias de la ciudad de Mérida

Una de las características de la III Encuesta Nacional de Presupuestos Familiares es que la muestra considerada en la ciudad de Mérida (municipio Libertador), tuvo un factor de expansión. Con ello se pretende abarcar un número representativo de familias de la ciudad de Mérida para efectuar estudios socio-económicos especializados para el municipio y establecer comparaciones con resultados a nivel nacional. En la III ENPF, la muestra de hogares recabada en el estado Mérida consta de 369 familias, con 302 pertenecientes al municipio Libertador.

La identificación de patrones de gastos familiares mediante el procedimiento expuesto anteriormente para la formación de particiones o grupos de observaciones, permitió extraer siete modalidades de gasto de la muestra de 302 familias que viven en el municipio Libertador, cuyas proporciones se presentan a continuación:

Patrón de gastos	N° total de observaciones	Porcentaje del total
A	153	50,66%
B	57	18,87%
C	29	9,6%
D	28	9,27%
E	20	6,62%
F	12	3,97%
G	3	0,99%
Total	302	100,00%

Tabla 6.15: Patrones de gastos en las familias merideñas y su presencia en la muestra de estudio.

Cabe destacar que las familias clasificadas bajo la estructura de gasto tipo G describen un comportamiento muy particular de desembolso, con lo que esta estructura de gasto no representa a un patrón de gastos como tal dentro de las familias merideñas, pues como se verá posteriormente, estas familias fueron clasificadas bajo un patrón de gastos nacional que posee muy baja frecuencia y que se caracteriza por sus altos desembolsos.

Al examinar las magnitudes desembolsadas y las prioridades otorgadas a las distintas variables de gasto bajo cada uno de los patrones de gastos presentes en las familias merideñas, la descripción de cada uno de estos modos de comportamiento se presenta en la siguiente tabla:

Patrón de gastos	Descripción	Gasto total mensual promedio (Bs.)
A	Desembolsos de baja magnitud, enfocados principalmente en bienes y servicios de primera necesidad.	862.681
B	Desembolsos moderados en los rubros de alimentos, vivienda y gastos básicos.	2.143.610
C	Desembolso alto en alimentos, vestido y calzado y moderado para los rubros restantes.	4.055.013
D	Gastos prioritarios en rubros referentes a la salud y la vivienda. Desembolsos moderados en los rubros restantes.	2.291.902
E	Gastos prioritarios y a gran escala en vehículos y servicios de transporte, con desembolsos moderados en los rubros restantes	3.908.367
F	Altos gastos en el rubro de transporte, salud, alimentos y gastos moderados en el resto de rubros.	2.796.944
G	Gasto excesivamente alto en el rubro de transporte y desembolsos moderados en el resto de bienes y servicios.	9.048.020

Tabla 6.16: Descripciones resumidas de los patrones de gasto de las familias de la ciudad de Mérida, municipio Libertador.

Del mismo modo como se planteó en el caso de los patrones de gastos a nivel nacional, es importante conocer cuales patrones de gastos presentan las mayores proporciones de desembolso por cada tipo de rubro considerado, con lo se identifican comportamientos distintivos para cada patrón. Después de analizar las proporciones otorgadas a cada variable de gasto por cada uno de los patrones de gastos locales (véase anexo III), se pueden extraer los siguientes veredictos acerca del comportamiento de gasto de las familias en la ciudad de Mérida:

1. A pesar de que las familias clasificadas con el patrón de gastos tipo A son las que presentan el menor gasto total mensual entre todas las familias, son las que invierten la mayor proporción del gasto en equipos y servicios de comunicaciones que el resto de familias clasificadas con los demás patrones de gasto.

-
2. Las familias clasificadas con el patrón de gastos B son las que invierten la mayor proporción en alimentos, en educación y en cultura.
 3. Las familias clasificadas con el patrón de gastos C son las que invierten en mayor proporción de su gasto total mensual en rubros innecesarios como bebidas alcohólicas y tabaco y otros. Estas familias también destacan respecto a las demás por ser las que invierten en mayor proporción en el rubro de vestido y calzado.
 4. Bajo el patrón de gastos tipo D, las familias otorgan mayor importancia a los gastos referentes a la vivienda y el mobiliario respecto a las familias clasificadas con otros patrones de gasto. Las familias con patrón de gastos D son las que comen con mayor frecuencia fuera del hogar (restaurantes, cafeterías, sitios de comida rápida, etc.) y a su vez son las que invierten la mayor proporción en el rubro de salud.
 5. Las familias clasificadas bajo el patrón de gastos F presentan las mayores proporciones de gastos en rubros relacionados con el entretenimiento y disfrute personal: artículos y efectos personales, recreación y esparcimiento y el rubro viajes. Son las familias con la mayor proporción de gasto dedicada al mantenimiento de la vivienda y a los gastos en servicios financieros, tributarios y legales.
 6. Las familias que presentan el patrón de gastos tipo G son las que invierten el mayor monto y proporción en el rubro de transporte.

6.5 Análisis comparativo entre los patrones de gastos familiares existentes en la ciudad de Mérida y los extraídos a nivel nacional

Dado que la población de cada región geográfica dentro de Venezuela guarda consigo características específicas, basadas primeramente en costumbres y necesidades a satisfacer, en este estudio se consideró conveniente identificar tanto las similitudes como las diferencias en el comportamiento de gasto entre las familias de la ciudad de Mérida respecto al comportamiento de gasto a nivel nacional. Mediante la extracción de los patrones de gastos se identificaron estructuras que otorgan distintas proporciones y montos desembolsados a las distintas variables de gasto, siendo cada una de estas estructuras modos de comportamiento familiar respecto al gasto.

Como las familias entrevistadas en la ciudad de Mérida presentan dos tipos de clasificaciones, es decir, el patrón de gasto local y el patrón de gasto otorgado a nivel nacional, una aproximación inicial entre las similitudes que guardan estos dos tipos de clasificaciones se basa en el análisis de las proporciones de familias clasificadas con los patrones de gastos nacionales para cada uno de los patrones de gastos locales, porcentajes que se pueden apreciar en la tabla 6.17:

Patrones de gastos locales	Patrones de gastos nacionales							
	A	B	C	D	E	F	G	H
A	60,78%	11,11%	28,10%	0,00%	0,00%	0,00%	0,00%	0,00%
B	0,00%	49,12%	14,04%	33,33%	3,51%	0,00%	0,00%	0,00%
C	0,00%	3,45%	0,00%	10,34%	82,76%	0,00%	0,00%	3,45%
D	0,00%	0,00%	78,57%	21,43%	0,00%	0,00%	0,00%	0,00%
E	0,00%	0,00%	0,00%	25,00%	5,00%	70,00%	0,00%	0,00%
F	0,00%	0,00%	41,67%	58,33%	0,00%	0,00%	0,00%	0,00%
G	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%

Tabla 6.17: Tabla de equivalencias entre los patrones de gastos locales y nacionales.

La tabla anterior evidencia un mejor nivel económico de las familias de la ciudad de Mérida respecto a la familia venezolana, pues los patrones de gastos locales más frecuentes guardan una mayor relación con patrones de gastos nacionales que describen mayores desembolsos, un ejemplo de ello es el patrón de gastos local tipo A. Como los patrones de gastos tipo A corresponden a las familias con los menores desembolsos totales, registrados tanto a nivel nacional como en la ciudad de Mérida, la tabla 6.17 refleja como un porcentaje cercano al 40 % de las familias que fueron clasificadas con el patrón de gasto local tipo A presentan una clasificación distinta al patrón de gastos nacional tipo A, presentando estas familias niveles de desembolsos superiores. Del mismo modo, el patrón de gastos local tipo B, que corresponde al segundo grupo familiar con más observaciones para la ciudad de Mérida, solo equivale en un 49,12 % al patrón de gastos nacional tipo B, perteneciendo el porcentaje restante a patrones de gastos nacionales con mayores desembolsos, principalmente el patrón de gastos nacional tipo D. Por otro lado, las proporciones de familias merideñas clasificadas con los patrones de gastos nacionales tipo G y H respalda la afirmación acerca de las mejores condiciones económicas presentes en éstas familias, pues estos patrones constituyen comportamientos de gasto excesivo en los rubros transporte y alimentos, cuya frecuencia es muy baja.

Para identificar las similitudes existentes entre los patrones de gastos locales y los extraídos a nivel nacional, las comparaciones hechas entre los distintos modos de comportamiento de estos dos tipos de clasificaciones se centran en examinar las proporciones y los montos desembolsados en los distintos rubros entre los patrones de gastos con mayor similitud, por ejemplo, comparaciones entre el patrón de gastos local tipo D y el patrón de gastos nacional tipo C, o la comparación entre el patrón de gastos local tipo E con el patrón de gastos nacional tipo F. Mediante el análisis comparativo efectuado como se describe anteriormente, se extrajeron los siguientes veredictos:

1. A pesar de que los patrones de gastos clasificados con la letra A pertenecen a las familias con los menores desembolsos totales, las familias de la ciudad de Mérida incurren en mayores gastos: 862.681 Bs. mensuales vs. 570.786 Bs. mensuales promediados por las familias clasificadas con el patrón de gastos nacional tipo A. Mientras que las familias con los menores desembolsos a nivel nacional otorgan mayor prioridad a los bienes y servicios de primera necesidad, entre ellos alimentos, transporte y vestimenta; las familias que presentan los menores desembolsos en la ciudad de Mérida otorgan mayor prioridad a los gastos en alimentos, vivienda y sus servicios, equipos y servicios de comunicaciones que a los rubros transporte, vestido y calzado. Por otro lado, a nivel nacional existe la tendencia de las familias con los menores desembolsos a interesarse por el consumo de artículos de uso personal y para el mantenimiento de la vivienda, a diferencia de las familias que presentan los menores desembolsos en la ciudad de Mérida, las cuales tienen un mayor enfoque en bienes y servicios para la educación y cultura, la recreación y el esparcimiento.

2. Los patrones de gastos familiares tipo B presentan comportamientos distintos: mientras que las familias merideñas clasificadas con el patrón de gastos local tipo B priorizan altamente los gastos en alimentos, vivienda y sus servicios, la educación y cultura; las familias con el patrón de gastos nacional tipo B otorgan mayores prioridades a los rubros alimentos, vestido, calzado y transporte. Las familias merideñas invierten mayor proporción del gasto total en los rubros salud y recreación, a diferencia de la familia venezolana, cuyo interés de segundo orden se basa la adquisición de artículos personales, mobiliario y gastos en mantenimiento de la vivienda.
3. El patrón de gastos local tipo C describe una estructura de gasto similar a la del patrón de gastos nacional tipo E, pues las prioridades fundamentales para estos modos de comportamiento radican en los gastos en alimentos y bebidas no alcohólicas, vestido, calzado y transporte; mientras que los gastos en los rubros vivienda y sus servicios, salud y educación presentan desembolsos similares. Estos patrones de gastos se diferencian en mayor medida debido a la gran proporción del gasto total otorgada a los gastos en equipos y servicios de comunicaciones por parte de las familias con patrón de gastos local tipo C, mientras que en el patrón de gastos nacional tipo E es el desembolso en mobiliario y equipos del hogar la segunda prioridad de gasto. Otra diferencia remarcable es el mayor desembolso en comidas fuera del hogar por parte de las familias merideñas clasificadas con el patrón local tipo C, mientras que las familias con el patrón de gastos nacional tipo E invierten más en artículos, efectos personales y bienes para la recreación.
4. El patrón de gastos local tipo D y el patrón de gastos a nivel nacional tipo C priorizan los gastos en la vivienda, sus servicios y en los gastos para la salud, presentando a su vez proporciones similares del gasto total en los rubros

educación y cultura, recreación, mantenimiento de la vivienda y los gastos en vestido y calzado. A pesar de que el 78,57 % de las familias clasificadas con el patrón de gastos local tipo D fueron a su vez clasificadas con el patrón de gastos nacional tipo C, existen diferencias considerables entre estos dos comportamientos de gasto familiar: bajo el patrón de gastos local tipo D las familias tienden a tener mayores desembolsos en artículos y efectos personales, en comidas fuera del hogar, gastos en equipos y servicios de comunicaciones y gastos en mobiliario; mientras que las familias bajo el patrón de gastos nacional tipo C incurren en mayores gastos en el rubro transporte.

5. El patrón de gastos tipo E de las familias en la ciudad de Mérida y el patrón de gastos a nivel nacional tipo F se caracterizan por presentar desembolsos a gran escala en el rubro transporte, otorgando a su vez proporciones similares del gasto total a los siguientes rubros: artículos, efectos personales y servicios diversos, bebidas alcohólicas y tabaco, comidas fuera del hogar, equipos y servicios de comunicaciones, gastos en vestido, calzado y gastos para la salud. En cuanto a las diferencias, bajo el modo de comportamiento descrito por el patrón de gastos nacional tipo F existe la tendencia a un mayor gasto en alimentos, educación y cultura, la recreación y en los gastos concernientes a la vivienda, su mantenimiento y equipamiento que en el modo de comportamiento de gasto descrito por el patrón de gastos local tipo E.
6. Los patrones de gastos local tipo F y nacional tipo D destinan proporciones similares de gasto total a los rubros comidas fuera del hogar, equipos y servicios de comunicaciones, mantenimiento de la vivienda y servicios financieros, tributarios y legales. Estos modos de comportamiento de gasto otorgan mayor proporción del gasto a los rubros alimentos y bebidas no

alcohólicas, salud, transporte, vivienda y sus servicios; sin embargo, los montos desembolsados en alimentos y bebidas no alcohólicas, salud y vivienda son mayores bajo el patrón de gastos nacional tipo D, mientras que el desembolso en el rubro transporte es mucho mayor para las familias merideñas clasificadas con el patrón de gastos local tipo F. Otras diferencias existentes entre estos dos modos de comportamiento de gasto familiar radican en los mayores desembolsos en los rubros educación y cultura, mobiliario, vestido y calzado por parte de las familias con el patrón de gastos nacional tipo D; mientras que los gastos en artículos y efectos personales, recreación y viajes son superiores bajo el patrón de gastos tipo F en las familias merideñas.

Esta sección finaliza estableciendo como hechos comunes entre los distintos patrones de gastos familiares, tanto nacionales como locales, el alto nivel de prioridad que poseen los rubros alimentos y bebidas no alcohólicas, transporte, vestido y calzado, vivienda y sus servicios; del mismo modo como comparten los niveles más bajos de relevancia los gastos en servicios financieros, tributarios y legales, viajes, bebidas alcohólicas y otros gastos. En cuanto a las diferencias generales en el comportamiento de gasto entre la familia merideña y la familia venezolana, destaca la tendencia de las familias merideñas a presentar mayores desembolsos en bienes y servicios para la educación y la cultura, gastos en equipos y servicios de comunicaciones, gastos en la vivienda y sus servicios y en artículos para la recreación y el esparcimiento, hechos característicos en una ciudad con gran población estudiantil; mientras que la familia venezolana tiene la tendencia al gasto en artículos, efectos personales y servicios diversos, vestido y calzado, mobiliario, equipos para el hogar y bienes necesarios para el mantenimiento de la vivienda.

6.6 Características socio-económicas de las familias clasificadas bajo los distintos patrones de gastos familiares

Recordando que el comportamiento de gasto familiar está condicionado por factores económicos, sociales, geográficos, etc.; en este estudio se consideró la inclusión de distintas variables socio-demográficas y variables de ingreso para determinar su relación con cada patrón de gastos familiar. Comportamientos de gasto distintos deben ser resultado de características familiares diferentes, por lo que la caracterización de los perfiles familiares permite verificar, en cierto modo, la validez de cada uno de los patrones de gastos familiares extraídos.

En el presente estudio se caracterizaron los perfiles familiares, correspondientes a cada uno de los patrones de gastos, a través del análisis de estadísticas descriptivas en las variables de ingreso y las variables socio-económicas medidas de la III ENPF. En las siguientes secciones se describen, de manera general, las principales características de las familias clasificadas bajo los patrones de gastos nacionales y locales. Una mejor descripción acerca de las frecuencias y montos promedio de ingreso en cada uno de estos grupos familiares puede verse en los anexos [II](#) y [III](#).

6.6.1 Patrones de gastos familiares y sus perfiles familiares asociados

Las características sociales y económicas de las familias venezolanas que describen comportamientos de gasto similares a nivel nacional, es decir, aquellas clasificadas bajo los mismos patrones de gastos, se resume a continuación:

Patrón de gastos nacional tipo A: Desembolsos de baja magnitud, enfocados principalmente en bienes y servicios de primera necesidad.

- **Perfil familiar asociado:** Familias pertenecientes en su mayoría a los estratos bajos de la sociedad. Por lo general viven en casas (71,9 %) y no poseen vehículo (70,8 %). Sus mayores fuentes de ingreso monetario son sueldos y salarios, transferencias recibidas en dinero e ingresos por trabajo independiente, alcanzando promedios mensuales aproximados de 741.200 Bs., 334.400 Bs. y 2.065.200 Bs., respectivamente. Reciben beneficios por misiones sociales.

Patrón de gastos nacional tipo B: Desembolsos moderados, con prioridad en el rubro de alimentos, bienes y servicios básicos.

- **Perfil familiar asociado:** Familias con tendencia a pertenecer a las clases media y baja de la sociedad. Viven en ciudades medianas y en ciudades de gran importancia, con residencia generalmente en casas (73,9 %) y en menor medida en apartamentos (11,1 %). Cerca de la mitad de estas familias poseen por lo menos un vehículo (56,4 %). Son familias con el mayor número de miembros (5,1 personas en promedio). Las principales fuentes de ingreso son sueldos y salarios con 1.268.000 Bs. mensuales y el aporte por trabajo independiente, que en promedio alcanza los 2.821.500 Bs. mensuales en las familias que perciben este tipo de ingreso.

Patrón de gastos nacional tipo C: Desembolsos moderados con alta prioridad en la vivienda, la salud y los servicios básicos.

- **Perfil familiar asociado:** Familias pertenecientes en su mayoría a la clase media y media alta. Viven en el área Metropolitana de Caracas y ciudades importantes. Los apartamentos son el tipo de vivienda de preferencia (55,3 % de los casos). Alto porcentaje de estas familias pagan alquiler de la vivienda (27 %). El núcleo familiar es reducido, promediando 3,6 personas. Más de la mitad de estas familias poseen al menos un vehículo (63,3 % de los casos). Los ingresos por concepto de sueldos y salarios tienden a un total mensual de 1.712.600 Bs., seguido por la contribución debido a transferencias recibidas en dinero e ingresos ocasionales, que superan en frecuencia observada a los ingresos por trabajo independiente.

Patrón de gastos nacional tipo D: Altos desembolsos en bienes y servicios básicos, la salud y la educación, con desembolsos moderados para los rubros restantes.

- **Perfil familiar asociado:** Familias con tendencia a pertenecer a las clases alta y media alta sociedad, viven en su mayoría en las ciudades importantes del país. El 48,5 % viven en apartamentos de edificios, seguidos por un 27,2 % de familias que viven en casas. El 86,9 % de las familias tienen al menos un vehículo, siendo los casos más comunes la posesión de un vehículo (43,9 %) o dos (31,9%). Los ingresos monetarios totales por sueldos y salarios ascienden a una media de 3.320.400 Bs., cantidad comparable con el promedio en el ingreso por trabajo independiente, solo que este tipo de ingreso es menos frecuente. Destaca una considerable contribución en el ingreso total familiar los beneficios provenientes de ingresos ocasionales y transferencias en dinero.

Patrón de gastos nacional tipo E: Gastos altos con prioridad a los rubros alimentos, bebidas no alcohólicas, vestido y calzado.

- **Perfil familiar asociado:** Familias que viven en ciudades importantes o ciudades medianas, ocupando en su mayoría los estratos bajos y medios de la sociedad. Generalmente viven en casas o apartamentos de edificios. El número promedio de miembros de la familia es de 4,96 personas y el 63,8 % de estas familias tiene un vehículo como mínimo. Aproximadamente la mitad de estas familias tiende a percibir grandes ingresos monetarios por concepto de trabajo independiente, promediando 4.532.000 Bs. mensuales. El monto promedio de ingreso por sueldos y salarios alcanza el promedio de 2.386.400 Bs. mensuales.

Patrón de gastos nacional tipo F: Gastos prioritarios y a gran escala en vehículos y servicios de transporte, con desembolsos moderados en los rubros restantes.

- **Perfil familiar asociado:** Familias con tendencia a vivir en ciudades importantes (56,6 %) y en el área Metropolitana de Caracas (25,8 %). En su mayoría estas familias pertenecen a las clases alta y media alta de la sociedad, con proporciones de 41,1 % y 28,1 % respectivamente. Existe un elevado porcentaje de estas familias que viven en quintas (24,5 %), seguido de las familias que viven en apartamentos de edificios (41,4 %) o en casas (33,1 %). Son las familias con mayor propensión a la posesión de un vehículo (96,7 %). Generalmente estas familias poseen dos vehículos. Los ingresos por sueldos y salarios ascienden a una media de 3.087.000 Bs. mensuales para el 82,45 % de estas familias. Las fuentes alternativas de ingresos de estas familias son el aporte monetario por trabajo independiente, los ingresos ocasionales y las

transferencias recibidas en dinero, que promedian 2.914.500 Bs., 1.498.000 Bs. y 1.234.200 Bs. mensuales.

Patrón de gastos nacional tipo G: Gastos excesivamente altos en el rubro de transporte y desembolsos altos en los rubros restantes.

- **Perfil familiar asociado:** Familias que residen en su mayoría en ciudades importantes (60,6 %) y en el área Metropolitana de Caracas (23,9 %). Son en su mayoría familias de la clase alta (87,3 %). Gran parte vive en quintas o en apartamentos. Todas estas familias poseen vehículo, siendo los casos más comunes la posesión de dos o de tres vehículos. Los ingresos por sueldos y salarios promedian la alta cantidad de 4.921.300 Bs. mensuales y los ingresos por trabajo independiente 5.752.000 Bs. mensuales.

Patrón de gastos nacional tipo H: Gastos excesivamente altos en el rubro alimentos y bebidas no alcohólicas, con desembolsos moderados para los rubros restantes.

- **Perfil familiar asociado:** La mayoría de estas familias viven en ciudades importantes o en ciudades medianas. La clase social de este grupo de familias es algo heterogénea pues 57,5 % pertenece a las clases bajas, el 17,5 % a la clase media y el 25 % pertenece a las clases altas. Las viviendas de estas familias son por lo general casas. El 70 % de estas familias posee un vehículo como mínimo. La principal fuente de ingresos monetarios proviene del trabajo independiente, que en promedio alcanza 10.811.000 Bs. mensuales.

6.6.2 Patrones de gastos existentes en la ciudad de Mérida y sus perfiles familiares asociados

Las principales características sociales y económicas de las familias clasificadas bajo los patrones de gastos extraídos en la ciudad de Mérida, municipio Libertador, se resumen a continuación:

Patrón de gastos local tipo A: Desembolsos de baja magnitud, enfocados principalmente en bienes y servicios de primera necesidad.

- **Perfil familiar asociado:** Familias que viven en su mayoría en apartamentos (64,1 %) o en casas (26,1 %). Un número considerable de estas familias carece de vehículo (56,2 %). Las principales fuentes de ingreso monetario son los ingresos por sueldos y salarios y las transferencias recibidas en dinero, alcanzando los promedios de 1.130.519 Bs. y 574.000 Bs., respectivamente. Solo el 15 % de estas familias desempeña algún tipo de trabajo por cuenta propia y cerca del 7 % trabaja en actividades agropecuarias.

Patrón de gastos local tipo B: Desembolsos moderados en los rubros de alimentos, vivienda y gastos básicos.

- **Perfil familiar asociado:** La mayoría de estas familias habitan en apartamentos de edificios (54,4 %), en quintas (19,3 %) o en casas (17,5 %). Generalmente la familia es propietaria de la vivienda (73,7 %). La disponibilidad de estas familias de poseer al menos un vehículo aumenta al 75,4 %. El promedio de miembros del núcleo familiar es 4,16. La principal fuente de ingreso proviene de los sueldos y salarios, cuyo promedio mensual alcanza los 2.073.300 Bs. en el 78,95 % de estas familias. Los ingresos por

transferencias recibidas en dinero e ingresos ocasionales presentan frecuencias de 56,14 % y 43,86 %, respectivamente. El 26 % de estas familias perciben ingresos por concepto de trabajo independiente. Estas familias no se desempeñan en actividades agropecuarias ni reciben ingresos por misiones de carácter social.

Patrón de gastos local tipo C: Desembolso alto en alimentos, vestido y calzado; con desembolsos moderados para los rubros restantes.

- **Perfil familiar asociado:** Estas familias habitan en su mayoría en apartamentos de edificios (48,3 %), en casas o en quintas (20,7 % para las opciones anteriores). En el 69 % de los casos la familia es propietaria de la vivienda. El porcentaje de estas familias que carece de vehículo propio es de 24,1 %, mientras que el 31 % de estas familias posee un vehículo y el 34,5 % posee dos vehículos. Son familias numerosas (con 4,21 miembros en promedio), que devengan la mayor parte de sus ingresos monetarios debido a sueldos y salarios (3.269.400 Bs. mensuales en promedio) y al trabajo independiente (6.772.900 Bs. mensuales como promedio). Las fuentes alternativas de ingreso para estas familias son las transferencias recibidas en dinero y los ingresos ocasionales, sin embargo, dado el elevado porcentaje de familias que perciben ingresos por trabajo independiente (48,28 %), estas familias tienen la tendencia a dedicarse a áreas como el comercio de bienes y servicios.

Patrón de gastos local tipo D: Gastos prioritarios en rubros referentes a la salud y la vivienda. Desembolsos moderados en los rubros restantes.

- **Perfil familiar asociado:** La mayoría de estas familias habitan en apartamentos de edificios (71,4 %). La mitad de estas familias habitan en viviendas bajo condiciones de alquiler. Generalmente se trata de familias con tres miembros. El 53,6 % de las familias poseen solo un vehículo y el 10,7 % poseen dos vehículos. Estas familias reciben sus ingresos monetarios en mayor medida por concepto de sueldos y por transferencias recibidas en dinero, alcanzando promedios mensuales de 1.808.900 Bs. y 1.165.400 Bs., respectivamente. Cerca del 32 % de estas familias perciben ingresos por su desempeño en el trabajo independiente.

Patrón de gastos local tipo E: Gastos prioritarios y a gran escala en vehículos y servicios de transporte, con desembolsos moderados en los rubros restantes.

- **Perfil familiar asociado:** Estas familias viven en su mayoría en apartamentos de edificios (65 %) y son propietarios de la vivienda en el 70 % de los casos. El número promedio de miembros en estas familias es 3,5. La mitad de este grupo de familias carece de vehículo, mientras que 30 % restante posee un vehículo y el otro 20 % poseen dos vehículos. Los ingresos monetarios para estas familias provienen mayoritariamente de sueldos y salarios, alcanzando un promedio mensual de 3.383.200 Bs. La contribución de los ingresos monetarios debido al trabajo independiente solo abarca al 15 % de las familias, razón suficiente para inferir que se dedican muy poco a las actividades del comercio de bienes y servicios. Alrededor del 50 % de las familias gozan de altos ingresos por concepto de transferencias recibidas en dinero e ingresos ocasionales.

Patrón de gastos local tipo F: Altos gastos en el rubro de transporte, salud y alimentos. Gastos moderados en el resto de rubros.

- **Perfil familiar asociado:** La mayoría de estas familias viven en apartamentos de edificios (66,7 %) o en casas (16,7 %). Generalmente estas familias son dueños de la vivienda (75 %) o pagan algún tipo de alquiler o arrendamiento (25 %). La disponibilidad de vehículo propio es superior para estas familias, pues el porcentaje de familias que carecen de automóvil se reduce a 8,3 %. Las principales fuentes de ingreso monetario son los sueldos y salarios, que promedian 3.654.800 Bs. mensuales; y las transferencias recibidas en dinero, que alcanzan un promedio de 1.264.800 Bs. mensuales. Alrededor del 40 % de estas familias perciben beneficios por ingresos monetarios ocasionales, porcentaje superior al 25 % de familias que se dedican al trabajo independiente. Estas familias no reciben ingresos por misiones de carácter social, por trabajo en el sector agropecuario ni por ingresos provenientes del exterior.

Patrón de gastos local tipo G: Gasto excesivamente alto en el rubro de transporte y desembolsos moderados en el resto de bienes y servicios.

- **Perfil familiar asociado:** Viven en apartamentos de edificios o en casas. Son familias con pocas personas, pues el número de miembros varía entre 2 y 4. Todas las familias poseen por lo menos un vehículo. La principal fuente de ingreso monetario proviene de los sueldos y salarios, que en promedio alcanzan los 3.543.800 Bs. Estas familias no reciben ingresos por misiones de carácter social, ni por trabajo independiente (tanto en el sector del comercio

como en el sector agropecuario), y tampoco por ingresos provenientes del exterior.

Como se pudo apreciar en las secciones anteriores, los distintos patrones de gastos familiares, tanto locales como nacionales, describen perfiles familiares que guardan características particulares. Además, los valores de las distintas variables socio-económicas de cada uno de los grupos familiares guardan una correspondencia lógica con cada uno de los patrones de gastos: las familias pertenecientes a los estratos bajos de la sociedad y con menores ingresos describen comportamientos de gasto basados en los patrones de gastos A, B o C; mientras que las familias con mejores condiciones socio-económicas son clasificadas, con mayor frecuencia, bajo patrones de gastos de gran desembolso. De esta manera, esta sección concluye tanto con la validez de los patrones de gastos extraídos como en las diferencias que guardan entre sí las familias clasificadas bajo distintos patrones de gastos.

Capítulo 7

Experimentación con máquinas de vectores de soporte

Una vez obtenidos los distintos patrones de gastos tanto a nivel nacional como en las familias de la ciudad de Mérida, municipio Libertador, se procedió a la experimentación con Máquinas de Vectores Soporte, con la finalidad de construir modelos para la clasificación. Los patrones de gastos representan una clasificación extraída del comportamiento de gasto observado en las familias, por lo que un estudio sobre identificación de patrones de gastos debe considerar la construcción de modelos de clasificación que permitan asignar o predecir el modo de gasto de una familia conociendo los valores que toman sus variables de gasto, es decir los desembolsos totales en cada variable considerada.

Dada la naturaleza de la variable de salida (los patrones de gastos familiares), en el presente estudio se experimentó con modelos de MVS basados en el kernel lineal, el kernel de función de base Radial (Radial Basis Function) y el kernel polinomial, descartando de esta manera la experimentación con el kernel sigmoideal debido a su alta característica no lineal. La experimentación bajo esta técnica fue llevada a cabo utilizando la librería LIBSVM versión 2.88, la cual incluye una

recopilación de scripts y código escrito en varios lenguajes de programación (C++, Java, Python, etc.), que implementan esta técnica de aprendizaje automático así como también herramientas para el muestreo (tanto aleatorio como el estratificado) y la búsqueda de parámetros de los modelos.

7.1 Metodología para la experimentación con MVS

La obtención de modelos basados en MVS, cuya clase de salida son los patrones de gastos obtenidos, sigue un procedimiento genérico que comprende las siguientes cuatro etapas: 1) transformación de los datos de la vista minable al formato aceptado por el software utilizado, 2) creación de muestras de entrenamiento y de validación, 3) búsqueda de los parámetros más apropiados del modelo MVS para cada muestra de entrenamiento y la función núcleo asociada y 4) entrenamiento y validación del modelo MVS con los parámetros obtenidos en la etapa anterior. A continuación se explican con mayor detalle las consideraciones más relevantes.

7.1.1 Creación de muestras de entrenamiento y de prueba

Luego de transformar los datos al formato aceptado por la librería LIBSVM, el siguiente paso en la experimentación consistió en crear muestras de entrenamiento y de validación, para así determinar diferentes modelos de clasificación basados en MVS, evaluando el desempeño de la combinación de los valores de los distintos parámetros a considerar. El primer paso para extraer las muestras es determinar cuál será el número de observaciones a considerar por cada muestra, cifra que depende tanto de la precisión que se desea obtener en las estimaciones como del tipo de muestreo a ejecutar. En la selección de muestras útiles para el entrenamiento con MVS, con los patrones de gastos familiares como clase de salida, es necesario

conservar dentro de cada una de las muestras las proporciones observadas en la vista minable de estos patrones de gastos, razón para elegir al muestreo aleatorio estratificado como metodología para la selección de observaciones.

Al identificar a los patrones de gastos como los estratos para el muestreo, éstos pueden ser considerados como provenientes de una distribución multinomial, por lo que la elección del tamaño muestral debe considerar las distintas proporciones de los patrones de gastos. En este estudio se determina el tamaño muestral en base a las recomendaciones hechas por Tortora para el muestreo en poblaciones multinomiales (véase Medina, 2.002), quien plantea la ecuación 7.1 para determinar el número mínimo de observaciones necesarias para que la muestra guarde una proporción representativa de un estrato dado:

$$n_i = \frac{B * p_i * (1 - p_i)}{d^2} \quad (7.1)$$

En la expresión 7.1, p_i es la probabilidad del i -ésimo estrato (que es estimada con la proporción observada del i -ésimo estrato o patrón de gastos), d es el error máximo de estimación deseado en el estudio y B es el número que tiene un valor de α/N de probabilidad en la cola derecha de una distribución Chi-cuadrado con un grado de libertad ($N =$ número de estratos, $1 \leq i \leq N$). El número de observaciones a considerar en la muestra es el número máximo requerido entre los distintos estratos, es decir, el tamaño muestral equivale a $MAX\{n_i\}$.

Al obtener el número de observaciones por cada muestra dependiendo del error máximo de estimación que se esté dispuesto a incurrir, se extrajeron muestras de la vista minable mediante el uso de una herramienta para el muestreo aleatorio

estratificado, incluida en la librería LIBSVM 2.88. Adicionalmente se crearon muestras con valores de desembolsos escalados en el rango $[0,1]$, dado que varios autores recomiendan el cambio de escala en los datos de entrada para mejorar la precisión de los modelos basados en MVS. Para el cambio de escala en los datos de entrada se usó un script incluido en la librería LIBSVM 2.88.

7.1.2 Búsqueda de los parámetros apropiados del modelo

Dependiendo del tipo de función kernel que se use en el modelo de clasificación de MVS, será necesario ajustar distintos parámetros que regulan tanto el aprendizaje como el desempeño del modelo. En el caso de los modelos basados en el kernel lineal, solo es necesario definir el parámetro de regularización C , mientras que para los modelos basados en las funciones kernel de Base Radial, aparte del parámetro de regularización C , es necesario fijar el valor del parámetro Gamma (γ), el cual se encarga de regular la amplitud de la función de Base Radial. Por otro lado, los modelos basados en funciones kernel polinomiales son algo más complejos, pues es necesario fijar cuatro parámetros: el parámetro de regularización C , el parámetro de amplitud γ , el grado de la función polinomial y el coeficiente o constante en la expresión polinomial (que generalmente se fija en el valor 1).

Para agilizar la búsqueda de los parámetros en los distintos modelos de MVS, en este estudio se utilizó una herramienta de búsqueda de rejilla (grid search) implementada en la librería LIBSVM, la cual permitía obtener los valores más apropiados de los parámetros C y γ , evaluando el desempeño de las posibles configuraciones de estos parámetros a través de la validación cruzada. Una de las características de la búsqueda de rejilla implementada en la librería LIBSVM 2.88, es que genera un gráfico de contornos para el desempeño de las distintas

configuraciones de parámetros C y γ , representando en el espacio bidimensional las zonas que guardan las configuraciones con mejor desempeño. En la figura 7.1 se muestra un gráfico de contornos obtenido mediante la ejecución del script para la búsqueda de rejilla:

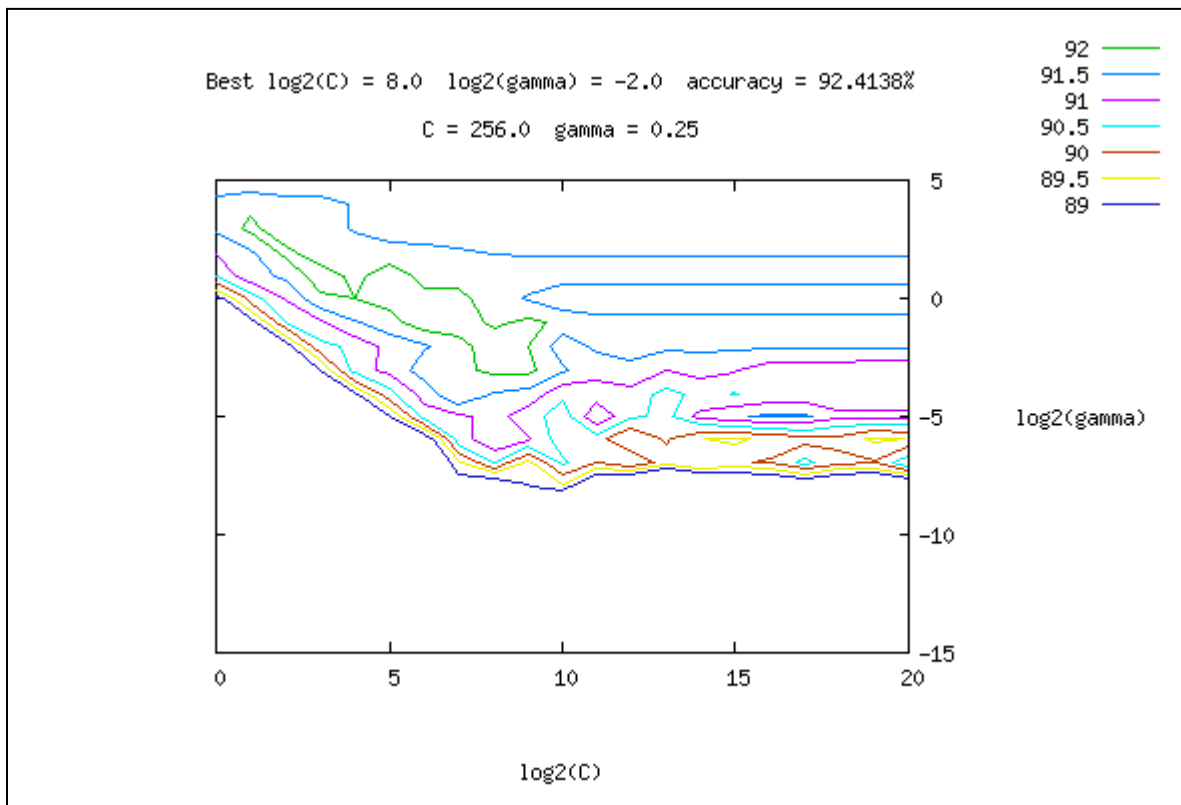


Figura 7.1: Ejemplo de gráfico de contornos. Búsqueda de rejilla en el espacio dado por los intervalos $[2^0, 2^{20}]$ para C y $[2^{-15}, 2^5]$ para γ .

A pesar de que esta herramienta de búsqueda resulta sumamente útil para los distintos entrenamientos y pruebas con modelos de MVS, solamente sirve para encontrar los valores de los parámetros C y γ , por lo que en el caso de los modelos basados en el kernel polinomial, el mejor grado y el mejor coeficiente fueron obtenidos mediante varias pruebas para configuraciones propuestas de estos dos

parámetros: pruebas con funciones de grado 2, 3, 5, 7 y 9; y pruebas con los coeficientes 1, 10, 50, 100 y 1.000.

7.1.3 Procedimiento general para la clasificación: entrenamiento y validación

Después de obtener la configuración ideal de los parámetros del modelo según el tipo de kernel mediante búsquedas de rejilla, el paso siguiente consistió en la construcción del modelo de clasificación a partir de la muestra de entrenamiento y la configuración de parámetros propuesta en la búsqueda de rejilla, para posteriormente probar el desempeño del modelo a la hora de predecir correctamente instancias provenientes de muestras de prueba. Este procedimiento se puede apreciar mediante el siguiente esquema mostrado en la figura 7.2:

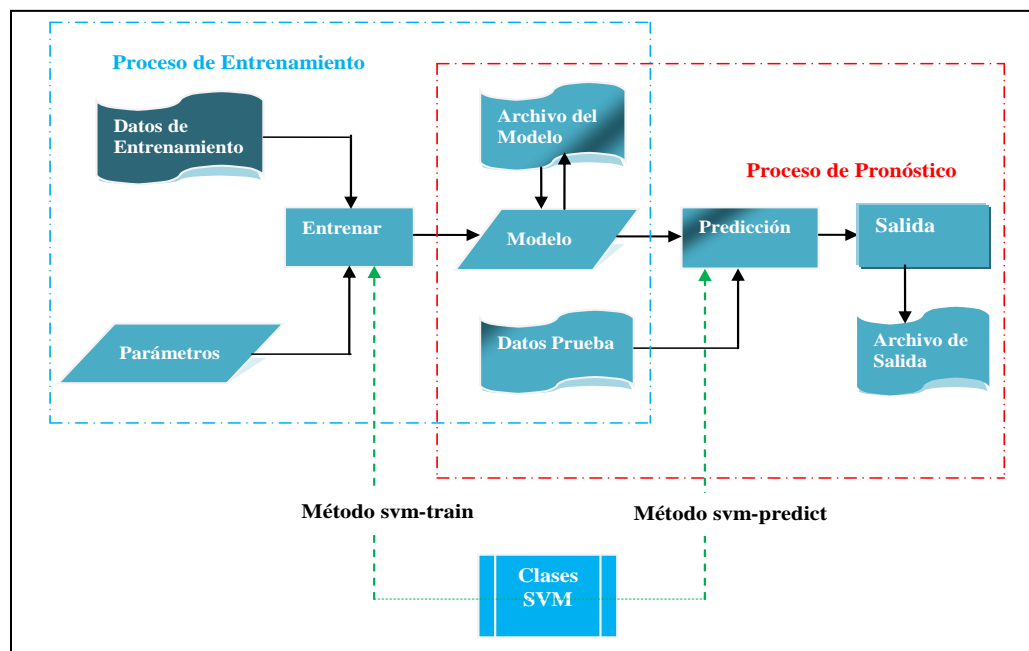


Figura 7.2: Esquema del procedimiento general para la construcción del modelo de clasificación de MVS.

7.2 Resultados obtenidos en las pruebas con MVS: patrones de gastos familiares a nivel nacional

Para la experimentación con modelos de máquinas de vectores soporte, cuya clase de salida fuesen los patrones de gastos a nivel nacional, se consideró un plan de dos fases: primero obtener los modelos con parámetros ideales para cada muestra de entrenamiento y cada tipo de función kernel, evaluando el desempeño de cada modelo mediante las proporciones de instancias clasificadas correctamente en las predicciones en tres muestras de validación y en todas las observaciones de la vista minable; y segundo, evaluar las mejores configuraciones encontradas en la primera fase para cada uno de los tipos de kernel probados (lineal, polinomial y función de base radial) a través del uso de muestras de distintos tamaños, esto con el fin de verificar cual es el efecto que ejerce el tamaño de la muestra de entrenamiento en la precisión obtenida por el modelo y decidir cuál es el tipo de función kernel más apropiado para la construcción de modelos de clasificación de patrones de gastos familiares (en este caso a nivel nacional).

Al aplicar la ecuación 7.1, los distintos tamaños de muestra para distintos niveles de error máximo de estimación deseado en el pronóstico se pueden apreciar en la tabla 7.1:

Patrón de gastos	N° Obs.	p_i	B	d				
				0,01	0,02	0,03	0,04	0,05
A	3.410	0,411	7,477	18.100,25	4.525,06	2.011,14	1.131,27	724,01
B	1.626	0,196	7,477	11.782,56	2.945,64	1.309,17	736,41	471,30
C	1.271	0,153	7,477	9.689,52	2.422,38	1.076,61	605,59	387,58
D	1.045	0,126	7,477	8.233,97	2.058,49	914,89	514,62	329,36
E	527	0,064	7,477	4.479,02	1.119,76	497,67	279,94	179,16
F	302	0,036	7,477	2.594,82	648,70	288,31	162,18	103,79
G	71	0,009	7,477	666,87	166,72	74,1	41,68	26,67
H	40	0,005	7,477	371,98	93	41,33	23,25	14,88
Max(n_i)				18.100,25	4.525,06	2.011,14	1.131,27	724,01
Tamaño muestral recomendado				18.101	4.526	2.012	1.132	725

Tabla 7.1: Tamaños muestrales para distintos niveles de error máximo permitido. Método de Tortora.

La tabla anterior indica que no es factible extraer una muestra a partir de la vista minable de modo que permita obtener un clasificador con un nivel de precisión mayor al 99%, pues para el caso de los ocho patrones de gastos nacionales se requerirían alrededor de 18.101 observaciones, cantidad que es superior al total de observaciones existentes en la vista minable: 8.292 familias; por esta razón solo se crearon muestras con 725, 1.132, 2.012 y 4.526 observaciones. Para agilizar la obtención de resultados y examinar la capacidad de generalización de los modelos de clasificación, las pruebas hechas en la primera fase se basaron todas en muestras de 725 observaciones: 5 muestras con datos de desembolsos sin cambios de escala y 5 muestras con las mismas observaciones de las primeras 5 muestras, pero con los datos de desembolsos escalados en el rango [0,1]. A continuación se muestran los resultados obtenidos en la primera fase de esta experimentación.

7.2.1 Primera fase de experimentación: MVS para los patrones de gastos familiares nacionales

Modelos MVS basados en el kernel lineal

Los resultados obtenidos en cada una de las muestras de entrenamiento y el desempeño en la predicción por cada uno de los modelos se muestran en las tablas 7.2 y 7.3:

Muestra	Resultados en búsqueda de rejilla	
	C	Tasa de aciertos (%)
Muestra1	5,82077E-11	93,79
Muestra2	7,27596E-12	94,76
Muestra3	1,45519E-11	94,48
Muestra4	2,91038E-11	94,07
Muestra5	5,82077E-11	95,04
Muestra1_Normalizado	4096	92,28
Muestra2_Normalizado	4096	92,69
Muestra3_Normalizado	16384	92,83
Muestra4_Normalizado	1024	92,55
Muestra5_Normalizado	16384	92,28

Tabla 7.2: Modelos MVS con kernel lineal. Patrones de gastos nacionales como clase de salida.

Muestra	Resultados en pruebas de validación (% de aciertos)				
	Prueba 1	Prueba 2	Prueba 3	Vista minable	Rendimiento
Muestra1	93,79	92,97	93,52	93,98	93,86
Muestra2	93,66	93,66	92,41	93,89	93,75
Muestra3	92,88	92,69	91,59	93,62	93,34
Muestra4	93,93	94,07	91,45	94,3	94,04
Muestra5	94,76	94,62	94,35	95,12	95
Promedio	93,804	93,602	92,664	94,182	94
Muestra1_Normalizado	92,55	91,45	93,38	93,38	93,17
Muestra2_Normalizado	92,41	90,07	91,04	92,28	92,03
Muestra3_Normalizado	93,1	92,41	92,14	93,6	93,37
Muestra4_Normalizado	92,14	92,69	90,76	92,73	92,54
Muestra5_Normalizado	93,93	93,79	92,41	93,83	93,73
Promedio	92,826	92,082	91,946	93,164	92,97

Tabla 7.3: Desempeño de los modelos MVS propuestos con kernel lineal. Patrones de gastos nacionales como clase de salida.

En la tabla 7.2, la columna *Tasa de aciertos* muestra el porcentaje de instancias correctamente clasificadas por el modelo en el conjunto de datos de entrenamiento a través de validación cruzada de 5 pliegues. Por otro lado, la columna *Rendimiento* de la tabla 7.3 se refiere al promedio ponderado de las tasas de aciertos obtenidas en las distintas pruebas, cuya expresión está dada por:

$$Rendimiento = (725 * P_1 + 725 * P_2 + 725 * P_3 + 7.567 * VM) / (3 * 725 + 7.567)$$

con P_i como la tasa de aciertos obtenida al evaluar las observaciones de la i -ésima muestra de prueba y VM la tasa de aciertos obtenida al evaluar las observaciones de la vista minable. Nótese que la ponderación otorgada a VM es de 7.567, pues se sustrae de las 8.292 observaciones las 725 utilizadas en el entrenamiento, mientras que la ponderación para los P_i es 725, cifra que coincide con el tamaño de las

muestras de prueba. A continuación se muestra el desempeño del kernel RBF en la construcción de modelos de clasificación basados en MVS.

Modelos MVS basados en la función kernel de base radial

Los resultados obtenidos mediante el procedimiento de búsqueda de los parámetros C y γ y el desempeño de los distintos modelos en las pruebas de validación, se muestran a continuación en las tablas 7.4 y 7.5:

Muestra	Resultados en búsqueda de rejilla		
	C	γ	Tasa de aciertos (%)
Muestra1	512	5,68434E-14	94,07
Muestra2	262.144	1,38778E-16	95,17
Muestra3	65.536	1,77636E-15	94,48
Muestra4	32	1,13687E-13	94,62
Muestra5	65.536	4,44089E-16	95,04
Muestra1_Normalizado	134.217.728	0,000488281	92,97
Muestra2_Normalizado	1.048.576	0,001953125	92,97
Muestra3_Normalizado	8.388.608	0,000488281	93,10
Muestra4_Normalizado	4.194.304	0,000244141	92,55
Muestra5_Normalizado	524.288	0,015625	92,83

Tabla 7.4: Modelos MVS con kernel RBF. Patrones de gastos nacionales como clase de salida.

Muestra	Resultados en pruebas de validación (% de aciertos)				
	Prueba 1	Prueba 2	Prueba 3	Vista minable	Rendimiento
Muestra1	93,79	93,1	93,79	94,22	94,07
Muestra2	93,38	93,24	92,14	93,83	93,63
Muestra3	94,21	93,52	92,69	94,74	94,46
Muestra4	93,1	93,24	91,45	93,84	93,56
Muestra5	94,9	94,62	94,35	95,1	95
Promedio	93,876	93,544	92,884	94,346	94,14
Muestra1_Normalizado	92,14	90,48	93,24	92,52	92,39
Muestra2_Normalizado	92,28	90,07	91,04	92,17	91,94
Muestra3_Normalizado	92,83	92	91,72	93,42	93,14
Muestra4_Normalizado	91,31	92,83	90,76	92,78	92,52
Muestra5_Normalizado	94,07	93,66	92,83	93,45	93,47
Promedio	92,526	91,808	91,918	92,868	92,69

Tabla 7.5: Desempeño de los modelos MVS propuestos con kernel RBF. Patrones de gastos nacionales como clase de salida.

Modelos MVS basados en el kernel polinomial

A continuación se muestran en la tabla 7.6 los resultados obtenidos mediante el procedimiento de búsqueda de los parámetros C y γ , para configuraciones propuestas de grados y coeficientes en la expresión polinomial usada como función núcleo; mientras que el desempeño de los distintos modelos en las pruebas de validación se muestran en la tabla 7.7:

Muestra	C	γ	Mejor grado	Mejor coeficiente	Tasa de aciertos (%)
Muestra1	1,45519E-11	1,13687E-13	9	50	94,069
Muestra2	1	9,09495E-13	3	10	95,0345
Muestra3	1.048.576	4,44089E-16	7	1	94,6207
Muestra4	0,000244141	1,11022E-16	9	10	94,6297
Muestra5	0,125	2,27374E-13	2	1000	95,1724
Muestra1_Normalizado	256	0,25	2	100	92,4138
Muestra2_Normalizado	4,76837E-07	0,125	7	50	93,3793
Muestra3_Normalizado	0,125	0,0078125	7	10	92,9655
Muestra4_Normalizado	524.288	0,000244141	9	1	92,5517
Muestra5_Normalizado	2	0,125	5	10	92,6897

Tabla 7.6: Modelos MVS con kernel polinomial. Patrones de gastos nacionales como clase de salida.

Muestra	Resultados en pruebas de validación (% de aciertos)				
	Prueba 1	Prueba 2	Prueba 3	Vista minable	Rendimiento
Muestra1	94,9	92,97	93,66	94,38	94,26
Muestra2	94,48	93,66	93,24	94,91	94,66
Muestra3	93,93	93,93	93,66	95,21	94,9
Muestra4	94,07	93,93	91,45	94,48	94,18
Muestra5	94,76	94,9	94,35	95,24	95,11
Promedio	94,428	93,878	93,272	94,844	94,62
Muestra1_Normalizado	92,69	92,41	93,52	93,84	93,62
Muestra2_Normalizado	92,55	90,76	90,62	92,52	92,25
Muestra3_Normalizado	93,79	92,69	91,86	93,95	93,69
Muestra4_Normalizado	92,28	92,83	91,45	93,03	92,84
Muestra5_Normalizado	94,48	93,66	92,97	93,66	93,67
Promedio	93,158	92,47	92,084	93,4	93,22

Tabla 7.7: Desempeño de los modelos MVS propuestos con kernel polinomial. Patrones de gastos nacionales como clase de salida.

Los resultados obtenidos en los distintos entrenamientos señalan que no es recomendable usar datos escalados, pues el desempeño de los modelos de clasificación basados en estos datos de entrenamiento es inferior al obtenido con los

modelos cuyos ejemplos de entrenamiento fueron datos sin cambio de escala. Al usar la función kernel polinomial, la precisión de los modelos de clasificación es ligeramente superior que los resultados obtenidos con los modelos con función kernel lineal y función kernel de base radial; sin embargo, los valores del parámetro de regularización C son considerablemente mayores a los registrados por las configuraciones propuestas con los modelos MVS con kernel lineal.

Una característica común de los modelos basados en los kernel RBF y polinomial, radica en que los valores del parámetro γ resultan ser prácticamente nulos. Como este parámetro γ controla el ancho de la función kernel, valores que tienden a cero estrechan la amplitud de dicha función, haciéndola que se comporte como una función lineal. Este hecho, aunado con el excelente desempeño de los modelos MVS con kernel lineal, refuerza la afirmación acerca de los comportamientos particulares que guardan cada uno de los patrones de gastos familiares a nivel nacional.

7.2.2 Segunda fase de experimentación: evaluación de los mejores modelos para cada tipo de función kernel

En la fase anterior se obtuvieron las configuraciones más recomendadas de los modelos de clasificación basados en las funciones kernel lineal, polinomial y la función de base radial, para cada conjunto de datos de entrenamiento, resultando como mejores configuraciones los modelos obtenidos al entrenar con los ejemplos de la muestra 5. En la presente fase de experimentación se determina cuál de estas configuraciones se traduce en el modelo de clasificación más apropiado, por lo que se crearon doce nuevas muestras de entrenamiento para evaluar tanto la precisión en el entrenamiento (mediante validación cruzada de 5 pliegues), como en la

predicción de todas las observaciones de la vista minable para distintos tamaños de muestra de entrenamiento. A continuación se muestran los resultados en el desempeño de cada una de las mejores configuraciones propuestas para cada tipo de kernel:

Desempeño general del mejor modelo MVS con kernel lineal

Recordando que el valor del parámetro de regularización C es $5.82077E-11$, el desempeño de esta configuración bajo distintos tamaños de muestra de entrenamiento se puede apreciar a continuación en la tabla 7.8:

Muestra de entrenamiento	Tamaño	% de aciertos en validación cruzada (5 pliegues)	% de aciertos en prueba de validación (Vista minable)
Entrenamiento 1	725	92,5517	95,2002
Entrenamiento 2	725	92,1379	94,6696
Entrenamiento 3	725	92,8276	95,3449
Promedio		92,5057	95,0716
Entrenamiento 4	1.132	93,3746	96,5991
Entrenamiento 5	1.132	95,583	95,6464
Entrenamiento 6	1.132	95,2297	96,2615
Promedio		94,7291	96,169
Entrenamiento 7	2.012	96,1233	97,8051
Entrenamiento 8	2.012	96,8191	97,4916
Entrenamiento 9	2.012	96,7197	97,6363
Promedio		96,5540	97,6443
Entrenamiento 10	4.526	97,8347	98,9146
Entrenamiento 11	4.526	97,7684	98,7217
Entrenamiento 12	4.526	97,5917	98,782
Promedio		97,7316	98,8061
Vista minable	8.292	98,5528	99,4573
Desempeño global		97,0786	98,2478

Tabla 7.8: Desempeño del mejor modelo MVS con kernel lineal para distintos tamaños muestrales.

En la tabla anterior se puede apreciar que el porcentaje de instancias clasificadas incorrectamente no es mayor al error máximo preestablecido para la definición de los distintos tamaños muestrales, hecho que respalda al método de Tortora como un procedimiento recomendable para el muestreo aleatorio estratificado en poblaciones multinomiales. El efecto del uso de muestras de entrenamiento con más observaciones se refleja en el incremento porcentual de las instancias clasificadas correctamente en alrededor de 1% entre los distintos tamaños muestrales, por lo que la construcción de modelos de clasificación depende del nivel de precisión deseado, sin embargo se recomienda usar muestras de entrenamiento grandes debido a la rapidez en la obtención de los modelos mediante esta técnica de

aprendizaje supervisado. Este modelo de clasificación goza de un excelente desempeño, hecho que se refleja tanto en el gran porcentaje de instancias correctamente clasificadas al entrenar al modelo con todos los datos, como en la uniformidad en las clasificaciones para cada uno de los patrones de gastos, resultado que se puede apreciar en la siguiente matriz de clasificación:

Patrones observados	Patrones predichos								% predicho correctamente
	A	B	C	D	E	F	G	H	
A	3.391	14	5	0	0	0	0	0	99,44%
B	10	1.596	10	4	6	0	0	0	98,15%
C	8	5	1.253	5	0	0	0	0	98,58%
D	0	7	8	1.023	4	3	0	0	97,89%
E	0	9	4	4	509	0	0	1	96,58%
F	0	0	0	3	1	295	3	0	97,68%
G	0	0	0	0	0	2	69	0	97,18%
H	0	0	0	0	0	0	0	40	100,00%
% del Total	41,2%	19,6%	15,3%	12,6%	6,4%	3,7%	0,9%	0,5%	97,93%

Tabla 7.9: Matriz de clasificación obtenida por el modelo MVS con kernel lineal.

Desempeño general del mejor modelo de MVS con función kernel de base radial

Para este modelo la configuración de los parámetros es como sigue: $C = 65.536$ y $\gamma = 4,44089E-16$. El desempeño de este modelo para varios tamaños muestrales se aprecia a continuación en la tabla 7.10:

Muestra de entrenamiento	Tamaño	% de aciertos en validación cruzada (5 pliegues)	% de aciertos en prueba de validación (Vista minable)
Entrenamiento 1	725	92,2759	95,1761
Entrenamiento 2	725	92,2759	94,6696
Entrenamiento 3	725	92,8276	95,3449
Promedio		92,4598	95,0635
Entrenamiento 4	1.132	93,3746	96,6353
Entrenamiento 5	1.132	95,4947	95,562
Entrenamiento 6	1.132	95,2297	96,3097
Promedio		94,6997	96,169
Entrenamiento 7	2.012	96,2227	97,8413
Entrenamiento 8	2.012	96,8688	97,5036
Entrenamiento 9	2.012	96,7694	97,6122
Promedio		96,6203	97,6524
Entrenamiento 10	4.526	97,8568	98,9146
Entrenamiento 11	4.526	97,7905	98,6975
Entrenamiento 12	4.526	97,5917	98,7578
Promedio		97,7463	98,79
Vista minable	8.292	98,5528	99,4573
Desempeño global		97,0905	98,2422

Tabla 7.10: Desempeño del mejor modelo MVS con kernel RBF para distintos tamaños muestrales.

La tabla anterior indica que el desempeño de este modelo para las distintas pruebas de validación es similar al obtenido con el modelo basado en el kernel lineal, aunque las medidas de precisión tienden a ser menores cuando las muestras de entrenamiento poseen pocas observaciones. A continuación se muestra la matriz de clasificación obtenida en el entrenamiento con todos los datos de la vista minable:

Patrones observados	Patrones predichos								% predicho correctamente
	A	B	C	D	E	F	G	H	
A	3.391	14	5	0	0	0	0	0	99,44%
B	10	1.596	10	4	6	0	0	0	98,15%
C	9	4	1.253	5	0	0	0	0	98,58%
D	0	9	7	1.019	7	3	0	0	97,51%
E	0	9	4	4	509	0	0	1	96,58%
F	0	0	0	3	1	295	3	0	97,68%
G	0	0	0	0	0	2	69	0	97,18%
H	0	0	0	0	0	0	0	40	100,00%
% del total	41,2%	19,6%	15,3%	12,6%	6,4%	3,7%	0,9%	0,5%	97,88%

Tabla 7.11: Matriz de clasificación obtenida por el modelo MVS con kernel RBF.

Desempeño general del mejor modelo MVS con función kernel polinomial

Este modelo de clasificación se caracteriza por tener como función kernel a la expresión $(\gamma \langle x, y \rangle + 1.000)^2$, con un valor de penalización C de 0,125 y un valor de $2,27374E-13$ para el parámetro γ . En la tabla 7.12 se muestra el desempeño de este modelo relativamente complejo en las distintas muestras de entrenamiento, especificadas en la segunda fase de experimentación de modelos MVS para los patrones de gasto a nivel nacional:

Muestra de entrenamiento	Tamaño	% de aciertos en validación cruzada (5 pliegues)	% de aciertos en prueba de validación (Vista minable)
Entrenamiento 1	725	92,5517	95,164
Entrenamiento 2	725	92,1379	94,6093
Entrenamiento 3	725	92,8276	95,3087
Promedio		92,5057	95,0273
Entrenamiento 4	1.132	93,3746	96,6233
Entrenamiento 5	1.132	95,6714	95,6343
Entrenamiento 6	1.132	95,1413	96,2856
Promedio		94,7291	96,1811
Entrenamiento 7	2.012	96,0736	97,7689
Entrenamiento 8	2.012	96,8688	97,5036
Entrenamiento 9	2.012	96,8688	97,6363
Promedio		96,6037	97,6363
Entrenamiento 10	4.526	98,0115	98,9749
Entrenamiento 11	4.526	97,8347	98,6493
Entrenamiento 12	4.526	97,5917	98,7699
Promedio		97,8126	98,7980
Vista minable	8.292	98,5166	99,4332
Desempeño global		97,1114	98,2354

Tabla 7.12: Desempeño del mejor modelo MVS con kernel polinomial para distintos tamaños muestrales.

A pesar que este modelo es más complejo que los modelos MVS basados en los kernel lineal y de función de base radial, la medida de desempeño global es ligeramente inferior respecto a la obtenida con estos modelos. Este modelo de MVS con kernel polinomial clasifica correctamente de manera uniforme a todos los patrones de gasto familiares, es decir, los porcentajes de instancias correctamente predichas son similares para los distintos patrones de gasto, hecho que se refleja en la tabla 7.13:

Patrones observados	Patrones predichos								% predicho correctamente
	A	B	C	D	E	F	G	H	
A	3391	14	5	0	0	0	0	0	99,44%
B	9	1597	10	4	6	0	0	0	98,22%
C	10	5	1251	5	0	0	0	0	98,43%
D	0	7	7	1024	4	3	0	0	97,99%
E	0	9	4	4	509	0	0	1	96,58%
F	0	0	0	3	1	295	3	0	97,68%
G	0	0	0	0	0	2	69	0	97,18%
H	0	0	0	0	0	0	0	40	100,00%
% del total	41,2%	19,6%	15,3%	12,6%	6,4%	3,7%	0,9%	0,5%	97,93%

Tabla 7.13: Matriz de clasificación obtenida por el mejor modelo MVS con kernel polinomial.

Mejor modelo de clasificación con MVS: Patrones de gastos familiares a nivel nacional

Luego de experimentar con cada uno de los tres tipos de kernel: lineal, función de base radial y polinomial, se considera como el mejor modelo de clasificación a la MVS con kernel lineal y parámetro de regularización C de valor $5.82077E-11$, pues a pesar que los modelos basados con las otras funciones kernel presentaron niveles satisfactorios en la predicción, resultan ser más complejos. Los modelos con kernel polinomial y de función de base radial se caracterizan por tener valores del parámetro γ prácticamente nulos, lo que obliga a este tipo de funciones a comportarse de manera similar a funciones lineales. En el caso del modelo basado en la función kernel de base radial, destaca el valor alto de penalización C de 65.536 ; por otro lado, el modelo basado en el kernel polinomial tiene un desempeño similar al modelo con kernel lineal, sin embargo su valor C de $0,125$ es mucho mayor que el del modelo lineal.

7.3 Resultados obtenidos en la experimentación con MVS: patrones de gastos familiares obtenidos en la ciudad de Mérida

La experimentación con modelos de clasificación, basados en los patrones de gastos extraídos en las familias de la ciudad de Mérida, se limitó en la construcción de modelos de MVS con todos los datos pertenecientes a dicha jurisdicción (302 observaciones), evaluando el desempeño de los modelos mediante la precisión obtenida a través de validación cruzada de 5 pliegues. No se consideró la creación de muestras para la predicción y prueba dado el número reducido de ejemplos.

El uso de la búsqueda de rejilla como herramienta para la elección de la configuración ideal de los parámetros para cada tipo de función kernel, permitió extraer los siguientes modelos de clasificación:

Modelo	Función kernel	Parámetros	% de aciertos en validación cruzada (5 pliegues)
Modelo 1	Lineal	$C = 1,45519152284E-11$	96,6887
Modelo 2	Función de base Radial	$C = 1 ; \gamma = 0,015625$	50,6623
Modelo 3	Polinomial	$C = 1,49012E-8 ;$ $\gamma = 4,4408920985E-16;$ Grado = 7 ; Coeficiente = 100	97,351

Tabla 7.14: Mejores modelos MVS para la clasificación de los patrones de gastos familiares en la ciudad de Mérida.

La tabla anterior indica que los modelos de clasificación basados en la función kernel de base radial tienen un desempeño deficiente cuando la variable clase corresponde a los patrones de gastos familiares de la ciudad de Mérida. Por otro lado, el desempeño de los modelos basados en las funciones kernel lineal y polinomial es equiparable, siendo ligeramente superior el modelo de clasificación bajo el kernel polinomial. De nuevo, el valor despreciable del parámetro de amplitud γ en el modelo con kernel polinomial da un indicio acerca del carácter linealmente separable que guardan entre sí las observaciones clasificadas bajo distintos patrones de gastos. Sin embargo, para este tipo de clase de salida se recomienda como mejor modelo a la MVS con kernel polinomial y con la configuración de parámetros dada en la tabla 7.14, pues clasifica uniformemente cada uno de los patrones de gastos, es decir, las proporciones en instancias predicas correctamente no difieren mucho entre distintos patrones de gastos, hecho que se puede constatar en la siguiente matriz de clasificación:

Patrones observados	Patrones predichos							% predicho correctamente
	A	B	C	D	E	F	G	
A	151	0	0	2	0	0	0	98,69%
B	1	56	0	0	0	0	0	98,25%
C	0	2	26	0	1	0	0	89,66%
D	1	1	0	25	0	1	0	89,29%
E	0	0	0	0	20	0	0	100,00%
F	1	1	0	0	0	10	0	83,33%
G	0	0	0	0	0	0	3	100,00%
% del total	50,66%	18,87%	9,60%	9,27%	6,62%	3,97%	0,99%	94,17%

Tabla 7.15: Matriz de clasificación obtenida por el mejor modelo MVS con los patrones de gasto de la ciudad de Mérida como clase de salida.

Capítulo 8

Conclusiones y recomendaciones

En este capítulo se presentan las conclusiones inherentes al estudio de patrones de gastos familiares extraídos a partir de los datos de la III ENPF y las recomendaciones para nuevas investigaciones basadas en la identificación de patrones de gastos familiares.

8.1 Conclusiones

- A través del método de agrupamiento de K-medias se definieron, a nivel nacional, ocho grupos de familias con comportamientos de gasto diferentes, tanto en las prioridades asignadas a los distintos rubros considerados como en las magnitudes desembolsadas; y siete modalidades de gasto en la muestra de familias que viven en la ciudad de Mérida, municipio Libertador. Estos patrones de gasto describen comportamientos particulares dentro de las familias venezolanas y merideñas, por lo que la estructura de cada uno de ellos difiere considerablemente a la de los demás patrones de gasto.
- El estudio de las características socio-económicas de las familias clasificadas bajo los mismos patrones de gasto permitió comprobar la validez de cada una de estas modalidades de propensión al gasto, pues existe una relación lógica

entre los valores que toman las distintas variables de ingreso y demás características familiares con la manera en la que la familia distribuye su presupuesto: familias de pocos ingresos describen comportamientos similares a la de los patrones de gasto con los menores desembolsos, mientras que las familias con mejores condiciones socio-económicas son clasificadas bajo patrones de gastos que presentan mayores desembolsos.

- Los rubros alimentos y bebidas no alcohólicas, transporte, vestido y calzado, vivienda y sus servicios constituyen los gastos de mayor relevancia para la mayoría de las familias (tanto venezolanas como en las merideñas), hecho que se refleja en las altas proporciones que presentan estos rubros respecto al desembolso total, así como también por los elevados montos desembolsados en estas variables bajo cada uno de los distintos modos de comportamiento del gasto familiar.
- El incremento del presupuesto familiar posibilita el aumento en el desembolso en equipos y servicios de comunicaciones, gastos en educación y cultura, gastos en bienes y servicios para la recreación, gastos en salud y la adquisición de mobiliario y equipos para el hogar. Patrones de gastos familiares correspondientes a mayores desembolsos usualmente presentan mayores proporciones invertidas en estos rubros que los patrones de gastos que describen menores desembolsos totales.
- Los rubros servicios financieros, tributarios y legales, viajes, bebidas alcohólicas y otros gastos constituyen bienes y servicios poco frecuentes, pues presentan tanto las menores proporciones como los más bajos montos de desembolsos en los distintos patrones de gastos familiares (tanto nacionales como locales).

-
- Las familias merideñas gozan de mejores condiciones económicas que la familia venezolana en general, pues por un lado los patrones de gasto locales promedian mayores desembolsos totales al mes en los grupos conformados por mayores observaciones, y por el otro, las equivalencias entre los patrones de gasto nacionales y los patrones de gasto locales indican que éstos últimos están conformados por patrones de gasto nacional que describen mayores desembolsos.
 - Las familias merideñas se caracterizan por la tendencia a presentar mayores desembolsos en bienes y servicios para la educación y la cultura, gastos en equipos y servicios de comunicaciones, gastos en la vivienda y sus servicios y en artículos para la recreación y el esparcimiento, hechos característicos en una ciudad con gran población estudiantil; mientras que la familia venezolana tiene la tendencia al gasto en artículos, efectos personales y servicios diversos, vestido y calzado, mobiliario, equipos para el hogar y bienes necesarios para el mantenimiento de la vivienda.
 - La excelente adecuación de modelos de clasificación basados en máquinas de vectores de soporte a la hora de establecer pronósticos constituye una prueba de las considerables diferencias existentes entre las familias clasificadas bajo distintos patrones de gastos. Por otra parte, el buen desempeño de la función kernel lineal y los valores prácticamente nulos que toma el parámetro γ en los modelos MVS con funciones kernel polinomial y RBF, constituye una prueba del carácter linealmente separable que gozan los ejemplos pertenecientes a distintos patrones de gastos en el espacio de entrada.

8.2 Recomendaciones

- Adaptar scripts y funciones como las escritas en el presente estudio para procesar de manera automatizada el vasto volumen de datos sobre gastos e ingresos monetarios, presentes en las distintas tablas de la base de datos de las Encuestas Nacionales de Presupuestos Familiares.
- En el caso concreto del software utilizado en el presente estudio para la construcción de modelos de clasificación de MVS (LIBSVM), se recomienda adaptar el código del programa de búsqueda de rejilla para habilitar la obtención del mejor grado y el mejor coeficiente en la función kernel polinomial.
- Evaluar el desempeño de modelos de regresión logística en estudios alternativos en el área de gastos e ingresos familiares, debido a que éstos facilitan la comprensión de la importancia de las variables (tanto categóricas como numéricas) en el fenómeno de estudio.
- Realizar estudios análogos al concluido pero utilizando los datos a obtener en la IV ENPF y las subsiguientes ENPF a realizarse en el país, permitiendo así la identificación de posibles cambios en las prioridades de los venezolanos respecto a sus gastos en el transcurso de los años.

Referencias bibliográficas

Allende, H., Benavides, M. y Seccatore, L. (2.006). Máquinas de soporte vectorial en regresión. Universidad Técnica Federico Santa María, Chile.

[1] Análisis factorial. Documento web recuperado el día 15 de Junio de 2.009. URL: <http://ciberconta.unizar.es/LECCION/factorial/FACTORIALEC.pdf>

Anido, D. (1.998). Sistema lineal del gasto: Especificación y estimación para la ciudad de Mérida, 1.986. Tesis de grado, Instituto de Investigaciones Económicas y Sociales, IEES. Universidad de Los Andes, Mérida, Venezuela.

Anido, D., Colmenares, G., Orlandoni, G. y Quintero, L. (2.007). Estructura del gasto y del ingreso familiar en la ciudad de Mérida, Venezuela: Un análisis basado en las encuestas de presupuestos familiares. Revista FeRMentUM, 17(50).

Aranguren, L. (2.008). Identificación de patrones de consumo de los venezolanos mediante maquinas de vectores de soporte. Tesis de grado, Escuela de Ingeniería de Sistemas, Departamento de Investigación de Operaciones. Universidad de Los Andes, Mérida, Venezuela.

Betancourt, G. (2.005) Las Máquinas de Soporte Vectorial (SVMs). Scientia et Technica, Año XI, N° 27, UTP. ISSN 0122-1701, Colombia.

- Chang, C. y Lin, C. (2.009). LIBSVM: a Library for Support Vector Machines. Documento Web recuperado el día 7 de Mayo de 2.009. URL: <http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>
- [2] III Encuesta Nacional de Presupuestos Familiares. Documento Web recuperado el día 5 de Abril de 2.009. URL: <http://www.slideshare.net/motorspade/encuesta-nacional-de-presupuestos-familiares-2005-bcv>
- Grajales, T. (2.000). El análisis factorial. Documento Web recuperado el día 10 de Junio de 2.009. URL: <http://tgrajales.net/estfactorial.pdf>
- Gondar, J. (2.000). Análisis factorial. Data Mining Institute. Madrid, España. Documento Web recuperado el día 10 de Junio de 2.009. URL: <http://www.estadistico.com/arts.html?20001106&PHPSESSID=e8d77cbf5297ff9fe7e43c75b5934a0a>
- Gonzales, R. (2.002). Reflexiones sobre el consumo: más allá de lo privado y más acá de la condena. Revista Propositiones, Nº 34. Santiago de Chile. Documento Web recuperado el día 12 de Abril de 2.009. URL: <http://www.sitiosur.cl/publicacionesdescarga.php?id=3332&nunico=15000034>
- Gunn, S. R. (1.997). *Support Vector Machines for Classification and Regression. Technical Report*. Image Speech and Intelligent Systems Research Group, Universidad de Southampton, Inglaterra.
- Hair, J., Anderson, R., Tatham, R. y Black, W. (1.999). Análisis Multivariante. Editorial Prentice-Hall. 5 Edición. Versión traducida al español. Universidad Autónoma de Madrid, España.

- Hernández, J., Ramírez, M. y Ferri, C (2.004). Introducción a la Minería de Datos. Editorial Pearson Educación, S.A. Madrid, España.
- Hsu, C., Chang, C. y Lin, C. (2.008). A practical guide to Support Vector Classification. Documento Web recuperado el día 18 de Mayo de 2.009. URL: <http://www.csie.ntu.edu.tw/~cjlin>
- Márquez, V. (2.004). Análisis estadístico de presupuestos familiares en Venezuela. Tesis de grado, Instituto de Estadística Aplicada, Universidad de Los Andes. Mérida, Venezuela.
- Medina, F. (2.002). Determinación del tamaño de muestra en encuestas complejas de propósitos múltiples. Comisión Económica para América Latina y el Caribe, CEPAL. Buenos Aires, Argentina. Documento Web recuperado el día 30 de Mayo de 2.009. URL: <http://www.eclac.cl/deype/mecovi/docs/TALLER10/4.pdf>
- Medina, F. (2.002). Tamaño óptimo de muestra en encuestas de propósitos múltiples. Comisión Económica para América Latina y el Caribe, CEPAL. Buenos Aires, Argentina. Documento Web recuperado el día 30 de Mayo de 2.009. URL: <http://www.intranus.net/archivos/Tamanooptimomuestraencuestas.pdf>
- Meza, O. (2.001). Notas metodológicas sobre la II Encuesta Nacional de Presupuestos Familiares 1.997/1.998. Cuadernos BCV, N° 15, Venezuela.
- Moro, Q. y Hurtado, A. (2.006). Introducción al diseño de experimentos para el reconocimiento de patrones, Máquinas de Vectores Soporte. Documento Web recuperado el día 30 de Abril de 2.009. URL:

[http://www.infor.uva.es/isaac/doctorado/Cap07 SVM.pdf](http://www.infor.uva.es/isaac/doctorado/Cap07_SVM.pdf)

[3] One-Sample Kolmogorov-Smirnov Goodness-of-Fit Test. Documento Web recuperado el día 12 de Junio de 2.009. URL:

<http://faculty.chass.ncsu.edu/garson/PA765/kolmo.htm>

Peña, D. (2002). Análisis de datos multivariantes. Editorial Mc Graw Hill. Madrid, España.

Pérez L. César (2004). Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS. Editorial Pearson Educación, S.A. Madrid, España.

[4] Resumen de los principales menús del SPSS y Análisis factorial. Documento web recuperado el día 11 de Junio de 2.009. URL:

http://www4.ujaen.es/~mramos/Cursos/CSPSS/CSPSS_14_Factorial_Com.pdf

Sosa, M. (2006). Estructura del gasto en el presupuesto familiar venezolano año 1988. Tesis de grado, Instituto de Investigaciones Económicas y sociales, Universidad de Los Andes. Mérida, Venezuela.

Anexos

Anexo I: Composición de los grupos de gasto definidos

Grupo 1: Alimentos y bebidas no alcohólicas

Código BCV-INE. III ENPF	Descripción del grupo original
10101 – 10124	Cereales y sus productos derivados
10201 – 10231	Carne de ganado vacuno
10301 – 10313	Carne de aves
10401 – 10416	Carne de ganado porcino (cerdo)
10501 – 10512	Otras carnes
10601 – 10634	Preparados de carne
10701 – 10775	Pescados, mariscos y crustáceos
10801 – 10837	Leche, queso y huevo
10901 – 10914	Grasas y aceites comestibles
11001 – 11062	Frutas
11101 – 11138	Hortalizas
11201 – 11210	Raíces feculentas y derivados (verduras)
11301 – 11334	Semillas oleaginosas y leguminosas
11401 – 11405	Azúcar
11501 – 11511	Café, té, cacao
11601 – 11683	Productos alimenticios varios
11701 – 11705	Alimentos especiales para niños
11801 – 11815	Bebidas no alcohólicas
12401 – 12439	Comidas elaboradas

Grupo 2: Bebidas alcohólicas y tabaco

Código BCV-INE. III ENPF	Descripción del grupo original
11901 – 11917	Bebidas alcohólicas
12201 – 12216	Bebidas alcohólicas tomadas fuera del hogar
12301 – 12306	Tabaco

Grupo 3: Comidas fuera del hogar

Código BCV-INE. III ENPF	Descripción del grupo original
12001 – 12014	Alimentos tomados fuera del hogar
12101 – 12123	Bebidas no alcohólicas tomadas fuera del hogar

Grupo 4: Vestido y calzado

Código BCV-INE. III ENPF	Descripción del grupo original
20101 - 20125	Ropa para caballeros
20201 - 20236	Ropa para damas
20301 - 20328	Ropa para niños
20401 - 20435	Ropa para niñas
20501 - 20532	Ropa para bebe
20601 - 20647	Accesorios
20701 - 20720	Accesorios
20801 -20813	Servicio de confección, reparación y alquiler de prendas de vestir , telas sin confeccionar
20901 - 20909	Calzado para caballeros
21001 - 21009	Calzado para damas
21101 - 21109	Calzado para niños
21201 - 21209	Calzado para niñas
21301 -21304	Calzado para bebés
21401 - 21411	Vestimenta, calzado y accesorios de trabajo
21501 - 21513	Servicio de confección y reparación de calzado
21601 - 21624	Materiales para la confección o reparación de prendas de vestir

Grupo 5: Vivienda y sus servicios

Código BCV-INE. III ENPF	Descripción del grupo original
30101 – 30104	Alquiler de vivienda y sus servicios
30201 – 30240	Servicios
30301 – 30304	Otros combustibles
40701 – 40722	Servicios para el hogar excepto servicios domésticos
40801 – 40809	Servicios domésticos
90320, 90334, 90336, 90337, 90404, 90405	Impuestos, gastos legales y seguros

Grupo 6: Mobiliario y equipos del hogar

Código BCV-INE. III ENPF	Descripción del grupo original
40101 - 40199, 42001 - 42011, 40280 - 40299	Muebles
40201 - 40243	Ropa y enseres del hogar
40301 - 40344	Electrodomésticos
40401 - 40422	Otros equipos del hogar
40501 - 40557	Otros equipos del hogar

Grupo 7: Mantenimiento de la vivienda

Código BCV-INE. III ENPF	Descripción del grupo original
30401 -30412	Mantenimiento de la vivienda
40601 - 40668	Gastos diversos del hogar
41001 -41017	Materiales y repuestos para la fabricación y reparación del mobiliario, ropa y enseres y equipos del hogar

Grupo 8: Salud

Código BCV-INE. III ENPF	Descripción del grupo original
50101 - 50173	Medicina en general
50201 - 50265	Artículos médicos
50301 - 50395	Salud y otros servicios conexos
50401 - 50404	Material y repuestos para la reparación de artículos médicos
90307, 90328, 90410, 90413, 90420	Impuestos, gastos legales y seguros diversos

Grupo 9: Transporte

Código BCV-INE. III ENPF	Descripción del grupo original
60101 - 60113	Vehículos de transporte personal
60201 - 60217	Piezas y accesorios para vehículos
60301 - 60307	Combustible y lubricantes para vehículos
60401 - 60414	Servicios de alquiler de vehículos de transporte
60501 - 60528	Servicio de transporte
60601 - 60617	Servicios y reparación de vehículos
60701 - 60705	Otros gastos por vehículos
60901 - 60907	Materiales y repuestos para la reparación de vehículos de transporte
90308, 90318, 90322, 90323, 90329, 90409, 90411, 90417	Impuestos, gastos legales y seguros diversos

Grupo 10: Equipos y servicios de comunicaciones

Código BCV-INE. III ENPF	Descripción del grupo original
60801 - 60849	Equipos y servicios de comunicaciones

Grupo 11: Educación y cultura

Código BCV-INE. III ENPF	Descripción del grupo original
70101, 70107, 70109, 70111, 70120 - 70125, 70128, 70131 - 70136, 70138 - 70140, 70144, 70149, 70150 - 70161, 70163 - 70172, 70175 - 70177, 70186 - 70191, 70194, 70195, 70199	Artículos, equipos y accesorios relacionados con la educación y el esparcimiento
70303, 70308, 70318 - 70321, 70323 - 70325, 70330	Otros equipos y accesorios para educación y esparcimiento
70412, 70413, 70415, 70417, 70427 - 70442, 70447 - 70453, 70456, 70464, 70471, 70472, 70475, 70478, 70479, 70485, 70488 - 70497	Servicios relacionados con educación, esparcimiento y deporte
70502, 70503, 70506	Materiales para la elaboración y reparación de artículos y equipos de educación, esparcimiento y deporte

Grupo 12: Recreación y esparcimiento

Código BCV-INE. III ENPF	Descripción del grupo original
70102 - 70106, 70108, 70110, 70112 - 70119, 70126, 70127, 70129, 70130, 70137, 70141 - 70143, 70145 - 70148, 70162, 70173, 70174, 70178 - 70185, 70196 - 70198	Artículos, equipos y accesorios relacionados con la educación y el esparcimiento
70201 - 70242	Equipos y artículos deportivos
70301, 70302, 70304 - 70307, 70309 - 70317, 70322, 70326 - 70329, 70331 - 70347	Otros equipos y accesorios para educación y esparcimiento
70401 - 70411, 70414, 70416, 70418 - 70421, 70423 - 70426, 70443 - 70446, 70454, 70455, 70457 - 70463, 70465 - 70470, 70473, 70474, 70476, 70477, 70481, 70486, 70487	Servicios relacionados con educación, esparcimiento y deporte
70501, 70504, 70505, 70507 - 70509	Materiales para la elaboración y reparación de artículos y equipos de educación, esparcimiento y deporte

Grupo 13: Artículos, efectos personales y servicios diversos

Código BCV-INE. III ENPF	Descripción del grupo original
80101 - 80187	Artículos y servicios de cuidado personal
80201 - 80227	Efectos personales
80301, 80302	Materiales para la fabricación o reparación de artículos y efectos personales

Grupo 14: Servicios financieros, tributarios y legales

Código BCV-INE. III ENPF	Descripción del grupo original
90101 - 90109, 90111, 90112	Servicios financieros
90301 - 90306, 90309, 90310, 90312, 90313, 90315 - 90317, 90319, 90321, 90324 - 90327, 90332, 90335	Impuestos, gastos legales, tasa y otros
90401 - 90403, 90406, 90407, 90412, 90414 - 90416, 90418, 90419, 90421, 90422	Seguros y previsión social
90501, 90502	Intereses sobre préstamos

Grupo 15: Viajes

Código BCV-INE. III ENPF	Descripción del grupo original
70422, 70480, 70482, 70483	Servicios relacionados con educación, esparcimiento y deporte
90311, 90314, 90330, 90331, 90333	Impuestos, gastos legales, tasa y otros
90408	Seguros y previsión social

Grupo 16: Otros gastos

Código BCV-INE. III ENPF	Descripción del grupo original
90201 - 90266	Servicios y otros pagos n.e.o.p
90601 - 90603	Otros gastos n.e.o.p

Anexo II: Patrones de gastos familiares a nivel nacional y frecuencias observadas en las distintas fuentes de ingreso monetario por cada grupo familiar.

Estructura del patrón de gasto nacional tipo A

RUBRO	Promedio mensual (Bs.)	% del gasto
Alimentos y bebidas no alcohólicas	180.748	31,67%
Transporte	57.639	10,10%
Vestido y calzado	55.171	9,67%
Vivienda y sus servicios	51.158	8,96%
Equipos y servicios de comunicaciones	39.006	6,83%
Salud	28.757	5,04%
Artículos, efectos personales y servicios diversos	26.025	4,56%
Comidas fuera del hogar	25.117	4,40%
Mobiliario y equipos del hogar	22.769	3,99%
Educación y cultura	21.151	3,71%
Mantenimiento de la vivienda	18.968	3,32%
Bebidas alcohólicas y tabaco	18.323	3,21%
Recreación y esparcimiento	16.556	2,90%
Otros	8.129	1,42%
Servicios financieros, tributarios y legales	757	0,13%
Viajes	511	0,09%
TOTAL	570.786	100%

Ingresos mensuales promedio de las familias con patrón de gasto nacional tipo A

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	2.311	67,77%	741.200
Transferencias recibidas en dinero	1.278	37,48%	334.400
Ingreso por trabajo independiente	748	21,94%	2.065.200
Ingresos ocasionales	492	14,43%	102.400
Ingresos por becas de misiones sociales	215	6,30%	186.300
Alquiler de propiedades	177	5,19%	274.400
Ingresos por trabajo en el sector agropecuario	72	2,11%	2.576.400
Ingresos provenientes del exterior	22	0,65%	4.993.500

Estructura del patrón de gasto nacional tipo B

RUBRO	Promedio mensual (Bs.)	% del gasto
Alimentos y bebidas no alcohólicas	497.475	37,46%
Vestido y calzado	127.931	9,63%
Transporte	113.395	8,54%
Equipos y servicios de comunicaciones	78.093	5,88%
Vivienda y sus servicios	74.472	5,61%
Artículos, efectos personales y servicios diversos	71.951	5,42%
Comidas fuera del hogar	58.191	4,38%
Mantenimiento de la vivienda	56.666	4,27%
Mobiliario y equipos del hogar	54.280	4,09%
Salud	53.223	4,01%
Educación y cultura	45.498	3,43%
Bebidas alcohólicas y tabaco	43.962	3,31%
Recreación y esparcimiento	30.108	2,27%
Otros	19.142	1,44%
Servicios financieros, tributarios y legales	2.654	0,20%
Viajes	1.013	0,08%
TOTAL	1.328.054	100%

Ingresos mensuales promedio registrados por las familias con patrón de gasto nacional tipo B

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	1.293	79,52%	1.268.240
Ingreso por trabajo independiente	570	35,06%	2.821.500
Transferencias recibidas en dinero	518	31,86%	542.000
Ingresos ocasionales	304	18,70%	144.900
Alquiler de propiedades	92	5,66%	355.100
Ingresos por becas de misiones sociales	90	5,54%	194.600
Ingresos por trabajo en el sector agropecuario	21	1,29%	5.181.600
Ingresos provenientes del exterior	12	0,74%	4.181.850

Estructura del patrón de gasto nacional tipo C

RUBRO	Promedio mensual (Bs)	% de gasto
Vivienda y sus servicios	241.360	15,52%
Alimentos y bebidas no alcohólicas	221.203	14,23%
Salud	218.056	14,03%
Transporte	212.131	13,64%
Equipos y servicios de comunicaciones	148.311	9,54%
Vestido y calzado	103.698	6,67%
Educación y cultura	101.053	6,50%
Comidas fuera del hogar	74.833	4,81%
Artículos, efectos personales y servicios diversos	61.316	3,94%
Mobiliario y equipos del hogar	46.706	3,00%
Recreación y esparcimiento	42.161	2,71%
Mantenimiento de la vivienda	35.843	2,31%
Otros	20.916	1,35%
Bebidas alcohólicas y tabaco	18.747	1,21%
Servicios financieros, tributarios y legales	5.778	0,37%
Viajes	2.615	0,17%
TOTAL	1.554.728	100%

Ingresos mensuales promedio registrados por las familias con patrón de gasto nacional tipo C

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	1.002	78,84%	1.712.600
Transferencias recibidas en dinero	528	41,54%	735.500
Ingresos ocasionales	327	25,73%	314.000
Ingreso por trabajo independiente	298	23,45%	2.350.000
Alquiler de propiedades	100	7,87%	506.000
Ingresos por becas de misiones sociales	26	2,05%	185.000
Ingresos provenientes del exterior	20	1,57%	6.267.700
Ingresos por trabajo en el sector agropecuario	5	0,39%	3.236.000

Estructura del patrón de gasto nacional tipo D

RUBRO	Promedio mensual (Bs.)	% del gasto
Alimentos y bebidas no alcohólicas	528.481	15,59%
Salud	454.629	13,41%
Transporte	448.015	13,21%
Vivienda y sus servicios	432.980	12,77%
Educación y cultura	330.779	9,75%
Equipos y servicios de comunicaciones	255.962	7,55%
Vestido y calzado	181.664	5,36%
Comidas fuera del hogar	172.850	5,10%
Artículos, efectos personales y servicios diversos	124.279	3,67%
Mobiliario y equipos del hogar	117.686	3,47%
Recreación y esparcimiento	113.427	3,35%
Mantenimiento de la vivienda	104.000	3,07%
Otros	47.793	1,41%
Bebidas alcohólicas y tabaco	46.951	1,38%
Servicios financieros, tributarios y legales	19.951	0,59%
Viajes	11.459	0,34%
TOTAL	3.390.906	100%

Ingresos mensuales promedio registrados por las familias con patrón de gasto nacional tipo D

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	886	84,78%	3.320.400
Transferencias recibidas en dinero	400	38,28%	1.098.300
Ingresos ocasionales	341	32,63%	573.300
Ingreso por trabajo independiente	307	29,38%	3.333.700
Alquiler de propiedades	118	11,29%	817.000
Ingresos provenientes del exterior	13	1,24%	5.411.500
Ingresos por becas de misiones sociales	12	1,15%	198.000
Ingresos por trabajo en el sector agropecuario	7	0,67%	12.394.500

Estructura del patrón de gasto nacional tipo E

RUBRO	Promedio mensual (Bs.)	% del gasto
Alimentos y bebidas no alcohólicas	936.390	26,24%
Vestido y calzado	815.383	22,85%
Transporte	244.275	6,84%
Vivienda y sus servicios	193.394	5,42%
Mobiliario y equipos del hogar	181.112	5,07%
Salud	177.199	4,97%
Equipos y servicios de comunicaciones	170.741	4,78%
Educación y cultura	160.947	4,51%
Artículos, efectos personales y servicios diversos	154.594	4,33%
Recreación y esparcimiento	145.679	4,08%
Comidas fuera del hogar	122.038	3,42%
Mantenimiento de la vivienda	116.236	3,26%
Bebidas alcohólicas y tabaco	96.679	2,71%
Otros	43.855	1,23%
Servicios financieros, tributarios y legales	6.344	0,18%
Viajes	3.861	0,11%
TOTAL	3.568.728	100%

Ingresos mensuales promedio registrados por las familias con patrón de gasto nacional tipo E

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	406	77,04%	2.386.400
Ingreso por trabajo independiente	281	53,32%	4.532.000
Transferencias recibidas en dinero	157	29,79%	880.400
Ingresos ocasionales	108	20,49%	370.000
Alquiler de propiedades	36	6,83%	927.800
Ingresos por becas de misiones sociales	17	3,23%	200.500
Ingresos por trabajo en el sector agropecuario	3	0,57%	4.627.000
Ingresos provenientes del exterior	2	0,38%	784.000

Estructura del patrón de gasto nacional tipo F

RUBRO	Promedio mensual (Bs.)	% del gasto
Transporte	2.631.830	51,28%
Alimentos y bebidas no alcohólicas	426.818	8,32%
Vivienda y sus servicios	371.189	7,23%
Salud	296.329	5,77%
Vestido y calzado	258.985	5,05%
Educación y cultura	221.659	4,32%
Equipos y servicios de comunicaciones	215.600	4,20%
Comidas fuera del hogar	136.325	2,66%
Mobiliario y equipos del hogar	126.970	2,47%
Artículos, efectos personales y servicios diversos	112.563	2,19%
Recreación y esparcimiento	97.582	1,90%
Mantenimiento de la vivienda	95.535	1,86%
Otros	69.900	1,36%
Bebidas alcohólicas y tabaco	40.350	0,79%
Servicios financieros, tributarios y legales	15.789	0,31%
Viajes	14.742	0,29%
TOTAL	5.132.165	100%

Ingresos mensuales promedio registrados por las familias con patrón de gasto nacional tipo F

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	249	82,45%	3.087.000
Ingresos ocasionales	134	44,37%	1.498.000
Transferencias recibidas en dinero	105	34,77%	1.234.200
Ingreso por trabajo independiente	85	28,15%	2.914.500
Alquiler de propiedades	28	9,27%	696.000
Ingresos por trabajo en el sector agropecuario	11	3,64%	7.114.000
Ingresos por becas de misiones sociales	1	0,33%	160.000
Ingresos provenientes del exterior	0	0,00%	0

Estructura del patrón de gasto nacional tipo G

RUBRO	Promedio mensual (Bs.)	% del gasto
Transporte	6.354.330	64,11%
Alimentos y bebidas no alcohólicas	587.720	5,93%
Vivienda y sus servicios	569.441	5,75%
Salud	437.986	4,42%
Educación y cultura	378.706	3,82%
Equipos y servicios de comunicaciones	319.877	3,23%
Vestido y calzado	270.354	2,73%
Comidas fuera del hogar	223.219	2,25%
Artículos, efectos personales y servicios diversos	158.037	1,59%
Recreación y esparcimiento	157.259	1,59%
Mobiliario y equipos del hogar	153.867	1,55%
Mantenimiento de la vivienda	134.443	1,36%
Otros	64.067	0,65%
Bebidas alcohólicas y tabaco	49.249	0,50%
Viajes	26.559	0,27%
Servicios financieros, tributarios y legales	26.548	0,27%
TOTAL	9.911.661	100%

Ingresos mensuales promedio registrados por las familias con patrón de gasto nacional tipo G

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	61	85,92%	4.921.300
Ingresos ocasionales	35	49,30%	2.212.000
Transferencias recibidas en dinero	21	29,58%	1.043.500
Ingreso por trabajo independiente	18	25,35%	5.752.000
Alquiler de propiedades	7	9,86%	2.070.000
Ingresos por trabajo en el sector agropecuario	1	1,41%	15.073.000
Ingresos provenientes del exterior	0	0,00%	0
Ingresos por becas de misiones sociales	0	0,00%	0

Estructura del patrón de gasto nacional tipo H

RUBRO	Promedio mensual (Bs.)	% del gasto
Alimentos y bebidas no alcohólicas	5.258.510	70,79%
Vestido y calzado	312.012	4,20%
Transporte	289.195	3,89%
Artículos, efectos personales y servicios diversos	269.473	3,63%
Mantenimiento de la vivienda	244.255	3,29%
Vivienda y sus servicios	188.357	2,54%
Salud	174.110	2,34%
Educación y cultura	131.781	1,77%
Comidas fuera del hogar	123.417	1,66%
Equipos y servicios de comunicaciones	105.680	1,42%
Otros	104.852	1,41%
Mobiliario y equipos del hogar	103.216	1,39%
Bebidas alcohólicas y tabaco	70.065	0,94%
Recreación y esparcimiento	48.939	0,66%
Servicios financieros, tributarios y legales	4.274	0,06%
Viajes	0	0,00%
TOTAL	7.428.135	100%

Ingresos mensuales promedio registrados por las familias con patrón de gasto nacional tipo H

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Ingreso por trabajo independiente	39	97,50%	10.811.000
Sueldos y salarios	15	37,50%	1.050.000
Ingresos ocasionales	11	27,50%	270.600
Transferencias recibidas en dinero	7	17,50%	366.000
Alquiler de propiedades	3	7,50%	183.400
Ingresos provenientes del exterior	1	2,50%	100.000
Ingresos por trabajo en el sector agropecuario	0	0,00%	0
Ingresos por becas de misiones sociales	0	0,00%	0

Anexo III: Patrones de gastos familiares extraídos en la ciudad de Mérida, municipio Libertador; y frecuencias observadas en las distintas fuentes de ingreso monetario por cada grupo familiar.

Estructura del patrón de gasto local tipo A

RUBRO	Promedio mensual (Bs.)	% del gasto
Alimentos y bebidas no alcohólicas	187.202	21,70%
Vivienda y sus servicios	138.381	16,04%
Equipos y servicios de comunicaciones	102.087	11,83%
Vestido y calzado	86.371	10,01%
Transporte	77.740	9,01%
Salud	69.587	8,07%
Comidas fuera del hogar	46.409	5,38%
Educación y cultura	34.734	4,03%
Mobiliario y equipos del hogar	31.995	3,71%
Artículos, efectos personales y servicios diversos	29.046	3,37%
Recreación y esparcimiento	26.059	3,02%
Mantenimiento de la vivienda	16.517	1,91%
Otros	7.547	0,87%
Bebidas alcohólicas y tabaco	5.392	0,62%
Servicios financieros, tributarios y legales	2.527	0,29%
Viajes	1.089	0,13%
TOTAL	862.681	100,00%

Ingresos mensuales promedio de las familias con patrón de gasto local tipo A

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	100	65,36%	1.130.500
Transferencias recibidas en dinero	81	52,94%	574.000
Ingresos ocasionales	35	22,88%	123.800
Ingreso por trabajo independiente	23	15,03%	2.235.500
Alquiler de propiedades	9	5,88%	526.700
Ingresos por becas de misiones sociales	2	1,31%	160.000
Ingresos provenientes del exterior	1	0,65%	7.200.000
Ingresos por trabajo en el sector agropecuario	0	0,00%	0

Estructura del patrón de gasto local tipo B

RUBRO	Promedio mensual (Bs.)	% del gasto
Alimentos y bebidas no alcohólicas	530.023	24,73%
Vivienda y sus servicios	246.816	11,51%
Educación y cultura	199.724	9,32%
Equipos y servicios de comunicaciones	191.775	8,95%
Vestido y calzado	188.130	8,78%
Transporte	165.311	7,71%
Salud	164.318	7,67%
Comidas fuera del hogar	106.062	4,95%
Artículos, efectos personales y servicios diversos	89.416	4,17%
Recreación y esparcimiento	86.647	4,04%
Mobiliario y equipos del hogar	61.586	2,87%
Mantenimiento de la vivienda	57.965	2,70%
Otros	26.431	1,23%
Bebidas alcohólicas y tabaco	19.204	0,90%
Servicios financieros, tributarios y legales	7.348	0,34%
Viajes	2.854	0,13%
TOTAL	2.143.610	100,00%

Ingresos mensuales promedio registrados por las familias con patrón de gasto local tipo B

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	45	78,95%	2.073.300
Transferencias recibidas en dinero	32	56,14%	1.086.100
Ingresos ocasionales	25	43,86%	286.200
Ingreso por trabajo independiente	15	26,32%	1.738.700
Alquiler de propiedades	3	5,26%	191.000
Ingresos provenientes del exterior	0	0,00%	0
Ingresos por trabajo en el sector agropecuario	0	0,00%	0
Ingresos por becas de misiones sociales	0	0,00%	0

Estructura del patrón de gasto local tipo C

RUBRO	Promedio mensual (Bs.)	% del gasto
Alimentos y bebidas no alcohólicas	980.229	24,17%
Vestido y calzado	968.897	23,89%
Transporte	301.792	7,44%
Equipos y servicios de comunicaciones	281.745	6,95%
Vivienda y sus servicios	266.171	6,56%
Salud	212.740	5,25%
Comidas fuera del hogar	189.204	4,67%
Mobiliario y equipos del hogar	185.128	4,57%
Educación y cultura	178.605	4,40%
Otros	165.464	4,08%
Artículos, efectos personales y servicios diversos	105.267	2,60%
Recreación y esparcimiento	103.350	2,55%
Mantenimiento de la vivienda	64.516	1,59%
Bebidas alcohólicas y tabaco	45.770	1,13%
Servicios financieros, tributarios y legales	6.134	0,15%
Viajes	0	0,00%
TOTAL	4.055.013	100,00%

Ingresos mensuales promedio registrados por las familias con patrón de gasto local tipo C

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	18	62,07%	3.269.400
Ingreso por trabajo independiente	14	48,28%	6.772.900
Transferencias recibidas en dinero	7	24,14%	2.018.900
Ingresos ocasionales	6	20,69%	551.800
Alquiler de propiedades	5	17,24%	266.400
Ingresos por trabajo en el sector agropecuario	2	6,90%	8.330.000
Ingresos por becas de misiones sociales	1	3,45%	145.000
Ingresos provenientes del exterior	0	0,00%	0

Estructura del patrón de gasto local tipo D

RUBRO	Promedio mensual (Bs.)	% del gasto
Salud	422.561	18,44%
Vivienda y sus servicios	400.048	17,45%
Mobiliario y equipos del hogar	241.151	10,52%
Equipos y servicios de comunicaciones	221.319	9,66%
Alimentos y bebidas no alcohólicas	205.214	8,95%
Educación y cultura	141.399	6,17%
Comidas fuera del hogar	141.096	6,16%
Transporte	130.719	5,70%
Artículos, efectos personales y servicios diversos	114.366	4,99%
Vestido y calzado	104.381	4,55%
Mantenimiento de la vivienda	58.442	2,55%
Otros	53.597	2,34%
Recreación y esparcimiento	43.183	1,88%
Bebidas alcohólicas y tabaco	9.100	0,40%
Servicios financieros, tributarios y legales	5.326	0,23%
Viajes	0	0,00%
TOTAL	2.291.902	100,00%

Ingresos mensuales promedio registrados por las familias con patrón de gasto local tipo D

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	23	82,14%	1.808.900
Transferencias recibidas en dinero	12	42,86%	1.165.400
Ingreso por trabajo independiente	9	32,14%	3.589.000
Ingresos ocasionales	6	21,43%	471.100
Alquiler de propiedades	1	3,57%	100.000
Ingresos por becas de misiones sociales	1	3,57%	140.000
Ingresos provenientes del exterior	0	0,00%	0
Ingresos por trabajo en el sector agropecuario	0	0,00%	0

Estructura del patrón de gasto local tipo E

RUBRO	Promedio mensual (Bs.)	% del gasto
Transporte	2.144.810	54,88%
Alimentos y bebidas no alcohólicas	302.828	7,75%
Vestido y calzado	273.875	7,01%
Vivienda y sus servicios	269.916	6,91%
Equipos y servicios de comunicaciones	220.933	5,65%
Salud	213.265	5,46%
Comidas fuera del hogar	155.780	3,99%
Educación y cultura	84.155	2,15%
Artículos, efectos personales y servicios diversos	83.174	2,13%
Mobiliario y equipos del hogar	46.483	1,19%
Bebidas alcohólicas y tabaco	36.135	0,92%
Otros	34.177	0,87%
Mantenimiento de la vivienda	20.628	0,53%
Recreación y esparcimiento	18.325	0,47%
Servicios financieros, tributarios y legales	3.883	0,10%
Viajes	0	0,00%
TOTAL	3.908.367	100,00%

Ingresos mensuales promedio registrados por las familias con patrón de gasto local tipo E

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	16	80,00%	3.383.200
Ingresos ocasionales	10	50,00%	1.736.300
Transferencias recibidas en dinero	9	45,00%	1.230.300
Ingreso por trabajo independiente	3	15,00%	1.933.400
Alquiler de propiedades	1	5,00%	500.000
Ingresos provenientes del exterior	0	0,00%	0
Ingresos por trabajo en el sector agropecuario	0	0,00%	0
Ingresos por becas de misiones sociales	0	0,00%	0

Estructura del patrón de gasto local tipo F

RUBRO	Promedio mensual (Bs.)	% del gasto
Transporte	721.934	25,81%
Salud	311.668	11,14%
Alimentos y bebidas no alcohólicas	310.497	11,10%
Vivienda y sus servicios	255.437	9,13%
Artículos, efectos personales y servicios diversos	218.062	7,80%
Equipos y servicios de comunicaciones	207.842	7,43%
Recreación y esparcimiento	191.622	6,85%
Comidas fuera del hogar	153.117	5,47%
Viajes	108.889	3,89%
Vestido y calzado	102.416	3,66%
Educación y cultura	91.552	3,27%
Mantenimiento de la vivienda	78.549	2,81%
Mobiliario y equipos del hogar	19.403	0,69%
Servicios financieros, tributarios y legales	15.286	0,55%
Otros	9.070	0,32%
Bebidas alcohólicas y tabaco	1.600	0,06%
TOTAL	2.796.944	100,00%

Ingresos mensuales promedio registrados por las familias con patrón de gasto local tipo F

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	9	75,00%	3.654.800
Transferencias recibidas en dinero	6	50,00%	1.264.800
Ingresos ocasionales	5	41,67%	343.500
Ingreso por trabajo independiente	3	25,00%	5.133.400
Alquiler de propiedades	1	8,33%	200.000
Ingresos provenientes del exterior	0	0,00%	0
Ingresos por trabajo en el sector agropecuario	0	0,00%	0
Ingresos por becas de misiones sociales	0	0,00%	0

Estructura del patrón de gasto local tipo G

RUBRO	Promedio mensual (Bs.)	% del gasto
Transporte	7.295.420	80,63%
Alimentos y bebidas no alcohólicas	437.699	4,84%
Salud	279.708	3,09%
Vivienda y sus servicios	242.771	2,68%
Equipos y servicios de comunicaciones	242.339	2,68%
Educación y cultura	144.854	1,60%
Vestido y calzado	127.000	1,40%
Artículos, efectos personales y servicios diversos	83.676	0,92%
Comidas fuera del hogar	65.333	0,72%
Recreación y esparcimiento	63.544	0,70%
Mantenimiento de la vivienda	44.329	0,49%
Servicios financieros, tributarios y legales	12.147	0,13%
Bebidas alcohólicas y tabaco	5.867	0,06%
Otros	3.333	0,04%
Mobiliario y equipos del hogar	0	0,00%
Viajes	0	0,00%
TOTAL	9.048.020	100,00%

Ingresos mensuales promedio registrados por las familias con patrón de gasto local tipo G

Concepto	Familias recibiendo ingreso	% de familias	Promedio relativo (Bs.)
Sueldos y salarios	3	100,00%	3.543.800
Transferencias recibidas en dinero	1	33,33%	100.000
Ingresos ocasionales	1	33,33%	5.834.000
Alquiler de propiedades	0	0,00%	0
Ingresos provenientes del exterior	0	0,00%	0
Ingreso por trabajo independiente	0	0,00%	0
Ingresos por trabajo en el sector agropecuario	0	0,00%	0
Ingresos por becas de misiones sociales	0	0,00%	0