



PROYECTO DE GRADO

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

CLASIFICACIÓN, CUANTIFICACIÓN Y REPRESENTACIÓN DEL
DÉFICIT DE VIVIENDAS EN EL ESPACIO QUE DEFINE EL ÁREA
METROPOLITANA DE MÉRIDA

Por

Br. Jean Carlos Escalona Delgado

Tutor: Gerardo A. Colmenares L.

Cotutor: Alexander Parra U.

Noviembre 2009

©2009 Universidad de Los Andes Mérida, Venezuela

Clasificación, cuantificación y representación del déficit de viviendas en el espacio que define el Área Metropolitana de Mérida

Br. Jean Carlos Escalona Delgado

Proyecto de Grado – Investigación de Operaciones, 144 páginas

Resumen: La finalidad de este trabajo es crear una aplicación que permita construir un pronóstico de clasificación del déficit habitacional en el Área Metropolitana de Mérida, detectarlo y mostrarlo espacialmente. La ausencia de medios que provean resultados sobre este tópico fue el motivo fundamental para desarrollar esta investigación. En consecuencia, este estudio permite medir el fenómeno del déficit de viviendas en el Área Metropolitana con datos provenientes de muestras provistas por la Encuesta de Hogares por Muestreo en el período 2001-2007. La exploración a dichos datos, mediante análisis multivariante y los resultados de aplicar Máquinas de Vectores de Soporte para clasificar y cuantificar el déficit habitacional, suministraron datos intermedios a un sistema geográfico, que a través de un análisis del déficit, facilitó su interpretación en el espacio. En definitiva, se creó una herramienta híbrida que contempló tanto Máquinas de Soporte Vectorial como conceptos de Sistemas de Información Geográfica. Los resultados permitieron inferir significativamente en toda el área tanto descriptiva como espacialmente. Sin embargo, estos resultados quedaron sujetos a la consistencia de los datos de la fuente original.

Palabras claves: Déficit de viviendas, Encuestas de Hogares por Muestreo, Máquinas de vectores de soporte, Sistema de Información Geográfica, Análisis espacial.

Índice

| | |
|---|------|
| Índice..... | iii |
| Índice de Tablas..... | vii |
| Índice de Figuras..... | viii |
| Agradecimientos | xi |
| Capítulo 1 Introducción..... | 1 |
| 1.1 Antecedentes..... | 4 |
| 1.2 Planteamiento del problema..... | 5 |
| 1.3 Objetivos..... | 7 |
| 1.3.1 Objetivo general | 7 |
| 1.3.2 Objetivos específicos..... | 7 |
| 1.4 Justificación | 8 |
| 1.5 Alcances y Limitaciones | 9 |
| Capítulo 2 Marco Referencial..... | 11 |
| 2.1 Conceptos de déficit de viviendas | 11 |
| 2.1.1 Déficit cuantitativo..... | 13 |
| 2.1.2 Déficit cualitativo | 13 |
| 2.2 Estructura de la Encuesta de Hogares por Muestreo..... | 14 |
| 2.2.1 Encuesta de Hogares por Muestreo | 14 |
| 2.2.2 Unidades estadísticas de la EHM | 15 |
| 2.2.3 Marco muestral y tipo de muestreo | 16 |
| 2.2.4 Variables investigadas..... | 17 |
| 2.3 Análisis Exploratorio de Datos (AED) | 18 |
| 2.3.1 Etapas del Análisis Exploratorio de Datos..... | 19 |

| | | |
|------------|---|----|
| 2.4 | Técnicas de Análisis Multivariante de Datos | 19 |
| 2.4.1 | Análisis de conglomerados..... | 19 |
| 2.4.1.1 | Medidas de proximidad y de distancia..... | 20 |
| 2.4.2 | Análisis de correspondencias..... | 20 |
| 2.5 | Máquina de Vectores de Soporte | 21 |
| 2.5.1 | MVS para el caso no linealmente separable..... | 22 |
| 2.5.1.1 | MVS con margen máximo en el espacio de características..... | 24 |
| 2.5.1.2 | MVS con margen blando | 26 |
| 2.5.2 | Ventajas de las MVS | 28 |
| 2.5.3 | Posibles problemas con las MVS | 28 |
| 2.6 | Fundamentos de los SIG..... | 29 |
| 2.6.1 | Concepto de SIG..... | 29 |
| 2.6.2 | Funciones de un SIG..... | 29 |
| 2.6.2.1 | Entrada de datos a un SIG..... | 29 |
| 2.6.2.2 | Análisis espacial..... | 30 |
| 2.6.2.3 | Salidas o producción de resultados | 30 |
| 2.6.3 | Componentes de un SIG | 31 |
| 2.6.4 | Funcionalidad de un SIG | 33 |
| 2.6.4.1 | Las Referencias Geográficas..... | 33 |
| 2.6.4.2 | Modelos de datos: Vectorial y Raster | 34 |
| 2.6.5 | Proceso de Georeferenciación | 35 |
| Capítulo 3 | Procesamiento y exploración de datos..... | 36 |
| 3.1 | Origen y disposición de los datos | 36 |
| 3.2 | Conformación del subconjunto de datos de la EHM..... | 37 |
| 3.2.1 | Registro final de datos..... | 37 |
| 3.2.2 | Registro de observaciones..... | 38 |
| 3.2.3 | Fusión de los registros de datos..... | 38 |

| | | |
|------------|--|----|
| 3.3 | Pre procesamiento de datos | 39 |
| 3.3.1 | Variables Complementarias | 39 |
| 3.3.2 | Indicadores de déficit | 40 |
| 3.3.2.1 | Indicador de déficit cuantitativo | 41 |
| 3.3.2.2 | Indicadores de déficit cualitativo | 41 |
| 3.4 | Definición de la variable de salida..... | 49 |
| Capítulo 4 | Experimentación y análisis de los resultados..... | 52 |
| 4.1 | Proceso de clasificación del déficit habitacional..... | 52 |
| 4.1.1 | Datos del modelo de clasificación | 53 |
| 4.1.1.1 | Atributos o categorías en las variables de entrada | 53 |
| 4.1.1.2 | Clases de la variable de salida | 54 |
| 4.1.2 | Trasformación de los datos al formato de la aplicación utilizada | 54 |
| 4.1.3 | Selección de los datos de entrenamiento y prueba | 56 |
| 4.1.4 | Selección de la función Kernel | 57 |
| 4.1.5 | Ajuste de los parámetros del modelo..... | 58 |
| 4.1.6 | Búsqueda de rejilla y validación cruzada..... | 59 |
| 4.1.7 | Entrenamiento de la MVS..... | 60 |
| 4.1.8 | Pruebas y Resultados..... | 65 |
| 4.2 | Inferencias sobre el déficit habitacional | 66 |
| Capítulo 5 | Representación y análisis espacial | 70 |
| 5.1 | Construcción de la aplicación SIG | 70 |
| 5.1.1 | Obtención de los mapas cartográficos | 70 |
| 5.1.2 | Proceso de georeferenciación | 71 |
| 5.1.3 | Vinculación de los datos a los mapas..... | 71 |
| 5.2 | Representación espacial de la información | 73 |
| 5.3 | Consulta y despliegue de la información espacial | 74 |
| 5.4 | Análisis espacial del déficit habitacional | 78 |

| | | |
|------------|---|-----|
| 5.5 | Visualización de los resultados vía web | 83 |
| 5.5.1 | Arquitectura de una aplicación de MapServer..... | 85 |
| 5.5.2 | Elementos característicos del visor cartográfico..... | 86 |
| 5.5.2.1 | El archivo mapa | 86 |
| 5.5.2.2 | El archivo plantilla..... | 87 |
| 5.5.2.3 | El conjunto de datos SIG..... | 87 |
| 5.6 | Resultados de configurar el servidor de mapas..... | 87 |
| 5.6.1 | Interfaz de usuario..... | 87 |
| 5.6.2 | Funcionalidades del visor web de mapas | 89 |
| Capítulo 6 | Conclusiones y Recomendaciones | 91 |
| 6.1 | Conclusiones | 91 |
| 6.2 | Recomendaciones | 93 |
| | Referencias Bibliográficas | 94 |
| | Anexo A..... | 98 |
| | Anexo B | 101 |
| | Anexo C..... | 107 |
| | Anexo D..... | 114 |
| | Anexo E | 124 |

Índice de Tablas

| | |
|---|----|
| Tabla 1. Categorías de la variable tenencia y sus frecuencias..... | 42 |
| Tabla 2. Variables para construir la indicador condición de la vivienda..... | 44 |
| Tabla 3. Número de observaciones en cada conglomerado..... | 45 |
| Tabla 4. Asociación de las categorías en sus respectivos conglomerados. | 45 |
| Tabla 5. Categorías para el indicador condición de la vivienda..... | 46 |
| Tabla 6. Variables para construir el indicador adecuación de la vivienda. | 47 |
| Tabla 8. Asociación de las categorías en sus respectivos conglomerados. | 48 |
| Tabla 9. Categorías de la Variable de salida..... | 51 |
| Tabla 10. Atributos del modelo de clasificación. | 54 |
| Tabla 11. Formato de los datos de entradas. | 55 |
| Tabla 12. Cantidad de observaciones para entrenamiento y prueba..... | 56 |
| Tabla 13. Tipos de <i>kernels</i> más comunes. | 57 |
| Tabla 14. Resultados de la clasificación sin ajuste de parámetros..... | 60 |
| Tabla 15. Ajuste de los parámetros C y γ con Rejilla 1..... | 61 |
| Tabla 16. Ajuste de los parámetros C y γ con Rejilla 2..... | 62 |
| Tabla 17. Mejor combinación de C y γ | 64 |
| Tabla 18. Medida de desempeño del modelo de clasificación..... | 65 |
| Tabla 19. Capas con información geográfica..... | 73 |

Índice de Figuras

| | |
|--|----|
| Figura 2.1. MVS Linealmente separable. | 23 |
| Figura 2.2. MVS No linealmente separable inducida por una función <i>kernel</i> | 24 |
| Figura 2.3. MVS con margen blando. | 26 |
| Figura 2.4. Componentes de un SIG. | 31 |
| Figura 2.5. Conjunto de capas en un SIG. | 33 |
| Figura 2.6. Datos representados a partir de sus coordenadas en latitud – longitud. | 34 |
| Figura 2.7. Representación gráfica de los modelos vectorial y raster. | 35 |
| Figura 3.1. Obtención del registro de observaciones. | 38 |
| Figura 3.2. Resultado del Análisis de Correspondencias Múltiple. | 50 |
| Figura 4.1 Esquema del proceso de clasificación. | 53 |
| Figura 4.2. Ejemplo del formato correcto de los datos de entrada. | 55 |
| Figura 4.3. Porcentajes de la condición de déficit habitacional. | 67 |
| Figura 4.4. Influencia de los indicadores de déficit habitacional. | 68 |
| Figura 5.1. Esquema del proceso de vinculación de los datos al mapa. | 72 |
| Figura 5.2. Capa de las parroquias de Mérida y su tabla atributiva. | 74 |
| Figura 5.3. Selección a partir de los atributos de las tablas. | 75 |
| Figura 5.4. Mapa para impresión y sus elementos. | 77 |
| Figura 5.5. Relación de las tablas y su cardinalidad. | 78 |
| Figura 5.6. Selección por atributos del control 1011. | 79 |
| Figura 5.7. Vista en las tablas de atributos de los elementos consultados. | 79 |
| Figura 5.8. Estadísticas del déficit habitacional para el control 1011. | 80 |
| Figura 5.9. Estadísticas del indicador Número de Hogares para el control 1011. | 81 |
| Figura 5.10. Estadísticas del indicador Tenencia para el control 1011. | 81 |

| | |
|---|-----|
| Figura 5.11. Estadísticas del indicador Hacinamiento para el control 1011..... | 82 |
| Figura 5.12. Estadísticas del indicador Condición para el control 1011. | 82 |
| Figura 5.13. Estadísticas del indicador Adecuación para el control 1011..... | 83 |
| Figura 5.14. Arquitectura de <i>MapServer</i> | 85 |
| Figura 5.15. Esquema del archivo mapa..... | 86 |
| Figura. 5.16. Página principal del visor de mapas. | 88 |
| Figura b.1. Número de conglomerados para variable Condición de la vivienda. ... | 101 |
| Figura b.2. Porcentaje dentro del conglomerado de la variable Tipo de vivienda. | 102 |
| Figura b.3. Porcentaje dentro del conglomerado de la variable Techo..... | 102 |
| Figura b.4. Porcentaje dentro del conglomerado de la variable Paredes. | 103 |
| Figura b.5. Porcentaje dentro del conglomerado de la variable piso..... | 103 |
| Figura b.6. Número de conglomerados para variable adecuacion de la vivienda. | 104 |
| Figura b.7. Porcentaje dentro del conglomerado de la variable Suministro de agua. | 104 |
| Figura b.8. Porcentaje dentro del conglomerado de la variable Eliminación de excretas. | 105 |
| Figura b.9. Porcentaje dentro del conglomerado de la variable Servicio eléctrico. | 105 |
| Figura b.10. Porcentaje dentro del conglomerado de la variable Recolección de basura..... | 106 |
| Figura b.11. Porcentaje dentro del conglomerado de la variable Contenedor de basura..... | 106 |
| Figura e.1. Interfaz gráfica del visor web. | 125 |
| Figura e.2. Detalle de la tabla de contenidos y leyenda. | 129 |
| Figura e.3. Detalle de la herramienta de búsqueda. | 130 |
| Figura e.4. Selección de los elementos tras la búsqueda realizada..... | 131 |
| Figura e.5. Tabla de Atributos del segmento 1126..... | 132 |
| Figura e.6. Gráfica de las estadísticas del indicador Tenencia..... | 133 |

A mi Familia;
Mis Abuelos, Padres y Hermanos.

Agradecimientos

Dedico estas líneas para agradecer de manera muy especial a aquellas personas quienes permitieron el buen desarrollo de esta investigación y aquellos quienes fueron de gran apoyo y motivación para lograr la culminación de tan importante proyecto:

A Dios Todo Poderoso y a la Virgen de Coromoto por darme el maravilloso don de la vida, guiar mis pasos y permitirme estar donde estoy.

A mi familia a quienes dedico este trabajo, por creer en mí, por su apoyo, amor y cariño incondicional. Por darme los más importante en la vida que son los valores y el buen ejemplo.

A mis tutores el profesor Gerardo Colmenares y Alexander Parra por su confianza, amabilidad, profesionalismo y por impartirme las herramientas necesarias para lograr tan anhelada meta.

Al personal del Instituto Nacional de Estadística, en especial al Licenciado Luis Muñoz, el Geógrafo Pablo Fernández y la Licenciada Doris Matheus, por su atención, dedicación y colaboración la cual fue de gran aporte para el desarrollo y culminación de este proyecto.

A mis amigos y compañeros de clase, por su amistad y compañerismo, por el apoyo moral y por el aprendizaje que me llevo de muchos de ustedes gracias por estar siempre allí, les tengo mucho cariño y admiración.

Capítulo 1

Introducción

La vivienda representa un componente fundamental en cualquier sociedad, ya que es el albergue y morada para un individuo o grupo de individuos, proporcionándoles protección, bienestar y seguridad. Cuando existe carencia de vivienda o no se encuentra en condiciones de ser habitada, se produce el fenómeno del déficit habitacional. Una situación problemática donde las necesidades de viviendas potencialmente requeridas por un cierto grupo social superan a las viviendas existentes.

Actualmente los indicadores que determinan en el país el déficit habitacional no son los más apropiados. La consecuencia es una ausencia de políticas por parte de las instituciones y organismos encargados de gestionar y planificar proyectos habitacionales que mitiguen o resuelvan la problemática. Una manera que contribuye a indicar cuánto y dónde mitigar el déficit es con una información confiable y métodos precisos para la medición de este tipo de déficit. Por tal razón, con esta investigación, se pretende crear una aplicación que a partir del procesamiento cuidadoso y sistemático de datos provenientes de la Encuesta de Hogares por Muestreo (EHM), se pueda realizar una medición eficiente y a su vez detallada del déficit habitacional en el Área Metropolitana de Mérida (AMM). El ámbito de estudio que define el AMM comprende los municipios Libertador, Campo Elías, Sucre y Santos Marquina. Según el último censo nacional de población

en el área circunscrita al AMM, se contabilizó 345491 individuos; esto representa el 48% de la población del todo el estado Mérida. Por otro lado, el número de viviendas en estos cuatro municipios es de 96.977, representando esta cifra el 48% de las unidades habitacionales existentes en todo el estado. El tamaño de la muestra de la EHM a nivel semestral para esta área comprende aproximadamente unas 1050 viviendas, por lo que vale la pena precisar que con esta investigación se pretende realizar estimaciones e inferencias de acuerdo a esta muestra en torno al fenómeno del déficit habitacional.

Son dos los conceptos del déficit habitacional sustentados en esta investigación. El primero de ellos tiene que ver con la carencia absoluta de la unidad habitacional, al cual se le define como déficit cuantitativo. El segundo tipo de déficit tiene que ver con la situación en que se encuentra la vivienda respecto a ciertos indicadores como lo son la condición física de la vivienda (infraestructura), tenencia, hacinamiento y adecuación o acceso a redes de servicios públicos; a este tipo de déficit se le conoce como déficit cualitativo.

Al abordar la problemática de la vivienda desde un enfoque que contemple la calidad de vida de la población, podrá aceptarse que el contar con un lugar para vivir es sólo uno de los varios factores necesarios para aproximarse a un nivel de bienestar aceptable. Los moradores de un inmueble requieren tener ciertas comodidades, ya que la construcción física a habitar se transformará en “un espacio de uso familiar y un bien de consumo básico al que todos tienen derecho si se busca una elemental justicia social”. Por esta razón, un hogar representa mucho más que la mera estructura material (García, 2006).

Existen diversos aspectos que son considerados para sustentar la condición de adecuación de la vivienda. Las referencias obligatorias que está relacionadas a este concepto social son: a) el basamento jurídico establecido en la Ley de Régimen Prestacional de Vivienda y Hábitat (Gaceta oficial N° 38.192 del 9 de Mayo de 2005) en su Capítulo III. De los objetivos de la Ley. Art. 13 "*Vivienda y Hábitat Dignos*" y de la Constitución de la República Bolivariana de Venezuela en su Artículo 82; b) el Contenido de Derecho discutido por la Comisión de los Derechos Humanos de las Naciones Unidas en la Observación N° 4, en las que se explica las condiciones en las cuales deben vivir los individuos, la habitabilidad de las viviendas para ser aceptadas por ellos (Gil, 2008).

De esta manera se define el ámbito conceptual que sustenta el desarrollo de este trabajo y que busca ofrecer mecanismos para afrontar el tema del déficit de viviendas en función de que se cumplan cada una de premisas previamente planteadas.

Para la consecución del objetivo deseado, como es la clasificación y la cuantificación del déficit habitacional, se enfoca una metodología que, en líneas generales, comprende el proceso de procesamiento y exploración de datos provenientes de la EHM para construir una serie de indicadores de déficit a través del uso de técnicas de análisis multivariante de datos. Adicionalmente, un algoritmo de aprendizaje automático a partir de las máquinas de vectores de soporte permitirá la construcción de un modelo de clasificación y pronóstico del déficit de viviendas, y por último, los resultados serán integrados a un Sistema de Información Geográfica con el objeto de realizar una representación y análisis espacial del déficit.

1.1 Antecedentes

Según INE (2008), la EHM es una investigación de naturaleza estadística y propósitos múltiples que se viene realizando en el país de forma ininterrumpida desde los años sesentas. Surgió como respuesta a la necesidad de disponer, en los períodos inter-censales, de información sobre la estructura y evolución del mercado de trabajo y las características socioeconómicas de la población, especialmente en relación con variables susceptibles a modificaciones significativas en el corto plazo. Una de las finalidades de la EHM es proporcionar a través de encuestas complementarias información sobre la persona, el hogar y la vivienda, para el análisis de temas específicos que apoyen en la decisión a los organismos oficiales y los usuarios que así lo requieran. Esta encuesta contiene datos históricos y contiene una serie de variables con información de gran importancia para el presente estudio. Parte del apoyo referencial para esta investigación lo proporciona una serie de investigaciones que han utilizado la EHM para la medición del déficit habitacional:

Gil (2008), desarrolló un estudio denominado "Pronóstico del déficit de viviendas en el Estado Mérida a través de redes neuronales artificiales", con el fin de determinar el déficit habitacional en el Estado Mérida, mediante métodos heurísticos que permiten proyectar la necesidad global existente.

Linares (2004), condujo un estudio en base a los resultados arrojados por los censos generales de población y vivienda, el último de los cuales se realizó en el año 2001. Con base en estos resultados, otros reportes y publicaciones del INE, precisó cifras aproximadas del déficit habitacional en el ámbito nacional para el año 2004. De manera que este trabajo permitió conocer las estimaciones del déficit habitacional

para los años 2001 y 2004 a fin de tener una aproximación al fenómeno que se desea estudiar.

Szalachman (1999), condujo una investigación la cual consistió en realizar estimaciones del déficit de vivienda tanto cuantitativo como cualitativo en Bolivia. Para ello se construyeron indicadores en base a conceptos nuevos y utilizando información que no se había empleado para este propósito con anterioridad. La información necesaria para el estudio se tomó a partir de las encuestas de hogares y de una serie de estudios relacionados con el tema en cinco países de América Latina: Bolivia, Brasil, Colombia, Chile y Uruguay. Es relevante este estudio debido a que proporcionó a esta investigación los conceptos fundamentales y orientó la formulación de los indicadores que se emplearon.

Es importante de igual manera hacer referencia a estudios relacionados al déficit habitacional donde se haya utilizado herramientas de análisis espacial como es el caso de (Quintero, 2000). En este trabajo se realizó un estudio detallado del déficit de viviendas en el espacio que constituye la ciudad de Bogotá y se estimó la demanda actual de viviendas, así como también el déficit relacionado a las viviendas de interés social. Cabe señalar que este trabajo de acuerdo con la revisión bibliográfica realizada y al enfoque que plantea no tiene precedente alguno en el ámbito nacional.

1.2 Planteamiento del problema

Actualmente el déficit habitacional supera el millón ochocientas mil viviendas, y sesenta por ciento de las existentes necesita ser mejorada o ampliada (INE, 2008). El

Estado Mérida no escapa de este fenómeno que afecta a sus habitantes y continúa avanzando a pesar de los intentos de los organismos gubernamentales para resolver tal situación. La carencia de procedimientos expeditos para la planificación, ejecución y posterior adquisición de unidades habitacionales en cualquier ámbito, (nacional, regional o local), han originado un fenómeno con altos índices de insatisfacción en todas las capas sociales en desmedro de perjudicar el nivel general de calidad de vida.

La problemática se acentúa, más aún, por el hecho que no existe una aplicación que permita detectar de manera eficiente cuáles son los sectores más afectados y cuál es el registro de la información de déficit asociado a estos casos. En el presente trabajo, se plantea un estudio que permitirá medir la densidad del déficit tanto estadística como especialmente. Es importante destacar que con Gil (2008) se abre una oportunidad a explorar más específicamente las condiciones de déficit y los espacios donde éste ocurre con mayor densidad.

Inicialmente, para el desarrollo del estudio se dispone de un conjunto de datos procedentes de la EHM que comprende el período 2001-2007. La información que contiene dichos datos es muy diversa e incluye características de gran utilidad para el estudio. Por ejemplo, información referida a la tenencia, al hacinamiento, a la adecuación de la vivienda y por último, a las condiciones estructurales.

Se desea realizar un análisis estadístico a dichos datos para, en primera instancia, determinar ciertos patrones o variables que posteriormente van a permitir realizar la clasificación del déficit habitacional. Mediante la técnica de clasificación denominada Máquinas de Vectores de Soporte (MVS) se medirá en qué condiciones

de déficit se encuentra una determinada vivienda a partir de las variables a considerar. Posterior a la generalización del modelo sobre el dominio global del problema real para el Área Metropolitana de Mérida, se aplicarán herramientas de análisis espacial a estos resultados para cuantificar la densidad en el espacio de este fenómeno. Al mismo tiempo, se creará una herramienta híbrida que contemple tanto MVS como de Sistemas de Información Geográfica (SIG), mediante la implementación de un sistema geográfico para la detección de déficit de viviendas.

1.3 Objetivos

1.3.1 Objetivo general

Clasificar, cuantificar y representar el déficit de viviendas en el espacio que define el Área Metropolitana de Mérida.

1.3.2 Objetivos específicos

1. Utilizar técnicas para el análisis estadístico de datos multivariante obtenidos de la EHM y extraer patrones sobre las variables que tengan dependencia con el déficit habitacional.
2. Aplicar la técnica de MVS para la construcción de un modelo que permita clasificar el déficit habitacional en el Área Metropolitana de Mérida.
3. Implementar un SIG para realizar la representación y a su vez el análisis espacial a la información clasificada previamente.

4. Utilizar los métodos de análisis espacial para medir el impacto de variables características del déficit (Adecuación de la vivienda, Condición de la vivienda, Hacinamiento de la vivienda, Tenencia de la vivienda, entre otras) en el espacio correspondiente al Área Metropolitana de Mérida.

5. Implementar una herramienta híbrida que contemple tanto MVS y SIG, para la detección de déficit de viviendas, mediante la formulación de un procedimiento de ejecución paso a paso.

1.4 Justificación

Linares (2004), menciona "... en la actualidad cuantificar el déficit de viviendas es un ejercicio que origina discrepancias y controversias, debido a que los diversos métodos empleados siempre tienen un margen de imprecisión y duda". Por otro lado, según García (2000), "... las técnicas de análisis de vivienda, en el ámbito nacional regional o local, con la finalidad de estimar la necesidad, el déficit y la demanda, han sido poco desarrolladas en el país. En gran medida, la principal razón se debe a la ausencia de información y de datos confiables para un estudio exhaustivo y detallado". Esta investigación se nutre de estas referencias para elaborar un procedimiento que mejore a lo expuesto por Linares y que mitigue la preocupación de García.

En tal sentido y por las razones mencionadas previamente, se implementará una aplicación que mida de manera práctica y efectiva el déficit cuantitativo y cualitativo de viviendas. De igual manera, brinde la posibilidad a organismos

públicos o privados a planificar y gestionar en cuanto a situación habitacional se refiere.

1.5 Alcances y Limitaciones

Este estudio tal como se plantea no sólo pretende hacer una estimación general de déficit de viviendas para el Área Metropolitana de Mérida, sino también, crear una herramienta eficaz, práctica y de vanguardia tecnológica para hacer una medición del impacto del déficit habitacional, combinando técnicas estadísticas y minería de datos con las bondades que ofrecen los sistemas de información geográfica. Esta combinación permitirá realizar la clasificación del déficit y medir su densidad, mediante la representación y la aplicación de análisis espacial, con el fin de detectar cuáles son las áreas más afectadas donde el déficit de viviendas tenga mayor influencia.

El valor agregado que proporciona esta combinación de técnicas puede ser aprovechado en el desarrollo de proyectos similares donde se requiera la combinación de estas técnicas para la planificación y gestión. No sólo del déficit de viviendas, sino de otros fenómenos que afecten este déficit mediante los resultados finales e intermedios provistos por las diversas técnicas estadísticas empleadas y expuestos en mapas específicos, georeferenciando la información resultante del análisis estadístico.

Entre las limitaciones se destaca la disposición de los datos de la EHM. La encuesta recopila un gran número de variables, muchas de ellas con variaciones en su denominación en el tiempo y en consecuencia, dificultan su organización e

interpretación. Esto conduce a una revisión cuidadosa de la información que contiene la encuesta y cada una de las variables, para cada período encuestado.

Otra limitación que vale la pena mencionar viene dada por el nivel de detalle de la representación espacial de la muestra. La EHM define su unidad muestral a través de Lotes (áreas geográficas de aproximadamente 15 viviendas); sin embargo, el detalle disponible de información permite obtener resultados hasta el nivel de los segmentos, que es el vínculo espacial o mínima área geográficas delimitada por el INE para el levantamiento de las encuestas. De igual modo dentro de los lotes existe una sustitución constante de viviendas debido a las rotaciones que se hacen a la muestra en cada recolección, conservando las características de la unidad pero, variando su contenido. Esta variabilidad limita la selección de la unidad específica y la condiciona a que tenga que ser realizada por lotes que no incluye una georeferencia que lo pueda ubicar en cada uno de los segmentos. De este modo, se condiciona el estudio espacial a un área mínima de evaluación a través de los segmentos. De ahí que los segmentos permanecen invariables en el tiempo, mientras las viviendas dentro del mismo pueden variar semestralmente.

Capítulo 2

Marco Referencial

Antes de abordar la parte experimental de los métodos y procedimientos que se aplicarán para la consecución de cada uno de los objetivos que plantea este estudio, es necesario tomar en cuenta una serie de consideraciones y aspectos que ayuden a entender la esencia del mismo. Por lo que en este capítulo se detallarán los fundamentos y consideraciones teóricas en las que se basa el presente trabajo.

Básicamente se abordan los tópicos más relevantes y se especifican aspectos fundamentales en cuanto al déficit habitacional, la estructura de la encuesta de hogares, así como también, las técnicas de procesamiento de datos, la metodología MVS, técnica empleada para la clasificación del déficit habitacional. Del mismo modo, se revisan algunas generalidades de los SIG y de las bondades de esta tecnología para la representación y análisis espacial de datos.

2.1 Conceptos de déficit de viviendas

Toda persona de manera individual o en conjunto, requiere de un espacio donde albergarse de la intemperie; no obstante, dicho espacio debe contar con requerimientos mínimos de salubridad y seguridad que le garanticen bienestar. Es por ello que la vivienda juega un aspecto muy importante en la calidad de vida de

las personas y conlleva, en la mayoría de los casos, al acceso a servicios hoy considerados esenciales para alcanzar niveles mínimos de bienestar.

El Comité de expertos de la Organización Mundial de la Salud (OMS) en Higiene de la Vivienda, señala en su primer informe relacionado a la vivienda: “Es la estructura material que el hombre emplea para cobijarse, y sus dependencias, es decir; todos los servicios, instalaciones y dispositivos necesarios o convenientes para el bienestar social y la salud física y mental del individuo y la familia.”

Basados en estas definiciones se tiene una noción de lo que puede ser una vivienda apta o en condiciones idóneas para albergar un hogar; el déficit resulta cuando esas condiciones no se satisfacen. Básicamente el déficit habitacional es el indicador cuantitativo y cualitativo del problema de vivienda. Se puede presentar alguna de estas condiciones que califican al déficit: familias que no disponen de una vivienda, familias que no tienen acceso al derecho social de tener condiciones óptimas de habitabilidad, o ambas.

Como se puede apreciar el déficit habitacional es un factor condicionante en la calidad de vida de las personas, y son diversos factores los que se consideran al momento de realizar una medición o estimación del mismo. De manera que el déficit de viviendas no sólo se mide o cuantifica a partir de la carencia absoluta de ella, sino que se debe tomar en consideración ciertas condiciones básicas que permitan garantizar la seguridad y el bienestar de quienes habitan en ella. Tal como se plantea el concepto de déficit existen dos tipos de déficit conforme a los dos aspectos sobre la situación de vivienda.

2.1.1 Déficit cuantitativo

Según Szalachman (1999), la definición tradicional de “déficit cuantitativo” de vivienda se basa en la comparación entre el número de hogares y el de viviendas existentes. El momento en el cual la primera de estas cifras supera la segunda es lo que en la mayoría de los textos se designa como déficit cuantitativo. Sin atenuar la importancia de la definición anterior, este trabajo sostiene que éste es tan sólo un primer acercamiento a lo que se desea medir.

Basados en esta definición convenientemente nos vamos a referir al déficit cuantitativo como el conjunto de viviendas que es necesario construir para dotar a los hogares y núcleos "allegados". Esta situación se manifiesta en la convivencia de dos o más hogares en una vivienda.

2.1.2 Déficit cualitativo

La definición clásica de déficit cualitativo engloba a aquellas viviendas cuya calidad es insatisfecha en relación con ciertos criterios básicos. Tres variables son tomadas generalmente en cuenta para caracterizar este déficit: la condición de la vivienda, el hacinamiento y la disponibilidad de servicios públicos (Szalachman, 1999). En este trabajo se incluye una cuarta variable y está relacionada con el grado de propiedad de la vivienda o tenencia. Para entender la influencia de cada una de estas variables en la medición del déficit cualitativo, a continuación se definirán cada una de ellas.

Condición de la vivienda: La condición de la vivienda tiene que ver con los componentes constructivos; de manera que para definir ésta variable se toma en cuenta los materiales con que fueron construidos su techo, paredes y piso, para así poder determinar en cuales condiciones de infraestructura se encuentra la vivienda.

Hacinamiento: La condición de hacinamiento resulta de relacionar la variable referida a número de personas que vive en una vivienda con la variable número total de habitaciones que tiene dicha vivienda, sin contar el baño, cocina ni pasillos. Se calcula como el cociente entre ambas variables. Para el estudio de déficit habitacional, se estimó que existe hacinamiento cuando este cociente es mayor a dos; es decir, cuando en una vivienda hay más de dos personas por dormitorio (Szalachman, 1999).

Tenencia: Es la condición jurídica de propiedad de la vivienda por parte del hogar o los hogares que habitan en ella. A modo de ejemplo una vivienda puede ser: propia, alquilada, cedida, propia pagándose, etc.

Adecuación o disponibilidad a servicios públicos: La adecuación de una vivienda la determina el acceso o disponibilidad de la misma a servicios públicos básicos tales como; agua, luz, aseo y eliminación de excretas.

Estos indicadores, aun existiendo la vivienda, son los que permiten medir más detalladamente si presenta las condiciones óptimas de habitabilidad. Las cuatro variables por sí solas o en conjunto son determinantes claves de la calidad de vida de las personas.

2.2 Estructura de la Encuesta de Hogares por Muestreo

2.2.1 Encuesta de Hogares por Muestreo

De acuerdo con la ficha técnica, suministrada por el INE respecto a la EHM, se establece que es una investigación de naturaleza estadística y propósitos múltiples, la cual surgió como respuesta a la necesidad de disponer, en los períodos ínter

censales, de información sobre la estructura, evolución del mercado de trabajo y las características socioeconómicas de la población, especialmente en relación con variables susceptibles a modificaciones significativas en el corto plazo. A partir de las variables investigadas en la encuesta, se producen indicadores como la tasa de actividad económica, tasa de desempleo, tasa de cesantía, porcentaje de ocupados en el sector informal. Por el contrario, no se observa indicadores relacionados a la vivienda y los hogares.

La EHM tiene como finalidad, proporcionar en forma oportuna y confiable, a los organismos de planificación nacional y regional, información sobre la estructura y evolución de la fuerza de trabajo y, en particular, sobre el fenómeno de empleo y el desempleo, como insumo para el análisis que pueda derivar en una toma de decisiones. También a través de encuestas complementarias, puede proporcionar información sobre la persona, el hogar, y la vivienda, para el análisis de temas específicos a los organismos oficiales y los usuarios que así lo requieran.

2.2.2 Unidades estadísticas de la EHM

Son cuatro las unidades estadísticas que forman parte de esta encuesta; la unidad de investigación, la unidad de análisis, la unidad de observación y las unidades de muestreo. La primera de estas se basa en el hogar, que es quien contiene a las personas que se analizan. La unidad de análisis viene dada por las personas, los hogares y las viviendas de las cuales se busca la información que se requiere en el estudio. La unidad de observación está conformada por el informante, definido por la EHM como cualquier miembro del hogar mayor de 14 años que se encuentre en capacidad de suministrar información relativa a las personas que habitan en su hogar. Por último, las unidades de muestreo que teóricamente debería ser la

vivienda, son definidos como lotes, los cuales son áreas geográficas de aproximadamente 15 viviendas en las cuales conforman los segmentos/sectores (INE 2008).

2.2.3 Marco muestral y tipo de muestreo

El marco muestral utilizado en la EHM está constituido por la Muestra Maestra del INE, construida a partir del registro de edificación (zona urbana) y expedientes parroquial rural (zona rural), información obtenida del Censo Nacional de Población y Vivienda que se realiza cada 10 años.

El tipo de muestreo utilizado para la selección de la muestra de la EHM fue probabilístico, el cual exige que todos los elementos del universo estadístico tengan una probabilidad conocida y no nula de ser seleccionados. La clase de muestreo fue estratificado bifásico. Los estratos de selección es el área rural de cada entidad y grupos dentro del área urbana de cada entidad.

De la muestra a nivel nacional se asigna una proporción por estados y la entrevista se aplica a todos y cada uno de los hogares pertenecientes a las viviendas y dentro de cada hogar, a cada uno de los habitantes habituales. Por lo tanto, los hogares que conforman la muestra son todos aquellos pertenecientes a las viviendas seleccionadas. De acuerdo con la ficha técnica de la EHM, el total de lotes en la muestra es de 3175, equivalente a 45.000 viviendas. Se considera que el tamaño de esta muestra es suficiente para generar resultados para la población total del país. Sin embargo, no se logra la entrevista para todas las viviendas, por alguno de los siguientes motivos: viviendas desocupadas, en ruinas, de uso vacacional, utilizadas

como almacén o negocio, o porque en aquellas que están ocupadas no está la persona que puede dar información o en caso contrario se niegan a darla.

La EHM se realiza con periodicidad semestral. Hasta el año 2001 en su primer semestre las viviendas permanecían en la muestra durante seis semestres consecutivos; a esto se le conoce cómo “Rotación de la Muestra”, el cual consistía en que cada semestre se reemplazaba 1/6 de las viviendas de la muestra, que fueron entrevistadas durante seis semestres seguidos; esta proporción sustituye las reemplazables en un nuevo grupo de vivienda, incorporándose al panel de muestra. A partir del segundo semestre del 2001 se cambió esta modalidad de rotación por una que es aplicada a aquellas viviendas que sólo hayan sido encuestadas durante los últimos tres años y ya muestren cansancio los informantes, rehusándose a dar información tal que esto implique un aumento en la tasa de no respuesta en dicha entidad.

2.2.4 Variables investigadas

Las variables investigadas en la encuesta se pueden clasificar en los siguientes grupos:

Características de la vivienda: se refiere a la tipología constructiva, tenencia y densidad habitacional.

Características del hogar: se trata de identificar los servicios disponibles para el hogar, características de la familia.

Variables sociodemográficas: involucran un conjunto de características referidas a las unidades familiares y a la población.

VARIABLES DE CARACTERÍSTICAS EDUCATIVAS: identifican el conjunto de condiciones de vida vinculables a la incorporación de la población al sistema escolar formal y a los sistemas no formales, particularmente los de capacitación para el trabajo.

VARIABLES DE CONDICIÓN DE ACTIVIDAD Y FUERZA DE TRABAJO: interesa identificar aquellos aspectos básicos socio-ocupacionales que se vinculan claramente con las condiciones de vida de los hogares.

De todas estas variables, sólo se tomarán en consideración para este estudio las primeras dos, que son las que aportan la información relevante con la que se pretende indagar y analizar el fenómeno del déficit de viviendas.

2.3 Análisis Exploratorio de Datos (AED)

De acuerdo a Spiegel (1961), se puede mencionar que el AED consiste en la aplicación de una serie de herramientas que permiten observar los datos de una manera informal, para obtener una impresión inicial de los mismos. Entre los objetivos de un AED cabe resaltar los siguientes:

- Describir los datos.
- Investigar acerca de la calidad de los mismos.
- Buscar estructuras o patrones sin realizar ninguna hipótesis acerca de la estructura de estas observaciones o variables.
- Examinar los datos previamente al uso de la aplicación de cualquier técnica estadística.
- Obtener un conocimiento básico de los datos y de las relaciones entre las variables.

2.3.1 Etapas del Análisis Exploratorio de Datos

Las etapas para la realización de un buen AED tienen que ver básicamente con el establecimiento de los objetivos de la investigación, recolección y preparación de datos, investigación de la estructura y calidad de los datos, análisis inicial de los datos, selección y aplicación de un análisis estadístico formal apropiado, validación e interpretación de los resultados. La aplicación de manera secuencial de cada uno de estos pasos permite la consecución de resultados en cualquier estudio de naturaleza estadística.

2.4 Técnicas de Análisis Multivariante de Datos

Hair (1999), asegura que sólo a través del análisis multivariante, las relaciones múltiples entre las variables podrán ser examinadas adecuadamente para obtener un entendimiento más completo y real del entorno que permita tomar las decisiones más adecuadas.

A continuación se presentan conceptos de algunos de estos métodos que fueron utilizados para la creación de algunas variables indicadoras de déficit.

2.4.1 Análisis de conglomerados

El análisis de conglomerados (clusters) tiene por objeto agrupar elementos en grupos homogéneos en función de las similitudes o similitudes entre ellos (Peña, 2002). Se convierte así en una técnica de análisis exploratorio diseñada para revelar las agrupaciones naturales de los datos.

Este análisis no hace ninguna distinción entre variables dependientes y variables independientes sino que calcula las relaciones interdependientes de todo el conjunto de variables. El método trata de situar los casos en grupos homogéneos, de

manera que objetos que puedan ser considerados similares y sean asignados a un mismo grupo, mientras que objetos distintos se sitúan en grupos diferentes (Pérez, 2004).

Una gran ventaja de este análisis es que puede utilizarse como instrumento de reducción general de datos a fin de desarrollar subgrupos de datos que sean más fáciles de manejar que las observaciones individuales. El análisis multivariante subsecuente se realiza con base en los subgrupos, en lugar de las observaciones individuales.

2.4.1.1 Medidas de proximidad y de distancia

Una vez establecidas las variables y los objetos a clasificar es necesario establecer una medida de proximidad o de distancia entre ellos que cuantifique el grado de similaridad entre cada par de objetos. Las medidas de proximidad y similitud miden el grado de semejanza entre dos objetos de forma que, cuanto mayor es su valor, mayor es el grado de similaridad existente entre ellos y con más probabilidad los métodos de clasificación tenderán a ponerlos en el mismo grupo. Por el contrario, las medidas de disimilitud o distancia miden la distancia entre dos objetos de forma que, cuanto mayor sea su valor, más diferentes son los objetos y menor la probabilidad de que los métodos de clasificación los pongan en el mismo grupo (Salvador, 2001).

2.4.2 Análisis de correspondencias

Técnica estadística cuya finalidad es poner de manifiesto gráficamente las relaciones de dependencia existentes entre las diversas modalidades de dos o más variables categóricas, es decir cualitativas, a partir de la información proporcionada por sus tablas de frecuencias cruzadas o tablas de contingencia. Para ello asocia a cada

modalidad un punto en el espacio \mathbf{R}^k , de forma que, cuanto más alejado del origen de coordenadas está el punto asociado a una modalidad de una variable, más diferente es su perfil condicional del perfil marginal correspondiente a las otras variables; además, los puntos correspondientes a dos modalidades diferentes de una misma variable estarán más cercanos cuanto más se parezcan sus perfiles condicionales y, finalmente, dichos puntos tenderán a estar más cerca de aquellas modalidades con las que tienen una mayor afinidad; es decir, aquellas en las que las frecuencias observadas de la celda correspondiente tiende a ser mayor que la esperada bajo la hipótesis de independencia de las variables correspondientes. Cuando involucra más de dos variables recibe el nombre de análisis de correspondencias múltiple (ACM) o análisis de homogeneidades. Esta técnica estadística es de gran utilidad puesto que la interpretación del resultado puede hacerse de manera sencilla a través de gráficas. Con este procedimiento se puede evidenciar de manera más perceptible el grado de relación entre las categorías de cada variable; de ahí el nombre de mapas perceptuales. Cuando el grado de asociación es alto, éstas parecerán en el diagrama relativamente juntas (Salvador, 2001).

2.5 Máquina de Vectores de Soporte

La teoría de las Máquinas de Vectores de Soporte (SVM por su nombre en inglés Support Vector Machine) fue desarrollada por Vapnik basado en la idea de Minimización del Riesgo Estructural (MRE), a diferencia de las Redes Neuronales Artificiales (RNA) que utilizan durante la fase de entrenamiento, el principio de minimización del riesgo empírico (mre). El principio de Minimización del Riesgo Estructural ha mostrado un mejor desempeño, las Máquinas de Vectores de Soporte aplicando este principio minimizan un límite superior al riesgo esperado a

diferencia del principio de minimización del riesgo empírico que minimiza el error sobre los datos de entrenamiento (Vapnik, 2000). La MVS mapea los puntos de entrada a un espacio de características de una dimensión mayor, para luego encontrar el hiperplano que los separe y maximice el margen entre las clases. Pertenecen a la familia de clasificadores lineales puesto que inducen separadores lineales o hiperplanos en espacios de características de muy alta dimensionalidad (introducidos por funciones núcleo o *kernel*) con un sesgo inductivo muy particular la maximización del margen (Carreras, Márquez y Romero, 2004). Sin embargo, la formulación matemática de las Máquinas de Vectores Soporte varía dependiendo de la naturaleza de los datos; es decir, existe una formulación para los casos lineales y, por otro lado, una formulación para casos no lineales. Es importante tener claro que, de manera general para clasificación, las máquinas de vectores soporte buscan encontrar un hiperplano óptimo que separe las clases.

Las MVS han sido desarrolladas como una técnica robusta para clasificación y regresión para grandes conjuntos de datos complejos con ruido; es decir, con variables inherentes al modelo, que para otras técnicas aumentan la posibilidad de error en los resultados pues resultan difíciles de cuantificar y observar. Algunas de las aplicaciones de clasificación o reconocimiento de patrones son: reconocimiento de firmas, reconocimiento de imágenes como rostros y categorización de textos y en este trabajo, clasificación del déficit habitacional.

2.5.1 MVS para el caso no linealmente separable

Las MVS ofrecen la ventaja de que las mismas pueden ser utilizadas para resolver tanto problemas lineales como no lineales. En el caso de ser linealmente separable, las MVS conforman hiperplanos que separan los datos de entrada en dos subgrupos

que poseen una etiqueta propia. En medio de todos los posibles planos de separación de las dos clases etiquetadas como $\{-1, +1\}$, existe sólo un hiperplano de separación óptimo, de forma que la distancia entre el hiperplano óptimo y el valor de entrada más cercano sea máxima (maximización del margen) con la intención de forzar la generalización de la máquina que se esté construyendo. Aquellos puntos o ejemplos sobre los cuales se apoya el margen máximo son los denominados vectores de soporte. Un ejemplo de este caso se puede observar en la figura 2.1.

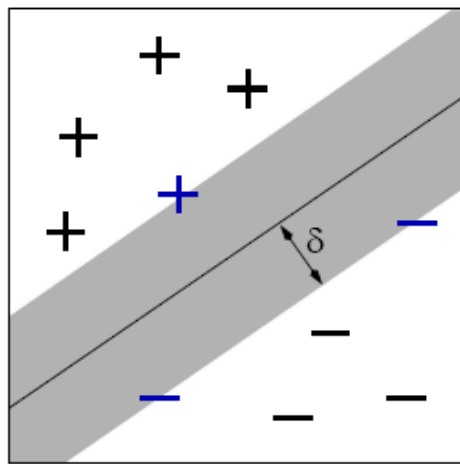


Figura 2.1. MVS Linealmente separable.

Para el caso no lineal existen dos casos que vale la pena mencionar, el primero de estos se presenta cuando los datos pueden ser separables con margen máximo pero en un espacio de características (el cual es de una mayor dimensionalidad y se obtiene a través de una transformación a las variables del espacio de entrada) mediante el uso de una función *kernel*. El segundo caso especial de las MVS denominado “*Soft Margin*” o margen blando, es utilizado cuando no es posible encontrar una transformación de los datos que permita separarlos linealmente, bien sea en el espacio de entrada o en el espacio de características.

2.5.1.1 MVS con margen máximo en el espacio de características

Hay casos donde los datos no pueden ser separados linealmente a través de un hiperplano óptimo en el espacio de entrada. En muchas situaciones los datos, a través de una transformación no lineal del espacio de entradas, pueden ser separados linealmente pero en un espacio de características y se pueden aplicar los mismos razonamientos que para las MVS lineal con margen máximo. La transformación de los datos de un espacio inicial a otro de mayor dimensión se logra mediante el uso de la función *kernel*. Una función núcleo o *kernel* es un producto interno en el espacio de características, que tiene su equivalente en el espacio de entrada (Gunn, 1997).

$$K(x, x') = \langle \Phi(x), \Phi(x') \rangle \quad (2.1)$$

En la ecuación 2.1 K es una función simétrica positiva definida que cumple las condiciones de Mercer (mayor detalle se puede ver en (Gunn, 1997)).

De manera gráfica se puede observar en la figura 2.2 como la función kernel permite realizar la separación y el traslado de los datos al espacio de características.

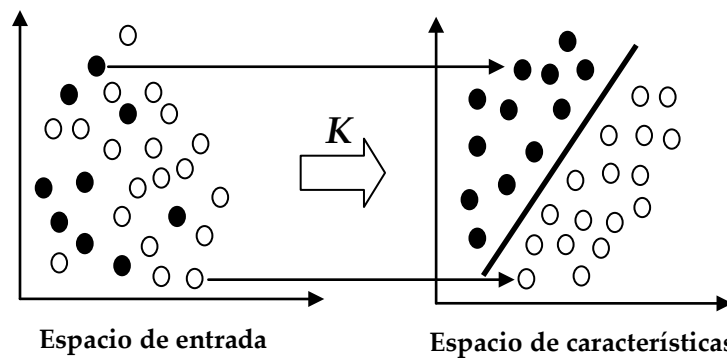


Figura 2.2. MVS No linealmente separable inducida por una función *kernel*.

Entre los *kernels* más comunes, se encuentran: la función lineal, polinomial, RBF (Radial Basis Function), ERBF (Exponential Radial Basis Function), entre otros (Gunn, 1997).

El problema de optimización a resolver para las MVS con margen blando está definido por un modelo de programación cuadrática con restricciones; es decir:

$$\begin{aligned} \text{Maximizar} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N y_i y_j \alpha_i \alpha_j K(x_i, x_j) \\ \text{Sujeto a:} \quad & \sum_{i=1}^N \alpha_i y_i = 0 \\ & \alpha_i \geq 0, 1 \leq i \leq N \end{aligned} \quad (2.2)$$

donde $K(x, y)$ es la función núcleo.

De acuerdo con Benavidez (2006), este problema de optimización se resuelve introduciendo los multiplicadores de Lagrange así los datos de entrenamiento sólo aparecerán en forma de una combinación de vectores y la resolución del problema, se puede hallar resolviendo el problema dual dado por la ecuaciones en 2.2.

Según Betancourt (2005), la principal característica de la formulación dual de las MVS es que, mediante el uso de multiplicadores de Lagrange es posible representar el hiperplano deseado como combinación lineal de los propios datos y no en términos de la base del espacio vectorial en que aparecen los datos. El análisis denominado Karush Kuhn Tucker (KKT) demuestra que la gran mayoría de los coeficientes de Lagrange son cero, y que sólo pueden ser distintos de cero para los vectores de soporte, puntos que se encuentran exactamente a la distancia marcada por el margen. Al dualizar el modelo de maximización del margen se transforma en

un problema de minimización de una función cuadrática convexa sujeta a restricciones lineales.

2.5.1.2 MVS con margen blando

Este tipo particular de las MVS trata aquellos casos donde existe datos de entrada erróneos, ruido o alto solapamiento de las clases en los datos de entrenamiento, donde se puede ver afectado el hiperplano clasificador, por esta razón se cambia un poco la perspectiva y se busca el mejor hiperplano clasificador que pueda tolerar el ruido en los datos de entrenamiento. Según Benavidez (2006) esto se logra relajando las restricciones presentadas en el caso lineal, introduciendo variables de holgura no-negativas $\xi_i \geq 0$ (ver figura 2.3).

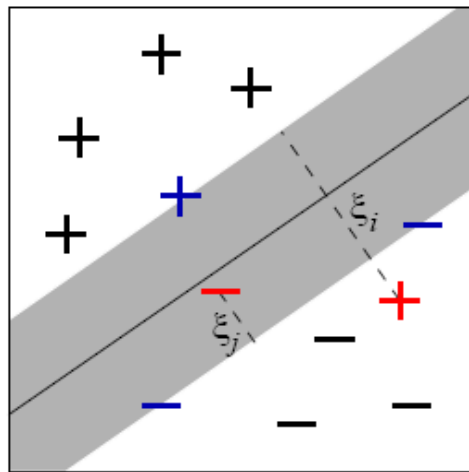


Figura 2.3. MVS con margen blando.

De manera que, matemáticamente el problema queda definido como:

$$\begin{aligned}
 & \text{Minimizar } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \\
 & \text{Sujeto a: } y_i (\langle \omega, x_i \rangle + b) \geq 1 - \xi_i \\
 & \quad \xi_i \geq 0, \quad i = 1, \dots, l
 \end{aligned} \tag{2.3}$$

En la función de optimización que se debe solucionar para el modelo de MVS con margen blando se incluye un término de regularización que depende de las variables de holgura, el cual, a su vez depende de la magnitud de las mismas y del margen. Además, este término incluye una constante C , que determina la holgura del margen blando. El valor de este parámetro C , el cual debe ser estimado a priori, depende del evaluador. La elección de un valor para este parámetro y el tipo de función *kernel* influyen en el desempeño de las MVS (Benavidez, 2006).

Siguiendo el mismo procedimiento utilizado en el caso linealmente separable, la resolución de este problema en 2.3, viene dada por la búsqueda de los multiplicadores de Lagrange, para esto se construye un Lagrangiano y se resuelve en el problema dual:

$$\begin{aligned}
 & \text{Maximizar} && \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N y_i y_j \alpha_i \alpha_j K(x_i, x_j) \\
 & \text{sujeto a:} && \sum_{i=1}^N y_i \alpha_i = 0 \\
 & && 0 \leq \alpha_i \leq C \quad 1 \leq i \leq N
 \end{aligned} \tag{2.4}$$

La función a maximizar es la misma que para el caso de margen máximo, a diferencia de la restricción $0 \leq \alpha_i \leq C$. Esto implica que los datos o patrones que cumplen la condición de tener valores $\alpha_i \geq C$ tienen el mismo comportamiento en la MVS con margen máximo, es decir, que las MVS con margen máximo se pueden obtener con $C = \infty$. Esto significa que mientras más grande es el valor de C , más estricta es la MVS al momento de permitir errores, penalizándolos con mayor rigurosidad (Betancourt, 2005).

2.5.2 Ventajas de las MVS

Las Máquinas de Vectores Soporten tienen ciertas características que las han puesto en ventaja respecto a otras técnicas populares de clasificación y/o regresión. Una de dichas características que vale la pena mencionar es que las mismas pertenecen a las disciplinas de aprendizaje automático o aprendizaje estadístico.

La diferencia más notable de las máquinas de soporte vectorial con respecto a otros algoritmos de aprendizaje, es la aplicación de un nuevo principio inductivo, que busca la minimización del riesgo estructural, además del uso de una función núcleo o *kernel*, atribuyéndole una gran capacidad de generalización, incluso cuando el conjunto de entrenamiento es pequeño. En otras palabras, se dice que tanto la capacidad de generalización como el proceso de entrenamiento de la máquina no dependen necesariamente del número de atributos, lo que permite un excelente comportamiento en problemas de alta dimensionalidad (Benavidez ,2006).

2.5.3 Posibles problemas con las MVS

Debido a que la MVS es una máquina de aprendizaje, estos algoritmos pueden presentar problemas durante el proceso de entrenamiento y validación, uno de los más comunes es lo que se conoce como "*Overtraining*" o sobre entrenamiento, el cual ocurre cuando se han aprendido muy bien los datos de entrenamiento pero no se pueden clasificar bien ejemplos nunca antes vistos (datos de verificación), es decir, una mala generalización del modelo. Otro problema que se puede presentar cuando no se ha aprendido muy bien la característica de los datos de entrenamiento, por lo que se hace una mala clasificación.

2.6 Fundamentos de los SIG

En la actualidad los Sistemas de Información Geográfica (SIG) se han convertido en herramientas muy efectivas de la ingeniería y de las ciencias básicas. Los datos espaciales despliegan información geográficamente referenciada con la finalidad de modelar el mundo real. Estudios científicos como, manejo de recursos, manejo de bienes, evaluaciones de impactos del medio ambiente, urbanismo, cartografía, entre otros, son algunas de las aplicaciones de los SIG. Por ejemplo, un SIG permitiría localizar zonas húmedas que necesitan ser protegidas de la contaminación; o a las empresas para identificar clientes potenciales similares a aquellos que ya tienen y así, proyectar sus ventas con miras.

2.6.1 Concepto de SIG

De acuerdo a NCGIA (1990), se define como “Un sistema de hardware, software y procedimientos diseñado para realizar la captura, almacenamiento, manipulación, análisis, modelización y presentación de datos referenciados espacialmente para la resolución de problemas complejos de planificación y gestión”.

2.6.2 Funciones de un SIG

2.6.2.1 Entrada de datos a un SIG

Se convierte la información geográfica, conservando todas las características iniciales de los datos espaciales, del formato analógico, el (mapas), al formato digital. Esto se puede realizar de varias maneras: digitalización, vectorización, importación, etc. Esta conversión se debe realizar (Bosque, 1992).

2.6.2.2 Análisis espacial

Según Bosque (1992), es el componente más importante en un SIG; principalmente se basa en utilizar información disponible para generar nueva información. Las siguientes funciones analíticas deben estar presentes:

- a) **Selección geográfica:** búsquedas tanto en el aspecto geográfico como en la base de datos. Se empleará en este estudio, a fin de poder hacer selecciones de los segmentos/sectores dentro del Área Metropolitana de Mérida y generar una serie de estadísticas descriptivas de cada uno de ellos que arrojen información respecto al déficit habitacional.
- b) **Proximidad:** corresponde a los elementos se encuentran próximos a otros basados en distancias.
- c) **Sobreimposición cartográfica:** combina las anteriores; además de añadir otras como es el caso de la reclasificación y las funciones por celdas, también conocido como superposición de capas de atributos.
- d) **Rutas óptimas:** selección de rutas con la menor distancia y resistencia posible. Pueden darse en medios como una red de carreteras, de drenajes, tuberías, etc.

2.6.2.3 Salidas o producción de resultados

Se refiere a las actividades que sirven para mostrar al usuario los propios datos incorporados en la propia base de datos del SIG, y los resultados de las operaciones analíticas efectuados sobre ellos. Permiten obtener mapas, gráficos, tablas numéricas y otro tipo de resultados en diferentes soportes: papel, pantallas graficas u otros (Bosque, 1992).

2.6.3 Componentes de un SIG

Principalmente se destacan el recurso humano, datos (espaciales y descriptivos), los métodos de análisis, el hardware y software. Todos ellos trabajando de manera organizada para automatizar, administrar y suministrar información a través de la representación geográfica (ESRI, 1999). La figura 2.4 ilustra tales componentes.



Figura 2.4. Componentes de un SIG.

a) Recurso humano

Son quienes se encargan del diseño y el uso del sistema. La tecnología SIG es de un valor limitado sin la gente que maneje el sistema y desarrolle aplicaciones para el mundo real. Entre los usuarios de SIG oscilan especialistas técnicos quienes diseñan y mantiene el sistema. Cuando se realiza un modelo geográfico, se construya una aplicación o cuando se escriba un manual de usuarios, es importante tener en cuenta a cuales usuarios está dirigido tal trabajo.

b) Los datos

Un SIG procesa cualquier tipo de datos que posea un componente espacial, en el caso de la EHM sería el segmento con sus viviendas correspondientes. La disposición de esta información puede ser muy diversa; fotografías aéreas, imágenes de satélites, mapas digitales, incluso base de datos con datos geográficos, como por ejemplo la EHM.

c) Métodos de Análisis

Métodos estadístico, numéricos y de consolidación de datos o funciones, procedimientos y reglas, son instrumentos de análisis para poder interpretar y procesar los datos y lo que se puede realizar con ellos.

d) Hardware y Software

En cuanto al hardware, es el equipo sobre el que un SIG opera (estaciones de trabajo, terminales, impresoras, plotters y tabletas digitalizadoras). Por otro lado el software de SIG empleado provee las funciones y las herramientas necesarias para almacenar, analizar, y mostrar información geográfica. Para este estudio se refiere al ArcGis Desktop 9.2¹. El concepto de base de datos es esencial y constituye la principal diferencia entre un SIG y un simple sistema de dibujo o de cartografía computacional. Cualquier SIG actual tiene un sistema de administración de base de datos que integra tablas de atributos, coberturas, imágenes, etc. Vincula datos espaciales con información descriptiva de alguna característica particular de un mapa.

¹ArcGis Desktop 9.2, es un software que permiten al usuario generar, importar, editar, consultar, cartografiar, analizar y publicar información geográfica.

2.6.4 Funcionalidad de un SIG

Un SIG almacena información sobre el mundo real como una colección de capas temáticas o de atributos que pueden ser vinculadas junto con la geografía. Este simple pero sumamente poderoso y versátil concepto ha sido probado para resolver muchos problemas del mundo real. El manejo de capas se puede ilustrar en la figura 2.5. Se observa un detalle de cómo un SIG podría proporcionar aspectos temático de manera independiente, a través de capas, pero referidos a un mismo espacio, Tal es el caso de topografía, redes de servicio, mapa base, etc.

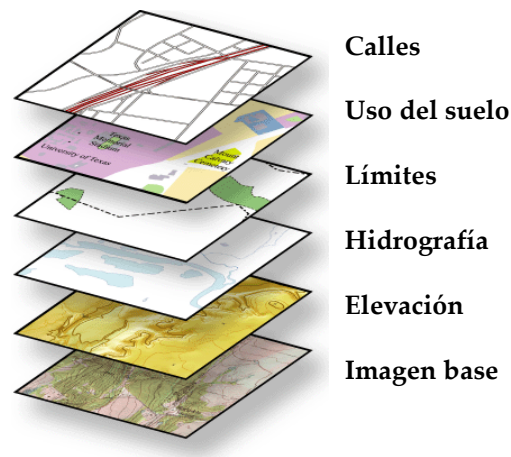


Figura 2.5. Conjunto de capas en un SIG.

2.6.4.1 Las Referencias Geográficas

Existe una explícita, tal como una latitud y longitud o las celdas cardinales y otra implícita, tal como una dirección, código postal o nombre de extensión de censo. Un proceso automatizado llamado *geocoding*, se usa para crear explicaciones de referencia geográfica (ubicaciones múltiples) desde referencias implícitas (descripciones tales como las direcciones). Estas referencias geográficas o georeferencias permiten ubicar aspectos, tales como la ubicación de un negocio o un

bosque, y sucesos, tales como un sismo, sobre la superficie de la tierra para su análisis (Posada, 1999).

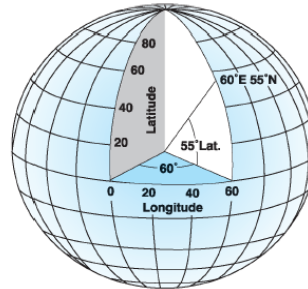


Figura 2.6. Datos representados a partir de sus coordenadas en latitud – longitud.

2.6.4.2 Modelos de datos: Vectorial y Raster

En el modelo vectorial se utilizan tres tipos de objetos espaciales para representar la geografía: puntos, líneas, y polígonos, estos son codificados y almacenados como una colección de pares de coordenadas (x, y). La ubicación de un punto, puede ser descrito por una única coordenada. Los aspectos lineales, tales como caminos y ríos, pueden almacenarse como una colección de coordenadas de punto. Los aspectos poligonales, tales como territorios, pueden almacenarse como un bucle cercano de las coordenadas. Este modelo es sumamente útil para describir aspectos discretos, pero menos útiles para describir aspectos continuos como accesibilidad o tipo de suelo y superficies territoriales. El modelo raster modela aspectos continuos. La imagen comprende una recopilación de celdas más bien como un barrido en un mapa. Ambos modelos son útiles para almacenar datos geográficos con sus propias bondades y limitaciones. Las aplicaciones de SIG modernas son capaces de manejar ambos modelos (Posada, 1999). Un aspecto de como son representados gráficamente, usando ambos modelos, se puede observar a continuación en la figura 2.6.

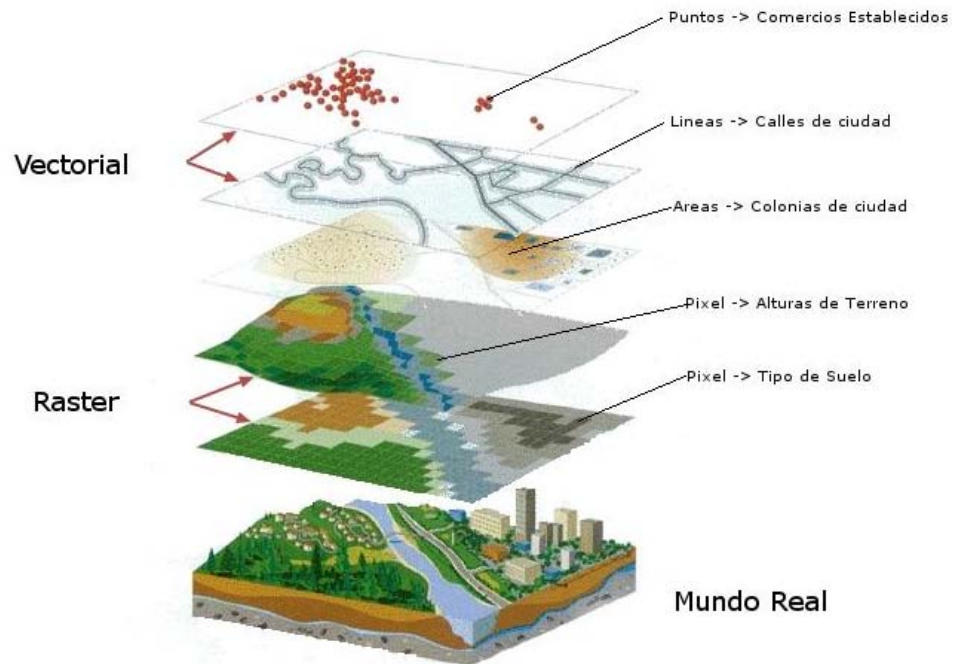


Figura 2.7. Representación gráfica de los modelos vectorial y raster.

2.6.5 Proceso de Georeferenciación

Establece una relación entre los datos desplegados en un software de SIG y su ubicación en el mundo real. Esto se logra usando un sistema de coordenadas. Para obtener resultados de análisis exactos de su base de datos SIG, se necesita entender y determinar el sistema de coordenadas. Un elipsoide datum, una proyección y sus unidades componen un sistema de coordenadas. Un datum horizontal es un marco de referencia usado para localizar elementos u objetos sobre la superficie terrestre. Se define por un elipsoide y la posición de ese elipsoide relativa a la tierra. Hay dos tipos de datum; con centro en la tierra y local. El centrado en la tierra tiene su origen colocado en el centro de la masa de la tierra actualmente conocido y es más exacto. El local se alinea de tal forma que corresponda lo más cercanamente posible a la superficie de la tierra, para un área particular y puede ser más preciso (Parra, 2007).

Capítulo 3

Procesamiento y exploración de datos

En este capítulo se describe el procedimiento de preparación de los registros de datos procedentes de la EHM, así como también el análisis exploratorio con el fin de obtener una serie de indicadores de déficit habitacional. Para el análisis exploratorio se utilizarán algunas de las técnicas de análisis multivariante tanto para la construcción de los indicadores como para la variable de salida.

3.1 Origen y disposición de los datos

Para este estudio particular se dispone de un conjunto de datos que provienen de información recolectada en la EHM en el período comprendido entre los años 2001 y 2007. Estos datos originalmente fueron suministrados por el INE almacenados en una base de datos con información de la encuesta a nivel nacional.

Es importante destacar que la base de datos de la EHM es bastante compleja. Produce tres bases de datos relacionales: de viviendas, de hogares y de personas. En este estudio se hará uso principalmente de la información relacionada a las viviendas y hogares; no obstante, la información respecto a las personas no será omitida. El número de campos que contiene la encuesta es considerable; sin embargo, se seleccionarán aquellos que muestren estrecha relación con el fenómeno

habitacional para realizar el reconocimiento de patrones y la construcción de nuevas variables indicadoras del déficit habitacional.

3.2 Conformación del subconjunto de datos de la EHM

3.2.1 Registro final de datos

Con el objeto de facilitar el manejo de la información de la encuesta, es necesaria la construcción de un registro único que contenga todos los datos que se requieran para el desarrollo de cada uno de los análisis que pretende este estudio. El propósito es compilar las variables o campos requeridos de los hogares y las viviendas para la fusión del registro resultante y la posterior creación de este registro final. La información de la encuesta se encuentra separada en semestres, y se requiere reunir todos estos bloques en uno sólo. La extracción desde la EHM del registro para el Estado Mérida se logra a través del código de la entidad federal (12 y 14 para este caso). Posteriormente, al ordenar los datos semestralmente para cada registro de acuerdo a ciertas variables (Entidad, Control, Área, Línea, Semestre), se fusionan los registros viviendas y hogares de manera semestral, y así, finalmente, compilar todos los registros de los bloques semestrales en un solo archivo, donde se destacan las variables (campos del registro) seleccionadas y como observaciones (registros), parte de las encuestas realizadas a la unidad de investigación durante cada semestre.

En vista que el espacio geográfico a considerar en esta investigación está delimitado por el Área Metropolitana de Mérida (AMM), se procede a seleccionar aquellas viviendas encuestadas en los cuatro municipios (Libertador, Campo Elías, Sucre, Santos Marquina) que conforman el AMM. La selección se hace en base a

información referencial aportada por el INE, donde se encuentra especificada la muestra de la encuesta de hogares cuyo levantamiento se realiza dentro del AMM.

3.2.2 Registro de observaciones

Este registro se crea a partir del registro final de datos y contiene la información para cada vivienda. En aquellos casos donde la vivienda posee hogares múltiples, se considera sólo uno de ellos, como se muestra en la figura 3.1 con los códigos resaltados. De este modo se elimina la redundancia en información respecto a una misma vivienda. La selección de las observaciones que conforman y se excluyen de este registro se hace a partir de una variable referencial llamada código único (Cod.Unico) la cual será definida más adelante.

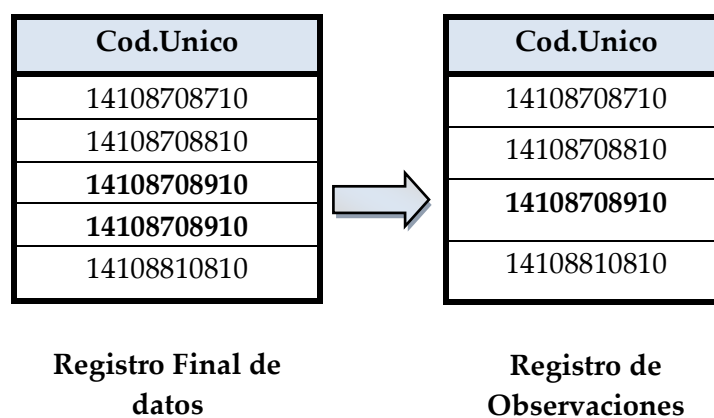


Figura 3.1. Obtención del registro de observaciones.

3.2.3 Fusión de los registros de datos

El procedimiento de unión de los registros vivienda y hogares se realizará mediante un tratamiento de datos que permite la fusión de ambos archivos; así como también

la agregación de nuevos casos. Esto resulta de gran utilidad, por el hecho de que permite la unión y relación de dos o más registros sin que ocurra pérdida de datos.² Esto se realiza con la finalidad de crear un registro único y compacto de datos para el posterior análisis estadístico.

3.3 Pre procesamiento de datos

Consiste en el uso de técnicas de análisis multivariante de datos para la obtención de variables con indicadores relacionados al déficit de viviendas, que posteriormente serán utilizadas como patrones de entrada para el proceso de clasificación.

En este apartado es importante mencionar que primero se realizó la reconstrucción de una serie de variables. Algunas de ellas porque se encontraban incompletas en la encuesta y otras porque era necesaria su construcción para facilitar tanto el manejo de los datos como su interpretación. Estas variables se denominaron complementarias. De igual manera para estudiar el déficit habitacional propiamente dicho, adicionalmente, se crearon variables que permitieran realizar la medición del déficit en distintos aspectos, a estas variables serán llamadas indicadores de déficit.

3.3.1 Variables Complementarias

Para el manejo ordenado y eficiente de los registros de datos se crea una nueva variable denominada código único (Cod.Unico) que permitirá identificar cada una de las observaciones. Se crea realizando la concatenación de 5 campos referenciales

² Para la fusión de los registros se empleó SPSS 17.

en la encuesta, estos son: entidad, control, área, línea, semestre. Los primeros cuatro campos son pertenecen a la encuesta mientras que el referido a semestre, fue creado con la finalidad de identificar el semestre en el que fue realizada la encuesta y en consecuencia el bloque de registro en el archivo final.

Otra de las variables complementarias es denominada ocupantes. Tiene que ver con el total de personas que habitan en una vivienda. La importancia de crear esta variable radica en que no estaba disponible en la encuesta, y es determinante para calcular el indicador que se refiere a hacinamiento.

3.3.2 Indicadores de déficit

El déficit habitacional tal como se plantea en este estudio tiene dos variantes. Una es la presencia de déficit cuantitativo y la otra tiene que ver con el déficit cualitativo. Ambos conceptos ya han sido descritos previamente. Ambas consideraciones de déficit es lo que va a determinar la situación en que se encuentra una determinada vivienda.

Para realizar la medición del déficit en una vivienda, es necesario determinar una serie de indicadores o variables que van a facilitar el análisis; puesto que cada una de ellas representa una condición o característica deficitaria particular. Para la creación de algunas de estas variables se hace uso de las técnicas de análisis estadístico multivariante, mientras que otras resultan directamente de la información de la EHM. Se va a detallar como es el proceso de obtención y selección de cada una de estas variables para así, tener mayor información de la influencia de cada una de ellas al momento de hacer la medición y clasificación del déficit habitacional. Básicamente se resumen en cinco las variables o indicadores: el número de hogares en la vivienda, la condición jurídica de la vivienda o tenencia, el

hacinamiento dentro de la vivienda, la condición física en que se encuentra la vivienda y la adecuación en cuanto a servicios de redes públicas. Debido a que la medición del déficit habitacional se basa en dos criterios o conceptos, déficit cuantitativo y cualitativo, se realiza una clasificación de estos indicadores en función a tales criterios.

3.3.2.1 Indicador de déficit cuantitativo

Número de Hogares

Esta variable indica el número de núcleos familiares que residen en una misma vivienda, con esto, se pretende medir la cantidad viviendas con hogares múltiples. Para la obtención de esta variable se realizó la fusión de los archivos viviendas y hogares de la encuesta de hogares, para luego contabilizar cuantos hogares por vivienda se encontraban. La relación existente entre esta variable y el número de viviendas es lo que determina el déficit cuantitativo. Los valores que toma esta variable van de 1 a 5, siendo 5 el valor número máximo de hogares por vivienda que maneja la encuesta.

3.3.2.2 Indicadores de déficit cualitativo

Tenencia

Indica el estado de propiedad o condición jurídica de la vivienda por parte del hogar o los hogares que habitan en la misma. La inclusión de este indicador básicamente radica en el hecho de poder estimar la cantidad de viviendas que están siendo potencialmente requeridas por familias que, aunque viven en buenas viviendas, no son de su propiedad. Se toma como referencia una variable de la encuesta de hogares, donde la tenencia originalmente toma un rango de valores nominales. En

la tabla 1, se muestra cada una de las categorías de la tenencia y sus respectivas frecuencias. Se redefinió la tenencia como una nueva variable de dos categorías. Así, se tiene que la tenencia de una vivienda queda categorizada como propia en función a las categorías 1 y 3 originales y se le asigna el valor (1), mientras que la segunda nueva categoría indica que una vivienda no es propia en función al resto de las categorías originales (2, 4, 5, 6, 7, 8) y se le asigna el valor (2).

| Categorías | Frecuencia | Porcentaje |
|------------------------------------|------------|------------|
| Propia pagada totalmente (1) | 7219 | 70% |
| Alquilada la vivienda (2) | 1605 | 15.6% |
| Propia pagándose (3) | 573 | 5.6% |
| Cedida por familiar (4) | 499 | 4.8% |
| Cedida por trabajo (5) | 151 | 1.5% |
| Alquilada parte de la vivienda (6) | 136 | 1.3% |
| Otra Forma (7) | 119 | 1.2% |
| Tomada (8) | 12 | 0.1% |
| Total | 10314 | 100% |

Tabla 1. Categorías de la variable tenencia y sus frecuencias.

Hacinamiento

Mide la relación existente entre el número de personas u ocupantes del hogar y el número de espacios usados como dormitorios en una vivienda, excluidos la cocina y el baño. Se calcula como el cociente entre ambas variables.

$$\text{Hacinamiento} = \text{Ocupantes} / \text{Nro_Dormitorios}$$

En este estudio se considera que viven en condiciones de hacinamiento los hogares donde hay más de dos personas por cuarto Szalachman (1999). Por tanto, se crean dos categorías, la primera de ellas designa el valor (2) para

aquellas viviendas en las cuales el número de ocupantes por cuarto sea mayor a 2, categorizándolas como hacinadas y por otro lado, las que están fuera de esta condición, asignándole el valor (1), no hacinadas.

Condición de la vivienda

Es importante destacar que el déficit habitacional no solamente se mide por la carencia absoluta de viviendas como ya se ha hecho mención, más allá de esto es necesario considerar si la vivienda a nivel de infraestructura cumple con las condiciones mínimas con las que pueda considerarse como habitable por un conjunto de individuos. Suele ocurrir que una importante cantidad de viviendas se construyen con materiales precarios o de una calidad poco deseable, lo que resulta en baja calidad de vida para los que se albergan en ella. Por tal razón, para la medición del déficit habitacional es necesario considerar, adicionalmente, la calidad de los materiales con que fueron construidos los elementos constitutivos de una vivienda (paredes, techo, piso), con el fin de determinar en base a estos materiales y al tipo de vivienda, las condiciones para morar sus ocupantes.

Los datos que reúnen la información acerca de las características de la vivienda son tomados directamente de la EHM, en la tabla 2 se listan las variables y sus posibles valores.

| Variables | Valores |
|-------------------------|---|
| Tipo de vivienda | Quinta[1], Casa[2], Apartamento en Edificio[3], Apartamento en casa o Quinta [4], Casa Vecindad [5], Vivienda Rústica[6], Rancho Campesino[7], Otro Tipo[8], Colectividad[9]. |
| Material de las Paredes | Bloque o ladrillo frisad (Acabado) [1], Bloque o ladrillo sin frisar (No acabado) [2], Madera Aserrada [3], Adobe-Tapia-Bahareque Frisado [4], Adobe-Tapia-Bahareque sin frisar [5] Otros (Caña, palma, tablas, etc.) [6] |
| Material del Techo | Platabanda[1], Teja[2], Láminas asfálticas[3], Fibrocemento, cemento ligero y similares[4], Laminas metálicas(Zinc y similares) [5], Asbesto y similares [6], Otros(Palmas, tablas y similares) [7] |
| Material del Piso | Mosaico, granito, vinil, cerámica, terracota, parquet, alfombra y similares [1], Cemento[2], Tierra[3], Otros[4] |

Tabla 2. Variables para construir la indicador condición de la vivienda.

El método que se utilizó fue el análisis de conglomerados para generar patrones que describan la condición de la vivienda. Se determinó el grado de asociación de las variables tipo de vivienda, material de las paredes, material del techo y material del piso.

Para la conformación de los conglomerados se aplicó el análisis de acuerdo al algoritmo de conglomerados en 2 fases. Se probó iterativamente el número de grupos hasta llegar al resultado que mejor describiera el comportamiento de los datos. De este modo, se obtuvieron 4 conglomerados, como se indica en las tablas 3 y descritos en detalle por atributos para cada conglomerado, en la tabla 4.

| Conglomerado | Nº de Observaciones | % de combinados | % del total |
|--------------|---------------------|-----------------|-------------|
| 1 | 2480 | 24.0% | 24.0% |
| 2 | 3248 | 31.5% | 31.5% |
| 3 | 2736 | 26.5% | 26.5% |
| 4 | 1850 | 17.9% | 17.9% |
| Combinados | 10314 | 100.0% | 100.0% |
| Total | 10314 | | 100.0% |

Tabla 3. Número de observaciones en cada conglomerado.

| Grupos | Tipo Vivienda | Tipo paredes | Tipo techo | Tipo piso |
|---------|---------------------------------------|--|--|--|
| Grupo 1 | Apartamento | Bloque o ladrillo frisado | Platabanda | Mosaico, granito, vinil, cerámica, terracota, parquet, alfombra y similares. |
| Grupo 2 | Casa Quinta Apartamento en casa | Bloque o ladrillo frisado | Platabanda Teja Laminas asfálticas Zinc | Mosaico, granito, vinil, cerámica, terracota, parquet, alfombra y similares. |
| Grupo 3 | Casa Rancho campesino | Bloque o ladrillo frisado Adobe tapia frisado Bloque o ladrillo sin frisar | Zinc Teja Asbesto Laminas asfálticas | Cemento Tierra |
| Grupo 4 | Casa Apartamento | Bloque o ladrillo frisado | Laminas asfálticas Platabanda | Cemento |

Tabla 4. Asociación de las categorías en sus respectivos conglomerados.

Como se puede observar cada grupo o conglomerado contiene categorías correspondientes a las variables originales que fueron reasignadas de acuerdo a los distintos grupos. La cantidad de grupos conformados logra una reducción

significativa de las variables originales logrando explicar lo más cercano posible la totalidad de observaciones. Así, observando la tabla anterior se puede constatar que las observaciones correspondientes al grupo uno, describen a una vivienda que se puede catalogar excelente en cuanto a su estructura física. El segundo grupo reúne aquellas viviendas que son consideradas en buenas condiciones, mientras que las observaciones correspondientes al grupo tres y cuatro son categorizadas como viviendas con condición mala y regular respectivamente, en el anexo B se muestran las gráficas de los porcentajes de cada variable dentro de cada conglomerado. Finalmente, se establece que las categorías de la nueva variable condición de la vivienda y sus valores quedan definidos como muestra la tabla 5.

| Conglomerado | Categoría | Valor |
|--------------|---------------------|-------|
| Grupo 1 | Condición Excelente | 1 |
| Grupo 2 | Condición Buena | 2 |
| Grupo 3 | Condición Mala | 3 |
| Grupo 4 | Condición Regular | 4 |

Tabla 5. Categorías para el indicador condición de la vivienda.

Adecuación de la vivienda

Otro aspecto fundamental relacionado con la vivienda, tiene que ver con el acceso a los servicios básicos: agua, electricidad, recolección de basura, eliminación de excretas. La carencia de alguno de ellos representa, sin duda, una situación de incomodidad que afecta el bienestar de los ocupantes del recinto habitacional.

La adecuación se determina en función de las variables antes mencionadas y los valores que las mismas pueden tomar. Se muestran en la siguiente tabla.

| Variable | Valores |
|-------------------------------------|---|
| Servicio de Agua | Acueducto (1) Pila pública (2) Camión (3) Otros medios (4) |
| Servicio de Eliminación de Excretas | Poceta a cloaca (1) Poceta a pozo séptico (2) Excusado a hoyo o letrina (3) No tiene poceta o Excusado (4) |
| Servicio Eléctrico | Si (1), No(2) |
| Recolección directa de Basura | Si (1), No(2) |
| Contenedor de basura | Si (1), No(2) |

Tabla 6. Variables para construir el indicador adecuación de la vivienda.

El procedimiento usado para construir el indicador de adecuación de la vivienda es similar al empleado para el de condición de la vivienda. Se tiene un conjunto de variables y se observa las relaciones entre ellas. De igual manera, se aplica el método de análisis de conglomerados, pero en este caso, para medir la asociación de las variables en la EHM que tienen que ver con el acceso a los servicios básicos. El análisis de conglomerados aplicado a las variables determinó 3 grupos, tal y como se aprecian en las tablas 7 y 8.

| Conglomerado | Nº de Observaciones | % de combinados | % del total |
|--------------|---------------------|-----------------|-------------|
| 1 | 8091 | 78.4% | 78.4% |
| 2 | 876 | 8.5% | 8.5% |
| 3 | 1347 | 13.1% | 13.1% |
| Combinados | 10314 | 100.0% | 100.0% |
| Total | 10314 | | 100.0% |

Tabla 7. Número de observaciones en cada conglomerado.

La mayor proporción de observaciones resultó en el conglomerado 1, donde se destacan todos los servicios, excepto la eliminación de basura. En la conformación de cada conglomerado se observó la asociación de las categorías de cada una de las variables relacionadas al acceso a los servicios básicos.

| Grupos | Suministro de Agua | Eliminación de Excretas | Servicio Eléctrico | Recolección directa de basura | Contenedor de basura |
|---------|---|--|--------------------|-------------------------------|----------------------|
| Grupo 1 | Acueducto | Poceta | Si | Si | No |
| Grupo 2 | Acueducto Otros medios Pila pública | Pozo séptico, No posee poceta Hoyo o letrina | Si | No | No |
| Grupo 3 | Acueducto Pila pública | Poceta | Si | No | No |

Tabla 8. Asociación de las categorías en sus respectivos conglomerados.

Con el propósito de determinar la nueva variable adecuación de la vivienda, y los patrones que se definen, se tiene que el primer conglomerado caracteriza una vivienda como adecuada asignándole el valor (1) ya que muestra deficiencia en sólo uno de los servicios, asociado a no poseer contenedor de basura. Las observaciones que corresponden al conglomerado 2 presentan deficiencia en todos los servicios por lo que se categorizan como inadecuadas y se les asigna el valor (2). Por otra parte el tercer grupo o conglomerado contiene observaciones con deficiencia en dos de los servicios, considerándolas regularmente adecuadas en términos de servicios básicos y se les asigna el valor (3). De igual manera que en el caso anterior en el anexo B se muestran las gráficas de los porcentajes de cada variable dentro de cada conglomerado.

3.4 Definición de la variable de salida

La creación de esta nueva variable dependiente o de salida se debió a la ausencia de medidas del déficit habitacional en la EHM. Para cada observación se obtuvo un valor cualitativo que caracterizaba su condición de déficit. De este modo, se construyó una variable que permitiera categorizar la situación deficitaria de cada una de las unidades de investigación, en función de los indicadores estudiados anteriormente.

Para construir dicha variable se hace uso de Análisis de Correspondencias Múltiple (ACM). Con este análisis, mediante mapas perceptuales, se observó que la reagrupación de las variables originales (número de hogares, tenencia, hacinamiento, condición y adecuación), en lo que respecta a sus categorías o atributos, determinó la nueva variable de carácter cualitativo. Esta variable fue el producto de la combinación del conjunto global de los atributos originales contenidos en cada variable original, en nuevas relaciones altamente diferenciables. El resultado fue la identificación del déficit habitacional, y en consecuencia, la construcción de la variable de salida. En conclusión esta nueva variable se creó con una nueva categorización que identificó la calificación del déficit habitacional. El resultado se ilustra en la figura 3.2.

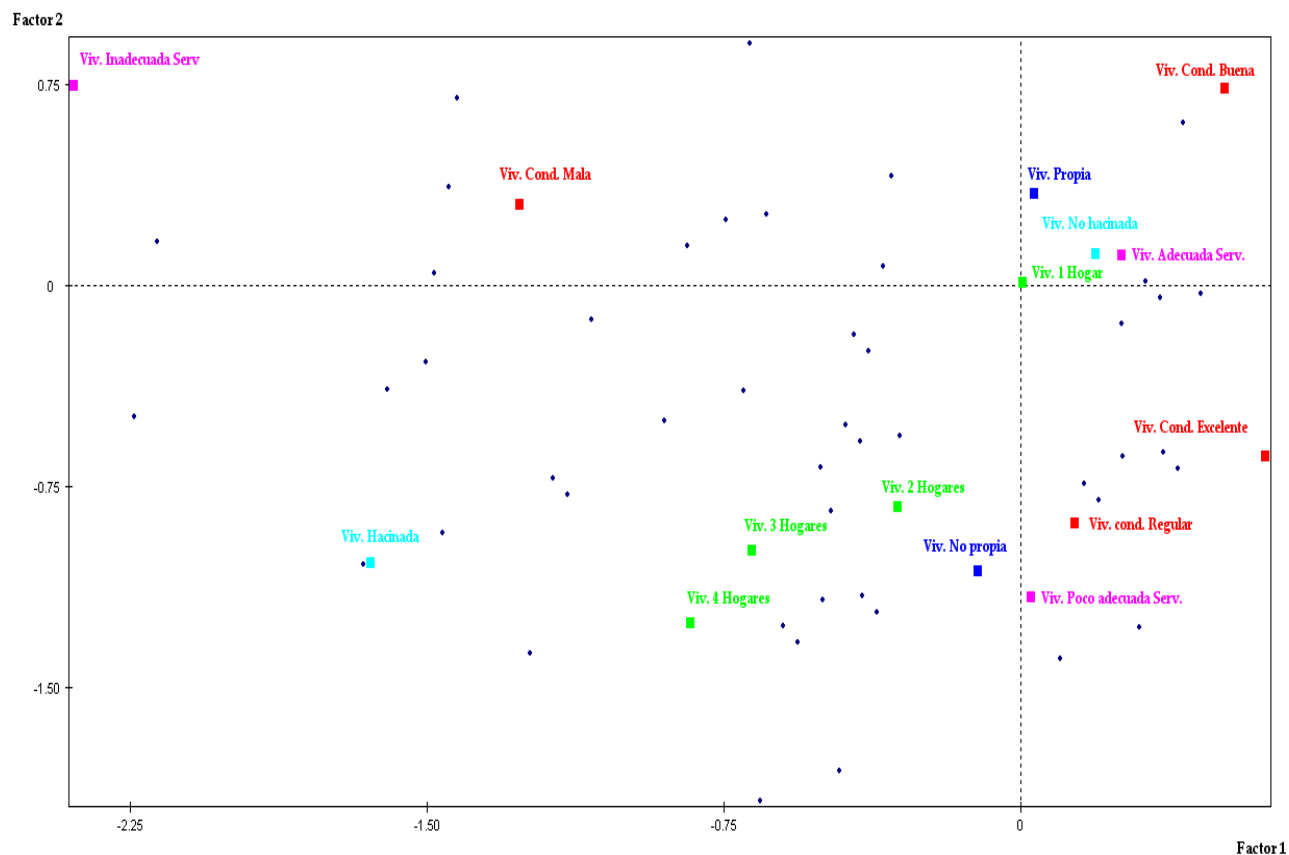


Figura 3.2. Resultado del Análisis de Correspondencias Múltiple.

Una de las ventajas del ACM es poder observar a través de gráficas, las relaciones que se establecen entre las distintas variables. En el gráfico se puede observar claramente estas relaciones. Para obtener esta variable se toma en consideración los conceptos de déficit habitacional que ya se conocen, como son el déficit cuantitativo y el cualitativo. En el lado derecho del eje de coordenadas se agrupan aquellas categorías que describen a una vivienda que alberga sólo un hogar, no presenta hacinamiento, la forma de tenencia es propia, la condición de la vivienda respecto a su estructura física varía entre regular y excelente y la vivienda es adecuada o a lo sumo poco adecuada respecto a la disponibilidad de servicios públicos, por lo que una vivienda con estas características se categoriza como no

deficitaria y se le asigna el valor (1). De la misma manera se observa en el lado izquierdo aquella vivienda que alberga dos o más hogares, presenta hacinamiento, no es propia, su condición física es mala y en cuanto a disponibilidad de servicios es inadecuada. A este tipo de vivienda se categoriza como deficitaria, y se le asigna el valor (2).

Adicionalmente, las relaciones entre las variables también se pueden establecer en uno u otro factor (Factor 1 ó Factor 2), con el fin de precisar el comportamiento de la variable de salida. Así, se puede decir que en la dimensión 1 o factor 1 (eje de las abscisas) hay mejor discriminación de las variables hacinamiento, condición de la vivienda y adecuación de la vivienda, puesto que ubica las características de viviendas deficitarias (Viviendas hacinadas, mala condición e inadecuadas) al lado izquierdo del eje y las no deficitarias (Viviendas no hacinadas, condición buena y excelente y adecuada) al lado derecho. De igual modo, se puede observar que las variables número de hogares y tenencia son mejor representadas por el factor 2 (eje de las ordenadas), el cual ubica las características de las viviendas deficitarias (más de dos hogares, no propia) por debajo del eje y las no deficitarias (1 hogar, propia) por encima del eje. En la tabla 9 se presenta un resumen de la nueva variable de salida.

| Categoría | Descripción | Valor |
|----------------|---|-------|
| No deficitaria | Vivienda donde no existe condición de déficit cuantitativo o cualitativo. | 1 |
| Deficitaria | Vivienda donde la condición de déficit cuantitativo o cualitativo se cumple en alguno de los casos. | 2 |

Tabla 9. Categorías de la Variable de salida.

Capítulo 4

Experimentación y análisis de los resultados

En este capítulo se pretende mostrar cómo se lleva a cabo la clasificación del déficit habitacional en el Área Metropolitana de Mérida, a partir de las variables o indicadores generados en el capítulo anterior. Para tal fin, se hará uso del algoritmo de aprendizaje que ofrecen las máquinas de vectores de soporte para la clasificación de datos, y así construir un modelo de pronóstico de clasificación del déficit habitacional.

4.1 Proceso de clasificación del déficit habitacional

Para lograr la clasificación del déficit de viviendas se utilizó el método estadístico de las máquinas de vectores de soporte, cuyo algoritmo de aprendizaje automático ha demostrado ser bastante eficaz para la clasificación de datos. La idea que hay detrás de este tipo de aprendizaje es de ir aprendiendo, a través de ejemplos, las salidas correctas para ciertas entradas. Los ejemplos que se utilizan para “enseñar” a la máquina se conocen como conjunto de entrenamiento y aquellos que se utilizan para validar se les denominan conjunto de prueba. La metodología de aprendizaje permite que las MVS puedan resolver problemas sin necesidad de conocer la distribución o naturaleza de los datos con los que se desea trabajar.

Para obtener la clasificación se llevará a cabo una secuencia de pasos, y de esta manera mostrar detalladamente el desarrollo del mismo. El esquema presentado en la figura 4.1 ilustra este proceso.

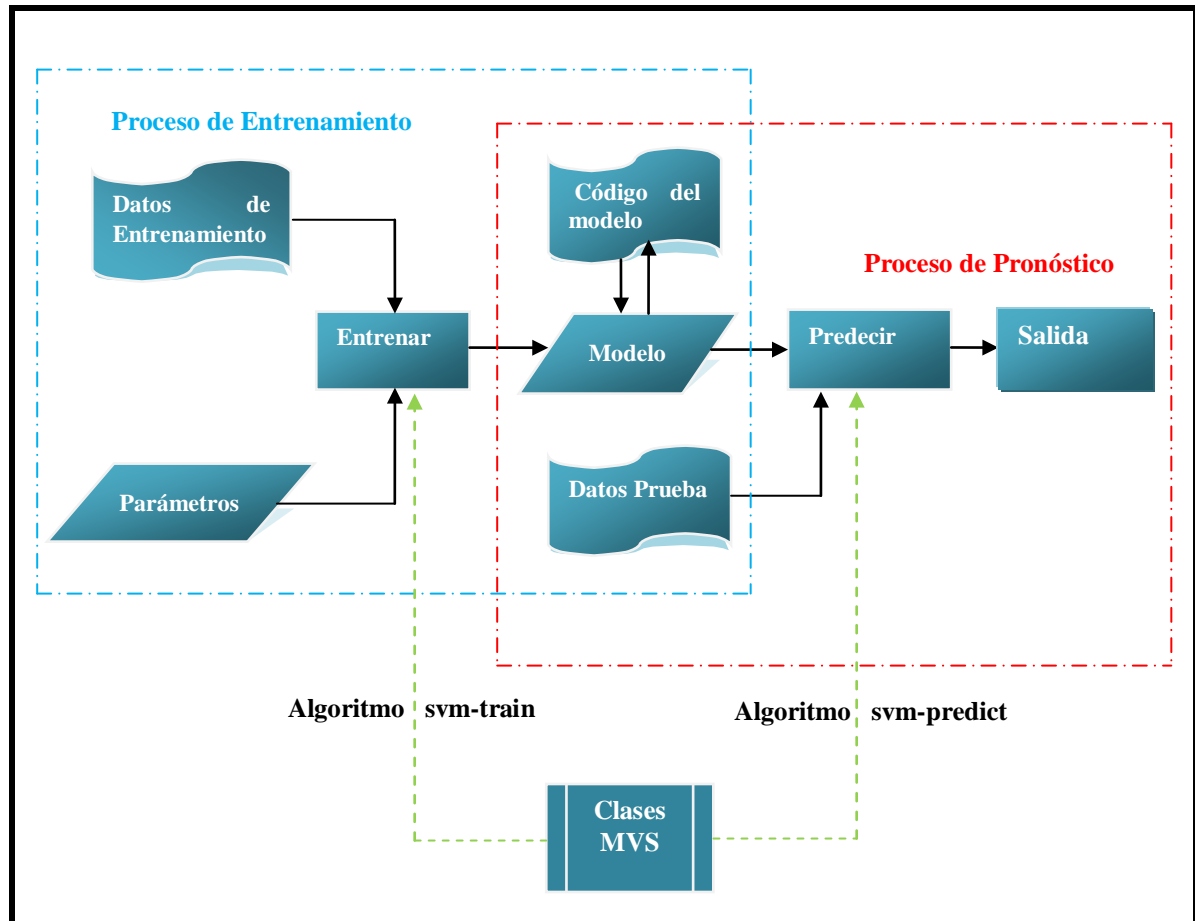


Figura 4.1 Esquema del proceso de clasificación.

4.1.1 Datos del modelo de clasificación

4.1.1.1 Atributos o categorías en las variables de entrada

Los atributos que el modelo de clasificación recibe, provienen del resultado del análisis exploratorio que se realizó a los datos que originalmente provenían de la EHM, tal como fueron descritos en los puntos 3.3.2.1 y 3.3.2.2 del capítulo anterior.

Estas entradas son los indicadores de déficit transformados en atributos construidos a partir de las técnicas de análisis multivariante. En la tabla 10 se muestra cada una de estas variables con sus correspondientes atributos.

| Variables | Nro_Hogares | Tenencia | Hacinamiento | Condición | Adecuación |
|------------|-------------|----------|--------------|-----------|------------|
| Categorías | 1,2,3,4 | 1,2 | 1,2 | 1,2,3,4 | 1,2,3 |

Tabla 10. Atributos del modelo de clasificación.

4.1.1.2 Clases de la variable de salida

La variable de salida clasifica el déficit habitacional en función a dos clases; la primera de ellas indica que una vivienda no es deficitaria tal como se describió en el capítulo anterior usando ACM, y por otro lado, en consecuencia, la segunda indica la condición deficitaria de la vivienda. Debido a que el número de clases es igual a dos, la clasificación es de tipo binaria.

4.1.2 Transformación de los datos al formato de la aplicación utilizada

Primeramente se debe especificar el software utilizado para realizar el procedimiento, y posteriormente definir el formato que se requiere para su funcionamiento. La herramienta computacional usada para resolver el problema de clasificación fue la LIBSVM versión 2.88³. Es muy recomendado ya que posee buena documentación y varios scripts para usuarios no expertos, además, ofrece herramientas que ayudan a la selección adecuada de los parámetros, aspecto que le agrega mucha utilidad y aplicabilidad.

³ Un paquete de libre distribución bajo LINUX, robusto, de fácil instalación y uso.

Este software requiere que los datos de entrada se encuentren en un formato especial, tal como se muestra en la tabla 11.

| Variable de salida | 1ª V.E | 2ª V.E | ... |
|--------------------|--------|--------|-----|
| 1 ó 2 | X:Y | X:Y | ... |

Tabla 11. Formato de los datos de entradas.

Los valores 1 y 2 son los atributos correspondientes a la variable de salida y para las variables de entrada (V.E), la disposición de los datos para cada observación sería el orden de la variable de entrada y la Y corresponde al valor atribuible a cada variable de entrada para esa observación, después del separador.

Un ejemplo de cómo sería para el caso de cinco variables de entrada y una de salida, se describe tal como sigue.

```

1  1:1  2:1  3:1  4:1  5:1
2  1:2  2:2  3:1  4:3  5:2
2  1:2  2:1  3:2  4:4  5:3
1  1:2  2:1  3:2  4:2  5:2

```

Figura 4.2. Ejemplo del formato correcto de los datos de entrada.

Las clases 1 y 2 corresponden a las respectivas salidas de la clasificación en la primera columna. Las columnas sucesivas muestran, para cuatro observaciones, las cinco variables de entrada con su correspondiente posición y su valor separados por dos puntos.

4.1.3 Selección de los datos de entrenamiento y prueba

La selección de los ejemplos se realizó de acuerdo a la disposición semestral de la EHM. Cada semestre constituye un estrato de observaciones y cada estrato se divide en 60% para entrenamiento y 40 % para generalización. De esta manera se tiene 13 estratos como se muestra en la tabla 12. La tabla contiene el número de observaciones en cada semestre o estrato y sus respectivos porcentajes para entrenamiento y prueba. Las muestras en cada estrato se extraen de manera aleatoria y fueron generadas por la LIBSVM a través de un script que posee para este propósito. La idea detrás de la estratificación en semestres viene dada por el grado de representatividad que se desea obtener de cada segmento al momento de hacer la representación espacial; es decir, que cada instancia dentro de cada segmento se encuentre clasificada correctamente y en consecuencia, esté representada en la muestra.

| Semestre | Nº Observaciones | Entrenamiento (60%) | Prueba (40%) |
|--------------|------------------|---------------------|--------------|
| 1 | 620 | 372 | 248 |
| 2 | 856 | 514 | 342 |
| 3 | 823 | 494 | 329 |
| 4 | 630 | 378 | 252 |
| 5 | 768 | 461 | 307 |
| 6 | 849 | 509 | 340 |
| 7 | 819 | 491 | 328 |
| 8 | 803 | 482 | 321 |
| 9 | 807 | 484 | 323 |
| 10 | 804 | 482 | 322 |
| 11 | 859 | 515 | 344 |
| 12 | 861 | 517 | 344 |
| 13 | 815 | 489 | 326 |
| Total | 10314 | 6188 | 4126 |

Tabla 12. Cantidad de observaciones para entrenamiento y prueba.

4.1.4 Selección de la función Kernel

Para llevar a cabo el proceso de entrenamiento de la MVS, es necesario definir el tipo de función *kernel* que se utilizará para entrenar las muestras. La selección del *kernel* es un paso fundamental en la aplicación de esta técnica, puesto que se busca separar los ejemplos en clases y , cuando no se logra de manera lineal en el espacio de entrada, hay que separarlo en un espacio de mayor dimensión a través de una de estas funciones.

Existen diversos tipos de funciones *kernels* que pueden ser usadas para entrenar los datos, la tabla 13 muestra los tipos más comunes.

| Kernel | Función | Parámetros a ajustar |
|------------|---------------------------------------|------------------------|
| Lineal | $k(x, y) = x * y$ | Ninguno |
| Polinomial | $k(x, y) = (x * y + 1)^d$ | d: Grado del polinomio |
| RBF | $K(x, y) = e^{-\gamma * \ x - y\ ^2}$ | γ : Gamma |

Tabla 13. Tipos de *kernels* más comunes.

El tipo de kernel que se decidió utilizar en este caso fue el RBF (Radial Basis Function) o *kernel* de función de base radial. La razón se debe fundamentalmente a que diversos autores, entre los cuales se destaca Moro y Hurtado (2006) y Hsu et al. (2008), lo sugieren como una elección razonable para clasificación con MVS debido a que ha demostrado buen desempeño en estudios previos. Además, dichos autores afirman que este *kernel* hace corresponder de manera no lineal ejemplos en un espacio de mayor dimensión. El *kernel* RBF tiene menos hiperparámetros (C y γ) a diferencia de otros, tal como el polinomial.

4.1.5 Ajuste de los parámetros del modelo

El ajuste de los parámetros del modelo es trascendental para realizar la clasificación, ya que de la elección de los mismos depende en gran manera su desempeño. El objetivo es identificar la mejor combinación de los parámetros de manera que el clasificador pueda predecir lo mejor posible ante datos desconocidos (datos de prueba), con un menor error de clasificación.

Para el *kernel* de función base radial el parámetro γ , es una constante de proporcionalidad cuyo rango de valores útiles debe ser estimado para cada aplicación particular, se le conoce como parámetro del nivel de no linealidad (Bautista, 2004). Es el que controla la amplitud de la función de base radial, como este parámetro controla el ancho de la función kernel, valores que tienden a cero estrechan la amplitud de dicha función. El parámetro C , se refiere al parámetro de penalidad para el error. Este parámetro C , es el que va a permitir la holgura para tener cierto error de clasificación a cambio de tener una mejor generalización de los datos. El valor de C varía de acuerdo a los datos que se desean clasificar; por lo tanto, se recomienda que se haga una búsqueda exhaustiva entre la C hasta alcanzar un valor adecuado que permita clasificar correctamente la mayor cantidad de instancias desconocidas como sea posible. Es importante tener en cuenta que con un valor muy elevado de C se tiene una alta penalización para puntos no separables y se tiende a tener muchos vectores soporte, lo que puede llevar al sobreajuste y, con esto, a la mala generalización. Por otra parte, un valor muy pequeño de C hace que el modelo sea muy rígido, lo que puede conducir a un subajuste.

En esta investigación tanto el parámetro de la función *kernel* (γ) como el parámetro de penalización C se determinan por el método de búsqueda en malla o

rejilla usando validación cruzada, ya que este método es de fácil implementación y evita el sobre entrenamiento.

4.1.6 Búsqueda de rejilla y validación cruzada

Los parámetros (C y γ) no son conocidos de antemano y el proceso de obtención de los mismos puede tornarse difícil, puesto que se tendrían que probar una gran cantidad de configuraciones de dichos parámetros hasta encontrar aquella que produzca el mejor resultado; además, el experimentador tendría que contar con mucha experiencia al momento de ajustar tales parámetros (Hsu et al, 2008). Por tal razón, se hace uso de un método de selección de parámetros como es el caso de la búsqueda en malla o rejilla usando validación cruzada, el cual consiste en probar combinaciones de C y γ , donde el primer parámetro se mantiene constante mientras se varía el segundo, luego el proceso se repite para el nuevo valor del primer parámetro y así en lo sucesivo. Se selecciona aquella combinación con la mejor tasa de validación cruzada.

Finalmente, se encuentran los valores más apropiados de ambos parámetros de tal manera que el modelo de la MVS se comporte lo mejor posible. En el caso del *kernel* RBF se ha demostrado que probando con secuencias exponencialmente crecientes de C y γ (por ejemplo $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ y $\gamma = 2^{-15}, 2^{13}, \dots, 2^3$), se obtienen buenos resultados. La idea de utilizar este método es probar inicialmente usando una rejilla gruesa, es decir un rango amplio tanto para C como para γ , y así identificar la mejor zona y posteriormente, utilizar una rejilla más fina sobre esa zona para determinar la mejor combinación C y γ . LIBSVM 2.88 ofrece un scripts que permite realizar tal búsqueda, de modo que es mucho más práctico y eficiente obtener la mejor configuración de los parámetros requeridos por el modelo.

4.1.7 Entrenamiento de la MVS

Se realizaron ciertas configuraciones previas hasta llegar a la mejor combinación de C y γ . Como ya se describió en secciones anteriores el conjunto de entrenamiento está conformado por el 60% de las observaciones de cada semestre. Inicialmente se realizó el entrenamiento con los parámetros por defecto sin previo ajuste, que permitiera ofrecer una idea general de con cuánta exactitud son clasificados los ejemplos. El *kernel* por defecto es el RBF y los resultados obtenidos en esta primera búsqueda se muestra en la tabla 14.

| Semestre | Desempeño de la Clasificación (%) |
|-----------------|-----------------------------------|
| Semestre 1 | 97.9839 |
| Semestre 2 | 98.8304 |
| Semestre 3 | 98.4802 |
| Semestre 4 | 99.6032 |
| Semestre 5 | 93.8111 |
| Semestre 6 | 99.1176 |
| Semestre 7 | 98.1707 |
| Semestre 8 | 99.0654 |
| Semestre 9 | 98.4520 |
| Semestre 10 | 97.5155 |
| Semestre 11 | 98.5465 |
| Semestre 12 | 99.4186 |
| Semestre 13 | 98.1595 |
| Promedio | 98.2427 |
| Varianza | 1.9586 |

Tabla 14. Resultados de la clasificación sin ajuste de parámetros.

Se puede observar que el porcentaje de desempeño de la clasificación es bueno a pesar de que no se ha hecho ajuste a los parámetros. El porcentaje de

desempeño más bajo se nota en el semestre 5 (93.8111%), esto ofrece una noción de cómo se encuentran los datos en este semestre, es decir, pueden presentar ruido o estar solapados. A pesar de que los porcentajes no son malos, es posible mejorar la clasificación ajustando los parámetros del modelo. Para la obtención de la mejor combinación de C y γ , se utilizó el método de búsqueda de malla o rejilla como ya se mencionó anteriormente. De esta manera, se probaron dos configuraciones de rejilla, con distintos rangos para C y γ . En la tabla 15, se puede apreciar los parámetros ajustados para cada semestre usando la búsqueda de rejilla con validación cruzada; la anchura con un rango de la rejilla equivalente a $C = [2^{-15}, 2^{15}]$ y paso = 2, $\gamma = [2^{-15}, 2^3]$ y paso = 2.

| Semestre | C | γ | Tasa de Validación Cruzada (%) | Desempeño de la Clasificación (%) |
|-----------------|-------------|----------|--------------------------------|-----------------------------------|
| Semestre 1 | 2.00 | 2.00 | 100.00 | 100.00 |
| Semestre 2 | 128.00 | 0.03 | 99.61 | 98.83 |
| Semestre 3 | 2048.00 | 0.03 | 99.60 | 99.70 |
| Semestre 4 | 32768.00 | 0.01 | 99.74 | 100.00 |
| Semestre 5 | 512.00 | 0.03 | 99.78 | 99.35 |
| Semestre 6 | 128.00 | 0.03 | 100.00 | 99.71 |
| Semestre 7 | 128.00 | 0.03 | 100.00 | 99.08 |
| Semestre 8 | 512.00 | 0.03 | 100.00 | 99.39 |
| Semestre 9 | 2.00 | 2.00 | 100.00 | 100.00 |
| Semestre 10 | 0.50 | 2.00 | 100.00 | 100.00 |
| Semestre 11 | 0.13 | 2.00 | 100.00 | 100.00 |
| Semestre 12 | 0.13 | 2.00 | 100.00 | 100.00 |
| Semestre 13 | 0.50 | 2.00 | 100.00 | 100.00 |
| Promedio | 2786.87 | 0.94 | 99.90 | 99.70 |
| Varianza | 75195656.45 | 0.97 | 0.02 | 0.1526 |

Tabla 15. Ajuste de los parámetros C y γ con Rejilla 1.

Además, se puede observar el porcentaje de la tasa de validación cruzada, así como también, el de la clasificación con el ajuste de los parámetros. Se aprecia una mejoría en la clasificación ya que el porcentaje de desempeño más bajo correspondiente al semestre 2, y es 98.8304%. Esta clasificación aunque muestra buenos resultados, no es la óptima. Existen valores de C muy elevados en algunos semestres, por lo que se está permitiendo un error de clasificación alto; por tal razón, se prosigue a realizar otra configuración con una rejilla más fina.

En esta nueva rejilla los rangos de C y γ para la esta rejilla fueron $C = [2^{-5}, 2^5]$ y paso = 1, $\gamma = [2^{-4}, 2^0]$, y paso = 1. En la tabla 16 se observan los resultados obtenidos para esta nueva configuración.

| Semestre | C | γ | Tasa de validación Cruzada (%) | Desempeño de la Clasificación (%) |
|-----------------|-------|----------|--------------------------------|-----------------------------------|
| Semestre 1 | 2.00 | 1.00 | 100.00 | 100.0000 |
| Semestre 2 | 4.00 | 0.25 | 99.61 | 98.8304 |
| Semestre 3 | 2.00 | 1.00 | 99.60 | 100.0000 |
| Semestre 4 | 32.00 | 0.125 | 99.74 | 99.6032 |
| Semestre 5 | 16.00 | 0.25 | 99.35 | 98.3713 |
| Semestre 6 | 2.00 | 1.00 | 100.00 | 100.0000 |
| Semestre 7 | 8.00 | 0.25 | 100.00 | 99.0854 |
| Semestre 8 | 1.00 | 1.00 | 100.00 | 99.6885 |
| Semestre 9 | 2.00 | 1.00 | 99.79 | 100.0000 |
| Semestre 10 | 0.25 | 1.00 | 99.79 | 100.0000 |
| Semestre 11 | 0.25 | 1.00 | 99.81 | 100.0000 |
| Semestre 12 | 0.25 | 1.00 | 99.81 | 100.0000 |
| Semestre 13 | 0.50 | 1.00 | 100.00 | 99.3865 |
| Promedio | 5.40 | 0.76 | 99.81 | 99.6127 |
| Varianza | 76.76 | 0.13 | 0.037285053 | 0.2722 |

Tabla 16. Ajuste de los parámetros C y γ con Rejilla 2.

Se puede apreciar de la tabla que nuevamente los datos son muy bien clasificados por la MVS, y en esta oportunidad, los valores del parámetro de penalización C son menos elevados, por lo que se está permitiendo un error de clasificación bajo, el valor máximo de C es 32. Además, los valores de γ no sufren mucha variabilidad entre semestres.

Esta configuración ofrece una idea más clara de cuáles son los parámetros a emplear para la generalización del modelo. Es importante observar que a pesar de sacrificar una variación no importante en la clasificación, se obtiene mejor desempeño por lograr un error de clasificación bajo, un valor de C aceptable y un γ sin mayor modificación.

Como se puede observar con la búsqueda de rejilla en pocos pasos se han obtenido los mejores parámetros del modelo para cada semestre. Lo ideal es obtener un modelo general que se aplique a todos los semestres, de esta manera se escoge la mejor combinación (C y γ) a partir de la tabla anterior, donde se considera como parámetro de regularización para el modelo de clasificación el valor $C=32$, debido a que se desea que la máquina generalice y prediga con el mayor desempeño ante nuevas instancias.

Respecto al valor de γ , este se determinó combinando el valor de C ya establecido, con los tres posibles valores de γ (0.125, 0.25, 1) en la tabla 16, de tal manera que se entrenó la máquina con cada de una de estas configuraciones, el resultado se ilustra en la tabla 17.

| Semestre | Desempeño de la Clasificación (%) (C=32, $\gamma=0.125$) | Desempeño de la Clasificación (%) (C=32, $\gamma=0.25$) | Desempeño de la Clasificación (%) (C=32, $\gamma=1$) |
|-----------------|--|---|--|
| Semestre 1 | 99.5968 | 100.0000 | 100.0000 |
| Semestre 2 | 99.8304 | 98.8304 | 99.1228 |
| Semestre 3 | 99.6960 | 99.6960 | 100.0000 |
| Semestre 4 | 99.6032 | 100.0000 | 99.6032 |
| Semestre 5 | 98.0456 | 98.3713 | 100.0000 |
| Semestre 6 | 99.7059 | 99.1176 | 100.0000 |
| Semestre 7 | 99.0854 | 99.0854 | 100.0000 |
| Semestre 8 | 100.0000 | 99.6885 | 99.6885 |
| Semestre 9 | 99.3808 | 99.3808 | 100.0000 |
| Semestre 10 | 99.6894 | 100.0000 | 100.0000 |
| Semestre 11 | 99.1279 | 99.4186 | 100.0000 |
| Semestre 12 | 100.0000 | 99.4186 | 100.0000 |
| Semestre 13 | 100.0000 | 99.3865 | 99.3865 |
| Promedio | 99.5201 | 99.4149 | 99.8308 |
| Varianza | 0.2639 | 0.2162 | 0.0791 |

Tabla 17. Mejor combinación de C y γ .

Finalmente se obtuvo la mejor configuración de los parámetros, dada por $C=32$ y $\gamma=1$, cuarta columna en la tabla anterior, con la cual se entrenan las muestras de cada uno de los semestres.

Como se puede observar las tres configuraciones ofrecen buenos porcentajes de desempeño, pero los parámetros elegidos ($C=32$, $\gamma=1$) ofrecen la menor varianza entre los semestres. De esta manera se han obtenido los parámetros y la clasificación de cada una de las instancias, de modo que ante nuevos datos se puede predecir a partir de los indicadores de déficit cual es la condición de una determinada vivienda.

4.1.8 Pruebas y Resultados

Las pruebas se realizaron con el 40% de las observaciones correspondientes a cada semestre (estos datos no fueron utilizados para entrenamiento). Se pretende medir la precisión del modelo ante datos desconocidos. En la sección anterior cada modelo fue probado con su respectiva configuración y el porcentaje de desempeño es el que indica el grado de generalización del modelo. El desempeño del modelo en cada caso, queda determinado por:

$$\% \text{ de desempeño de la Clasificación} = (\text{N}^{\circ} \text{ de instancias identificadas correctamente} / \text{total}) * 100$$

| Semestre | Nº de instancias identificadas correctamente | Total de instancias de prueba | Desempeño de la Clasificación (%) |
|-----------------|--|-------------------------------|-----------------------------------|
| Semestre 1 | 248 | 248 | 100.0000 |
| Semestre 2 | 339 | 342 | 99.1228 |
| Semestre 3 | 329 | 329 | 100.0000 |
| Semestre 4 | 251 | 252 | 99.6032 |
| Semestre 5 | 307 | 307 | 100.0000 |
| Semestre 6 | 340 | 340 | 100.0000 |
| Semestre 7 | 328 | 328 | 100.0000 |
| Semestre 8 | 320 | 321 | 99.6885 |
| Semestre 9 | 323 | 323 | 100.0000 |
| Semestre 10 | 322 | 322 | 100.0000 |
| Semestre 11 | 344 | 344 | 100.0000 |
| Semestre 12 | 344 | 344 | 100.0000 |
| Semestre 13 | 324 | 326 | 99.3865 |
| Promedio | | | 99.8308 |
| Varianza | | | 0.0791 |

Tabla 18. Medida de desempeño del modelo de clasificación.

De acuerdo a los porcentajes de desempeño obtenidos, queda demostrado el poder de generalización de la técnica utilizada, puesto que para instancias desconocidas se obtienen excelentes resultados.

Es en este punto donde se nota la aplicabilidad del modelo de clasificación utilizando las MVS. Más que un modelo de clasificación es un modelo de pronóstico, de manera que, por ejemplo, a futuro se desea clasificar nuevos datos correspondientes a un semestre 14, la máquina podrá clasificar estas nuevas instancias con un porcentaje de desempeño alrededor de 99.8308 con una variación de 0.0791.

4.2 Inferencias sobre el déficit habitacional

Una de las grandes ventajas de las MVS es que permite hacer pronóstico del déficit habitacional ante nuevas instancias de datos, como se demostró en la sección anterior. Sin embargo, esta técnica no ofrece detalles referentes a la cuantificación del déficit; es decir, de los porcentajes de viviendas que se encuentran en una u otra condición respecto al fenómeno del déficit de viviendas. Por esta razón, a partir de los resultados del ACM, se presentarán una serie de inferencias del fenómeno del déficit habitacional.

Es importante destacar que para medir el déficit habitacional se emplean dos conceptos del déficit, donde la contribución de ambos es lo que permite la medición exhaustiva del fenómeno en estudio. En este sentido, para lograr tal fin, se disponen de 10314 observaciones correspondientes a 13 períodos censales de la EHM, de manera que es preciso dejar claro que las observaciones, a pesar de que

semestralmente se repitan las viviendas encuestadas, son consideradas observaciones independientes.

Primero se mostrará de manera general para todo el conjunto de observaciones la situación de fenómeno habitacional, para posteriormente hacer un análisis detallado del déficit a partir de cada uno de los indicadores.

En la figura 4.3, se muestra una gráfica con los porcentajes acerca de la condición deficitaria de las viviendas encuestadas en el AMM. Observando los resultados se puede notar que el porcentaje de viviendas deficitarias es 49.58% y el porcentaje de viviendas que no presentan condición de déficit es 50.42%, una diferencia menor a un punto porcentual entre estas dos categorías; de manera que la proporción de viviendas en condición de déficit es similar a las que no poseen esta condición. Lo más relevante de este resultado tiene que ver con la cantidad de viviendas que son requeridas para compensar esta situación de déficit, ya sea por la necesidad de construir más recintos habitacionales o porque los que ya existen deban ser mejorados o ampliados.

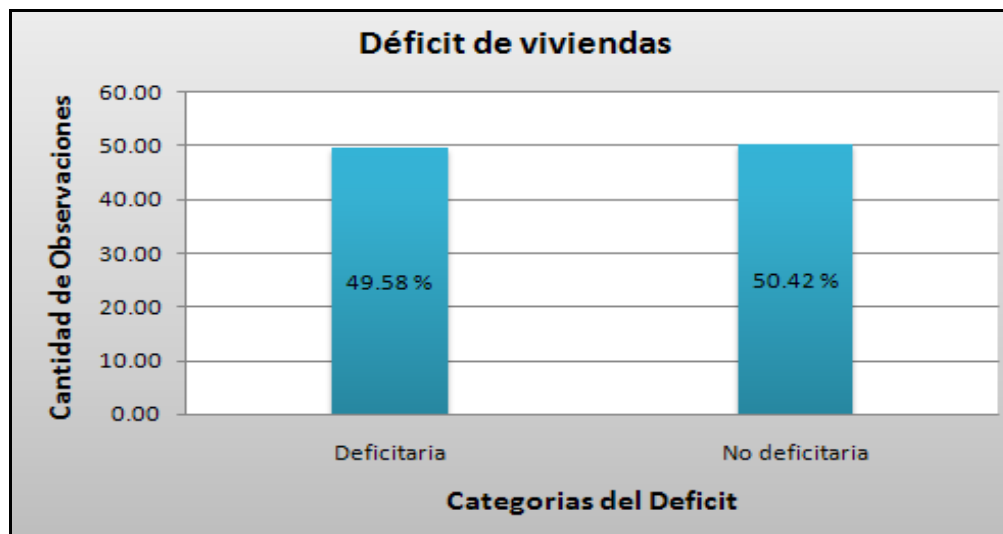


Figura 4.3. Porcentajes de la condición de déficit habitacional.

Luego de conocer estos resultados, es interesante indagar en cada uno de los indicadores de déficit de tal manera de poder observar la incidencia de cada uno de estos en la medición del déficit habitacional, ver la figura 4.4.

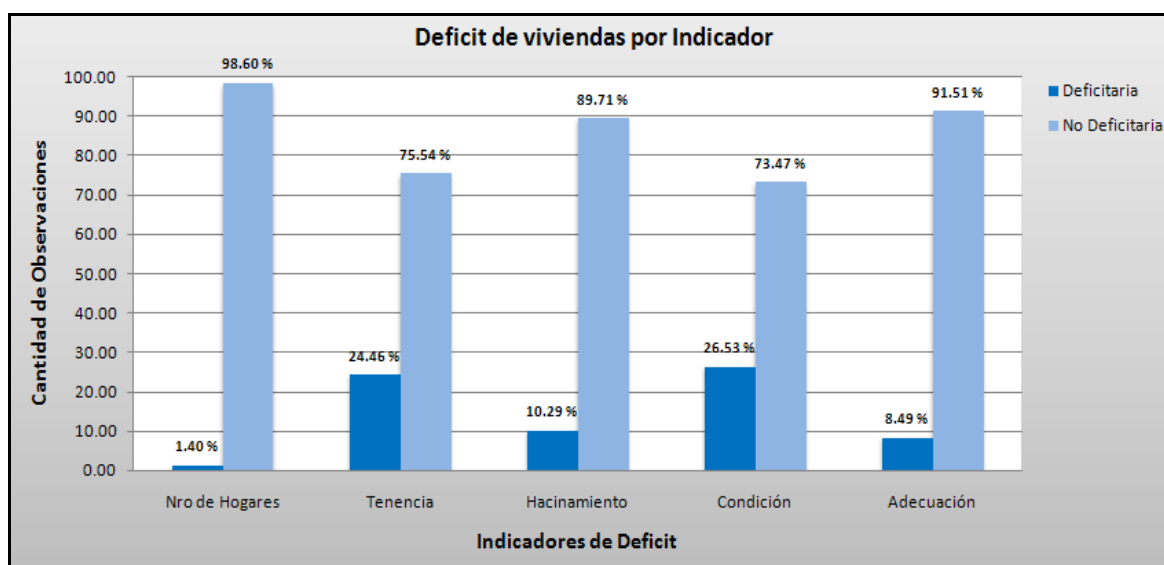


Figura 4.4. Influencia de los indicadores de déficit habitacional

La diferencia entre los porcentajes para la condición deficitaria y la no deficitaria es bastante notoria, lo que pareciera ser contradictorio respecto a la gráfica anterior. Se debe a que cada indicador por sí solo es condicionante del déficit habitacional, por lo que una vivienda es considerada deficitaria con el aporte de tan sólo un indicador o conjunto de estos.

Los indicadores que tiene mayor influencia en la condición deficitaria de la vivienda son la tenencia (24.46% Deficitaria, 75.54% No deficitaria) y la condición de la vivienda (26.53% Deficitaria, 73.43% No deficitaria); es decir, que con mayor peso una vivienda es considerada deficitaria, bien sea porque no es propia o porque presenta pésimas condiciones estructurales. Además, es de notar el bajo porcentaje

del indicador número de hogares, respecto a la situación deficitaria (1.46%); sin embargo, tal porcentaje es muy importante puesto que cuantifica las viviendas que albergan múltiples hogares. Respecto a los indicadores hacinamiento y adecuación de la vivienda su incidencia en la condición deficitaria es de 10,29% y 8.49% respectivamente.

La importancia de los indicadores es que ayudan en la toma de decisiones al momento de plantearse soluciones respecto al problema en estudio, ya que ofrecen un enfoque bien orientado de lo que posiblemente pueda estar ocasionando la situación deficitaria para aportar soluciones y medidas que sean eficaces.

Capítulo 5

Representación y análisis espacial

En este capítulo se muestra la influencia del déficit habitacional desde una perspectiva espacial mediante los Sistemas de Información Geográfica. Esto se realizará a partir de los resultados obtenidos de la aplicación de las distintas técnicas estadísticas para la medición del déficit habitacional, donde la componente espacial está determinada por la ubicación de aquellos sectores o segmentos de la muestra de la EHM dentro del Área Metropolitana de Mérida.

5.1 Construcción de la aplicación SIG

Este proceso consiste en la creación de una aplicación SIG que va a contener información geográficamente referenciada de los sectores o segmentos que conforman la muestra de la EHM para al Área Metropolitana de Mérida con el propósito de medir espacialmente el déficit habitacional. Para lograrlo se llevaron a cabo una serie de pasos que serán descritos detalladamente a continuación, con el fin de especificar cómo se lleva a cabo el proceso.

5.1.1 Obtención de los mapas cartográficos

Los mapas digitales son pieza fundamental al trabajar con un SIG ya que es la base para desplegar la información espacial. Cada mapa tiene asociada cierta información

atributiva, de manera que cada uno de ellos contiene cierta información espacial de interés; por lo que detrás de cada mapa se encuentra una base de datos relacional. Los mapas cartográficos que se utilizaron en esta investigación, fueron suministrados por el Instituto Nacional de Estadísticas (INE), ente oficial, encargado de realizar la división territorial y la selección de las muestras para la aplicación de las encuestas, en este caso la EHM. Algunos mapas complementarios fueron aportados por el Instituto de Fotogrametría de la Universidad de Los Andes.

5.1.2 Proceso de georeferenciación

Este proceso consiste en asignarle un sistema de coordenadas geográficas coherentes a la cartografía con el objeto de ubicarla en la región o territorio que le corresponde sobre el planeta tierra. Actualmente, para el caso de Venezuela el sistema de coordenadas geográfico utilizado se basa en el Datum Regven y el sistema de coordenadas proyectado por lo general es el sistema de coordenadas Transversal de Mercator o UTM por sus siglas en inglés, específicamente para la zona 19 en el hemisferio norte. Para el caso específico de este estudio se utilizó un software destinado al procesamiento de datos espaciales como lo es ArcGis Desktop 9.2, con el cual se le asignó una misma proyección geografía a cada uno de los mapas (layers).

5.1.3 Vinculación de los datos a los mapas

A partir de un conjunto de datos referenciales (atributos) tomados de la EHM, en relación a la ubicación espacial de las viviendas que son encuetadas, se logró la vinculación de estos datos a la cartografía digital organizada en el SIG ArcGis. La información atributiva contempla campos como: entidad federal, municipio, parroquia, control y segmento o sector. Cada uno de estos campos se describe en el

Anexo D. El nivel de detalle de la representación espacial es respecto a los segmentos. Para cada uno de ellos es seleccionado un grupo de 15 viviendas aproximadamente, a esto es lo que se denomina Lote o unidad de muestreo. La división del territorio en segmentos es aplicada por el INE en todo el país una vez realizado el censo nacional de población, donde cada segmento/sector constituye un cantidad aproximada de 180 a 250 viviendas para área urbana y entre 150 y 180 para área rural.

Una vez conocida la información referencial acerca de la ubicación de cada segmento, se le asocia un código único denominado control, el cual es un identificador que representa aquellos segmentos encuestados dentro del AMM. Luego, se usa este identificador con el objeto de vincular la información generada del análisis estadístico a los mapas, de manera que el código de control va a estar presente tanto en la tabla atributiva de la información espacial como en los datos que se manejaron para realizar los análisis estadísticos en la medición de déficit habitacional. La figura 5.1 ilustra este proceso.

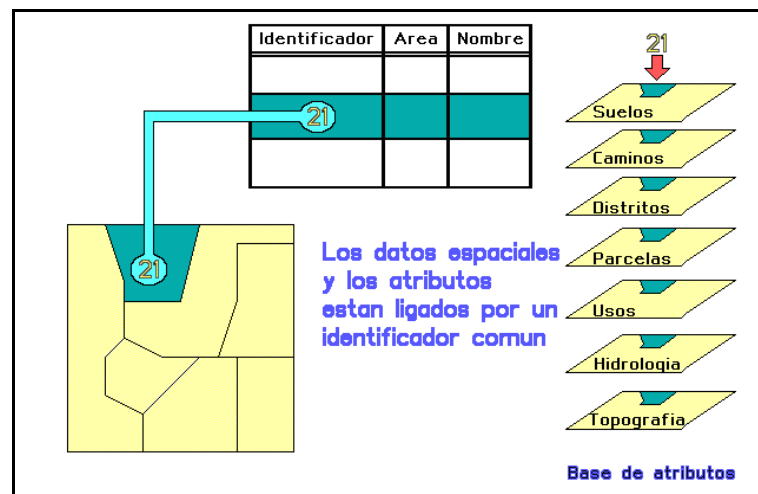


Figura 5.1. Esquema del proceso de vinculación de los datos al mapa.

De esta manera se tienen dos bases de datos que es importante destacar, una de ellas contiene la información espacial, mientras la otra contiene la información de cada vivienda; es decir, la información referente a los indicadores de déficit de cada vivienda.

5.2 Representación espacial de la información

Para la representación de los resultados en el mapa se dispone de un conjunto de capas (layers); cada capa representa un mapa distinto con una información geográfica particular, con el fin de tener una mejor percepción de la ubicación de cada control dentro del área que le corresponde. Las capas se encuentran en formato vectorial y pueden ser de tipo punto, línea o polígono, de acuerdo a la información que se esté representando. En la tabla 19 se describen cada una de las capas:

| Nombre | Tipo | Descripción |
|-------------------|----------|--|
| Estados | Polígono | Divisiones por estados de territorio Nacional |
| Parroquias Mérida | Polígono | Divisiones por parroquia del estado Mérida |
| AMM | Polígono | Los cuatro municipios que conforman el AMM |
| Segmentos | Polígono | Divisiones en segmentos de las parroquias del AMM |
| Controles | Polígono | Segmentos de la muestra que fueron encuestados |
| Vialidad | Línea | Principales arterias viales del municipio Libertador |
| Casas y Edificios | Polígono | Casas y edificios del municipio Libertador |
| Hidrografía | Línea | Ríos más importantes que atraviesan el AMM. |

Tabla 19. Capas con información geográfica.

Cada capa tiene asociada una base de datos (Tabla de atributos), de tal manera que se puede acceder en el caso de que sea requerido. En la figura 5.1 se muestra la capa de las divisiones por parroquias del estado Mérida y la tabla de atributos asociada a la misma.

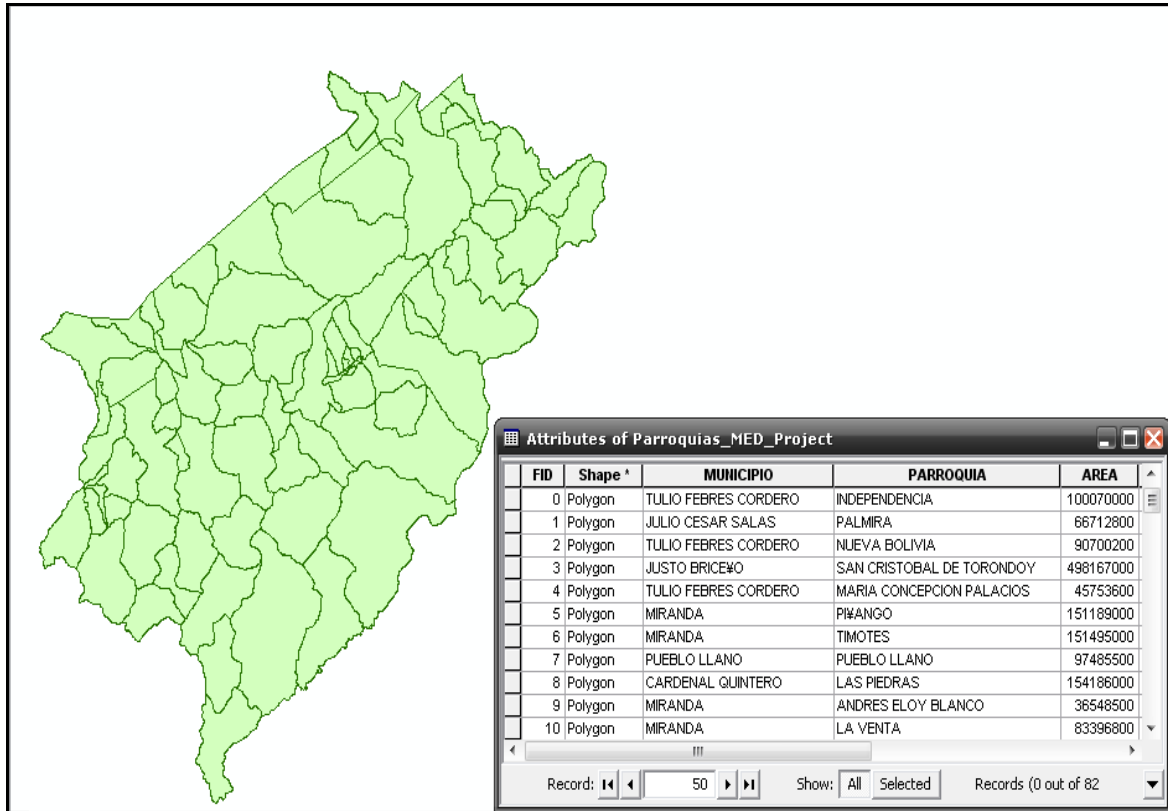


Figura 5.2. Capa de las parroquias de Mérida y su tabla atributiva.

5.3 Consulta y despliegue de la información espacial

A partir de la tabla atributiva es posible realizar consultas sobre cualquier capa y de esta manera poder seleccionar ciertos atributos de los cuales se desee obtener información y posteriormente visualizar el resultado obtenido. Para el caso de estudio, es posible realizar distintas consultas como por ejemplo; consultar los

controles que pertenecen a un determinado municipio o parroquia: si se conoce el código del control se puede consultar directamente por este campo. De la misma manera se procede para cualquier campo en la tabla de atributos de cualquier capa.

Para ilustrar esto en el SIG implementado, a manera de ejemplo, se consultaron aquellos controles que pertenecen sólo al municipio Santos Marquina, en este caso son dos los seleccionados (2091 y 1089) que son los únicos para este municipio, como se muestra en la figura 5.3.

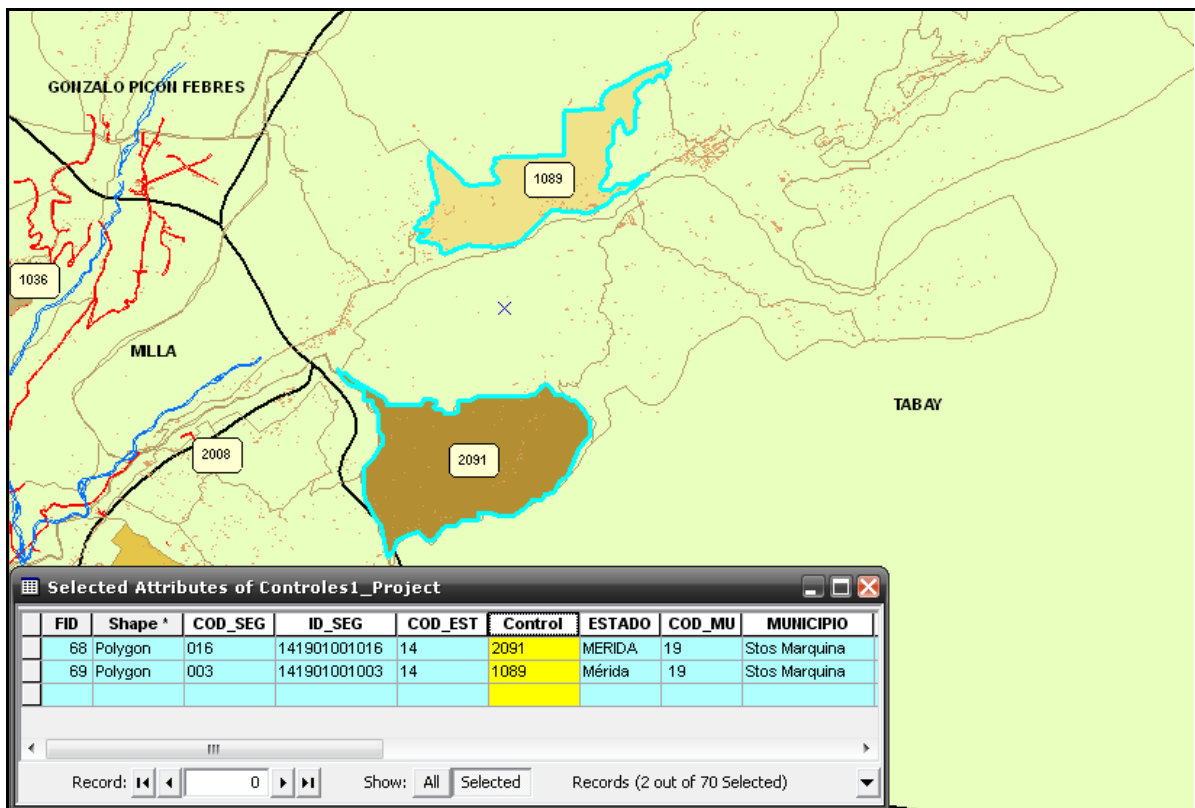


Figura 5.3. Selección a partir de los atributos de las tablas.

Como se observa más abajo, los dos controles ubicados en el municipio Santos Marquina han sido seleccionados tanto en el área de despliegue del mapa como en la tabla de atributos de la capa. Un SIG despliega datos espaciales y tabulares; ambos tipos de datos pueden ser desplegados en pantalla o bien generar mapas en impresora o plotter, de manera que el despliegue tiene que ver con la manera como se muestra información geográfica al usuario.

A manera de ejemplo, se generó un mapa de impresión de los controles ubicados en el municipio libertador con se muestra en la figura 5.4, para lograr esto, primero se realiza una consulta de los controles que pertenecen al municipio libertador, para luego agregar todos los elementos distintivos de un mapa.

Se observa el despliegue del mapa con la información espacial del déficit habitacional de todas las capas en forma de mapas temáticos; además, los elementos distintivos de un mapa, tales como, leyenda, flecha de norte, barra de escala y título. De esta manera es presentada la información espacial al usuario.

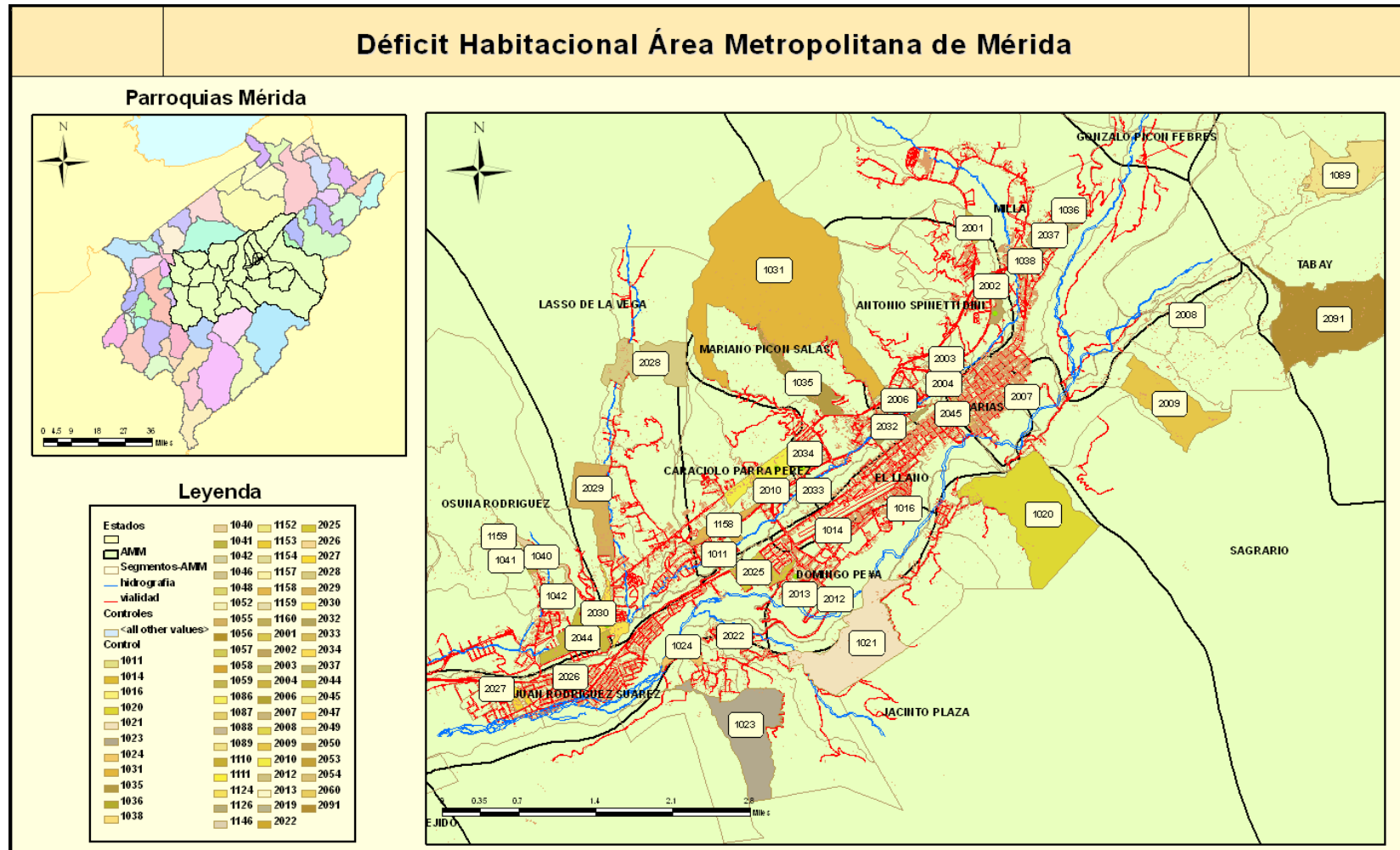


Figura 5.4. Mapa para impresión y sus elementos.

5.4 Análisis espacial del déficit habitacional

El análisis espacial es una de las funciones más relevantes de un SIG, puesto que permite generar nueva información a partir de la ya existente. Además, permite obtener respuestas y soluciones a un problema en particular.

En este caso particular, para la medición del déficit habitacional en el espacio geográfico que define el AMM, consistió en generar estadísticas en el mapa referente a la información contenida en cada uno de los controles que incluye un determinado número de viviendas e información para cada vivienda con los resultados estadísticos generados con anterioridad.

Luego de haber vinculado los controles a los mapas georeferenciados, y aplicado los métodos estadísticos donde se obtuvieron los indicadores y la clasificación del déficit habitacional, se establece una relación entre esta información y los datos espaciales referentes a los controles. Se trata de relacionar la base de datos espacial con una base de datos externa. Mediante el SIG se realizó una relación de tablas cuya cardinalidad es de uno a muchos (a un control se le asocian *n* viviendas), de esta manera se accede a la información de las viviendas encuestadas; la figura ilustra la relación.

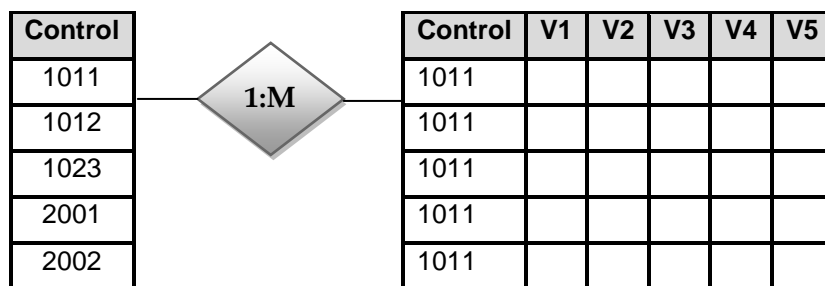


Figura 5.5. Relación de las tablas y su cardinalidad.

Si se desea conocer la condición de las viviendas referentes a un control específico, se procede a realizar una selección en la tabla de atributos específica, consultando el código del control sobre el que se desea obtener la información y automáticamente se selecciona la información de aquellas viviendas relacionadas al control. En la figura 5.6 se muestra la selección de un control (el control 1011), y en la figura 5.7 se muestra como es seleccionada la información en las tablas, además se observa claramente la relación que se establece entre las mismas.

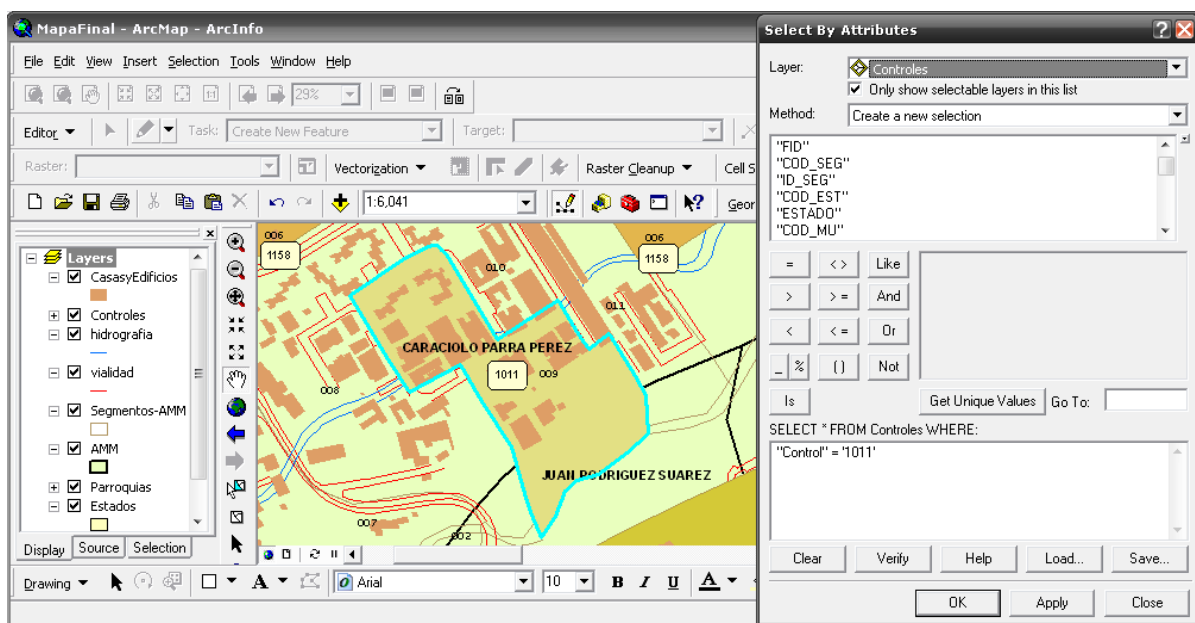


Figura 5.6. Selección por atributos del control 1011.

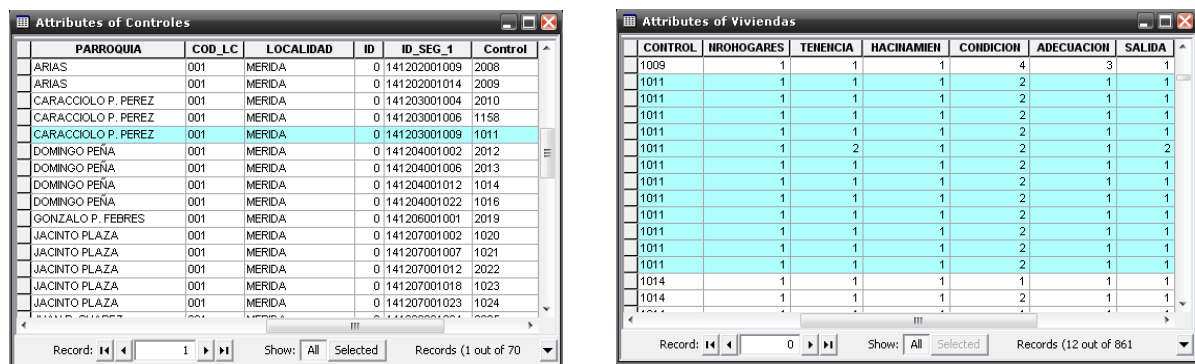


Figura 5.7. Vista en las tablas de atributos de los elementos consultados.

Se observa que el segmento identificado con el control 1011 es resaltado en el mapa y en la tabla atributiva. En el lado derecho de la figura 5.6 se muestra el constructor de consultas (*QueryBuilder*), provisto por el software SIG, con el fin de interactuar con las bases de datos que están asociadas a los mapas y así poder realizar consultas tanto simples como complejas, dependiendo de la necesidad y la relación que exista entre los datos. Una vez que se selecciona un control, automáticamente las instancias relacionadas al mismo también son seleccionados y resaltados en la tabla de atributos (base de datos externa).

A partir de la información seleccionada previamente, con el software SIG se pueden generar estadísticas por cada indicador para cada control; para el caso específico del control seleccionado en la figura anterior, las estadísticas generadas se podrán visualizar en los gráficos siguientes:

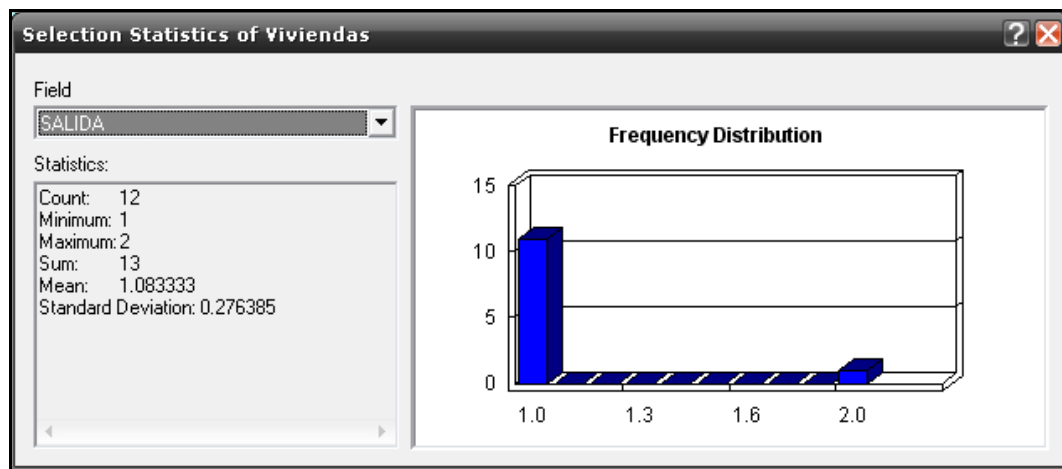


Figura 5.8. Estadísticas del déficit habitacional para el control 1011.

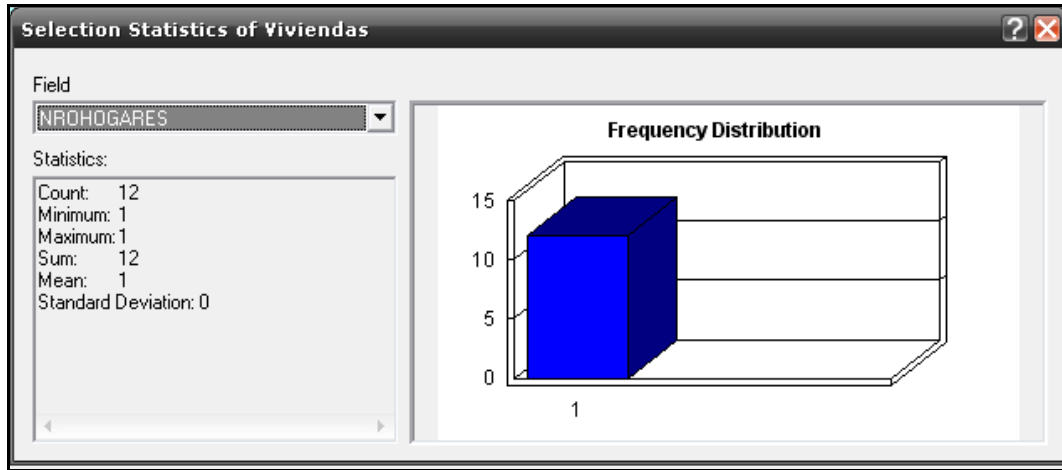


Figura 5.9. Estadísticas del indicador Número de Hogares para el control 1011.

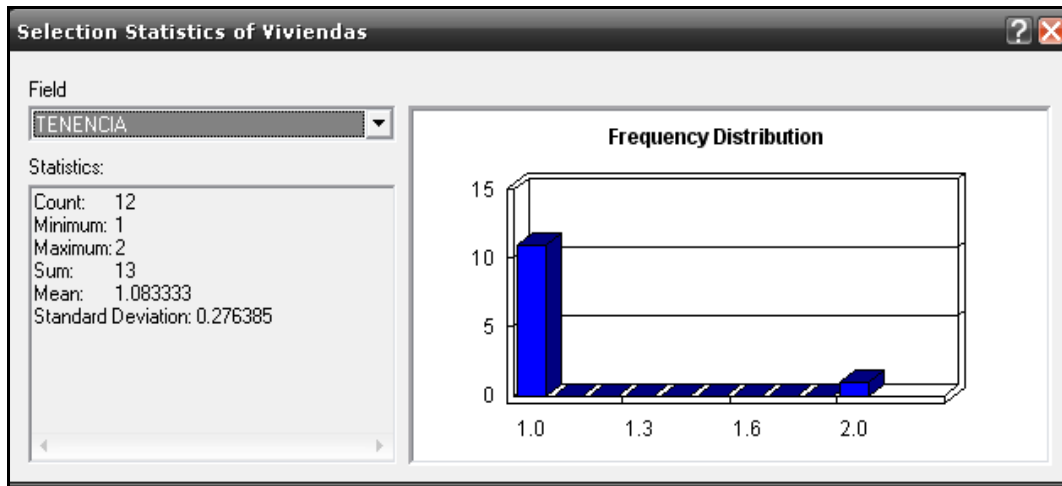


Figura 5.10. Estadísticas del indicador Tenencia para el control 1011.

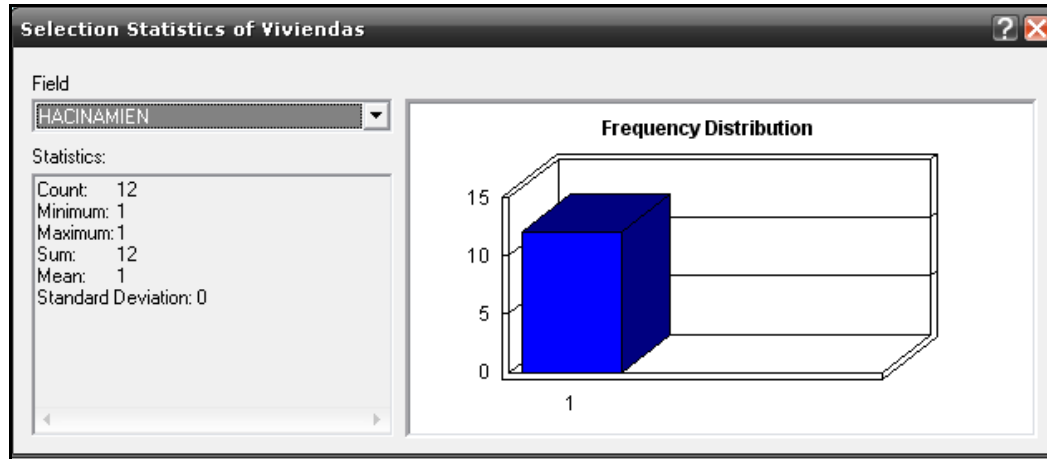


Figura 5.11. Estadísticas del indicador Hacinamiento para el control 1011.

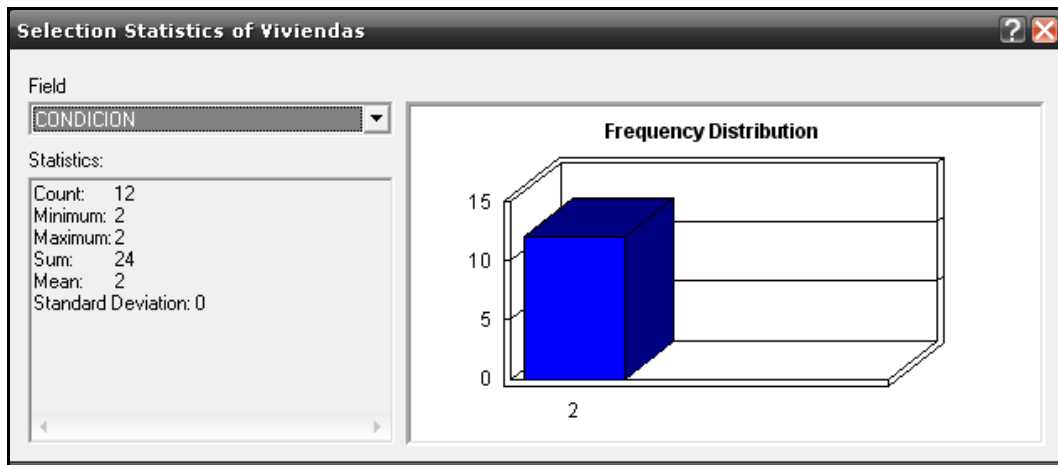


Figura 5.12. Estadísticas del indicador Condición para el control 1011.

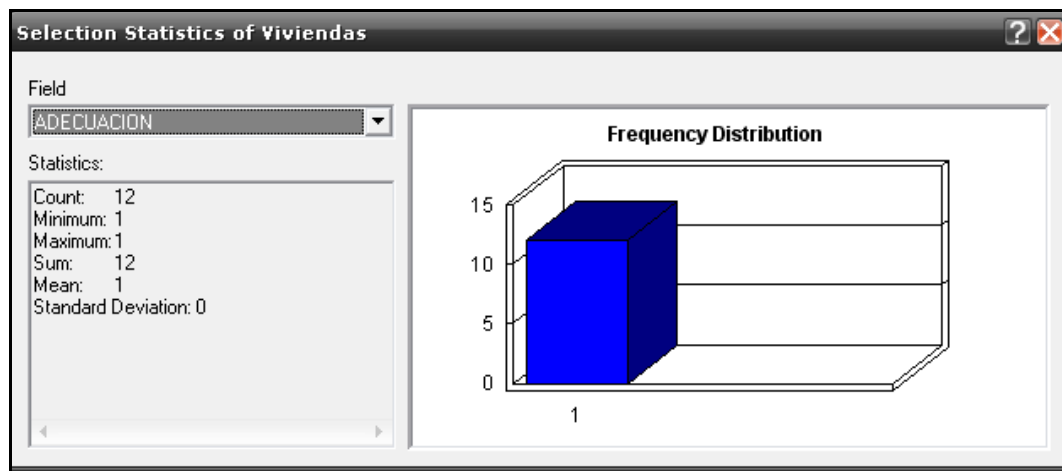


Figura 5.13. Estadísticas del indicador Adecuación para el control 1011.

En resumen, las graficas anteriores indican que este control está constituido por una muestra de 12 viviendas, donde sólo una de las viviendas es considerada en condición deficitaria, mientras que las 11 restantes no presentaron tal condición. Además del déficit habitacional, se muestra la influencia de cada indicador, en este caso particular, se observa que la vivienda es considerada deficitaria por concepto de tenencia. Se puede notar que las estadísticas a nivel de los controles ofrecen un mayor nivel de detalle al momento de medir el déficit habitacional. De cada control y de manera semestral es posible generar estas estadísticas, de este modo hacer comparaciones entre los controles e inferir cuales son los sectores más afectados por el problema del déficit habitacional dentro del Área Metropolitana de Mérida.

5.5 Visualización de los resultados vía web

Con el propósito de que cualquier usuario tenga acceso a los resultados del análisis espacial, se creó una interfaz sencilla mediante la configuración de un servidor cartográfico o visor de mapas. La función de este visor es ofrecer un servicio de

datos cartográficos y alfanuméricos a través de una dirección http, que el usuario puede incluir en un portal web, proporcionado unas herramientas y funcionalidades para la interacción con dicho visor, de esta manera no hay necesidad de que un usuario cualquiera tenga instalado un software SIG para poder visualizar la información espacial.

Es importante destacar que en este caso la función del servidor es sólo la de visualizar los resultados generados del análisis espacial; es decir, aquellos resultados producto de consultas ya realizadas en el SIG previamente creado. El hacer tales análisis directamente en el servidor como una aplicación web, no es el objetivo que se contempla en este estudio, pero que abre la brecha para que se desarrollen este tipo de aplicaciones. Las herramientas utilizadas para el desarrollo del visor fueron:

MapServer: Es un entorno de desarrollo en código abierto (*Open Source Initiative*) para la creación de aplicaciones SIG en Internet/Intranet con el fin de visualizar, consultar y analizar información geográfica a través de la red. No es un SIG propiamente dicho, sino que está destinado a renderizar datos espaciales (mapas, imágenes, datos vectoriales) para su publicación a través de la web.

Pmapper: es un entorno de desarrollo de aplicaciones *Web Map* eficazmente configurable. Está basado en *MapServer* y *PHP/MapScript*. Funciona con todos los formatos de datos soportados por *MapServer* y además implementa los estándares del *OpenGis Consortiums* para *Web Mapping Services* (WMS). El entorno *Pmapper* ofrece una gran facilidad y sencillez para configurar un entorno cliente/servidor así como una completa funcionalidad y múltiples modalidades para facilitar la configuración de una aplicación *MapServer*.

5.5.1 Arquitectura de una aplicación de MapServer

MapServer generalmente se ejecuta como una aplicación *Common Gateway Interface* (CGI)⁴ desde el servidor http. La aplicación *MapServer* CGI usa los siguientes recursos:

- Un servidor http como *Apache* o *Internet Información Server* (IIS).
- El software *MapServer*.
- Un archivo mapa (.map), cuya función es indicar a *MapServer* que debe hacer con los datos.
- Un archivo plantilla que controla la interfaz del usuario de la aplicación *MapServer* en una ventana *Web browser*.
- Un conjunto de datos SIG.

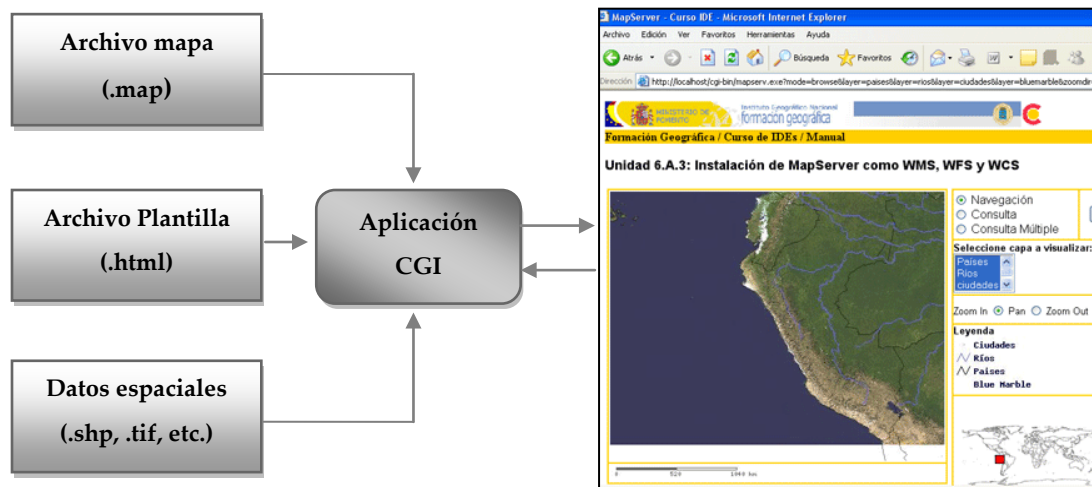


Figura 5.14. Arquitectura de *MapServer*.

⁴ CGI es una norma para establecer comunicación entre un servidor web y un programa, de tal manera que este último pueda interactuar con internet.

5.5.2 Elementos característicos del visor cartográfico

5.5.2.1 El archivo mapa

El archivo principal de configuración de *MapServer* es un archivo de texto, con extensión “.map”, en el se incluye una serie de parámetros que definen las capas disponibles en el servicio, el estilo con que se representarán, su simbología, formato en que se generará la imagen, el sistema de referencia, etc. Un archivo ‘mapa’ es jerárquico y consta de varias secciones.. El contenido de las secciones consiste en la definición de determinados parámetros del tipo atributo - valor. La sección principal es el objeto .map, la cual anida a otras secciones, como se observa en la siguiente figura. Para más detalle sobre los parámetros que pueden utilizarse en el archivo mapa consultar la información complementaria provista en el Anexo D.

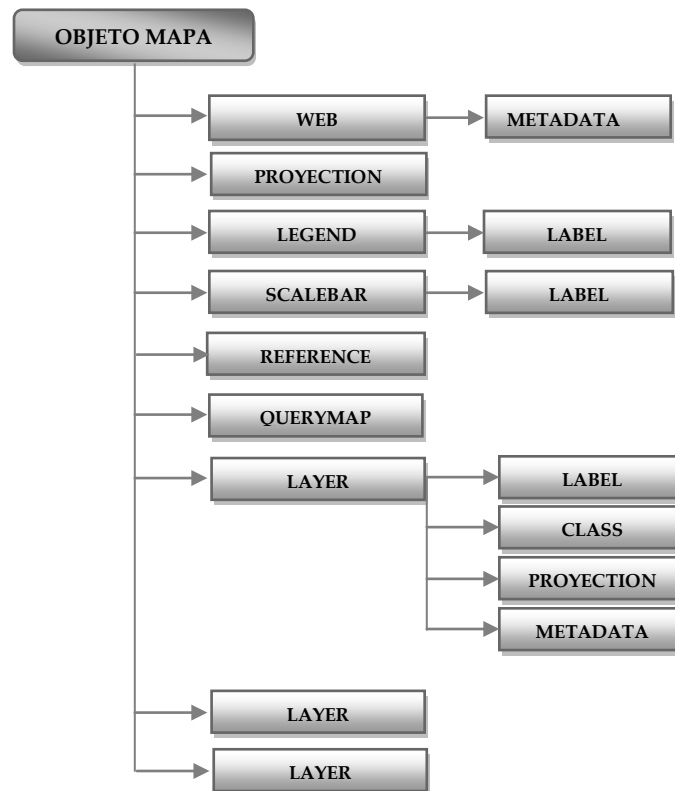


Figura 5.15. Esquema del archivo mapa.

5.5.2.2 El archivo plantilla

Este archivo controla las salidas de los mapas, leyendas y cualquier otra respuesta de *MapServer*, en una página HTML. Determina las formas en que el usuario puede interactuar con la aplicación (consultas, zoom, etc.). Opera como cualquier archivo HTML. A través de la plantilla se pasan los parámetros de peticiones al CGI, y luego *MapServer* reemplaza determinadas palabras claves en el archivo plantilla con las respuestas del servidor.

5.5.2.3 El conjunto de datos SIG

Son los datos a partir de los que se generan los mapas en las aplicaciones. Estos datos pueden ser tipo raster o vectorial. También es posible conectarse a bases de datos para consultar datos alfanuméricos.

5.6 Resultados de configurar el servidor de mapas

5.6.1 Interfaz de usuario

El diseño de la aplicación tiene la ventaja de que el usuario que potencialmente accederá a ella, no necesita ser un especialista en SIG para poder visualizar la información espacial, relacionada con el déficit de viviendas. Para ello, tanto la manera de acceder, como la utilización de las herramientas que aparecen en la aplicación, son bastante sencillas. En la figura 5.16 se muestra la pantalla inicial que se despliega al iniciar la aplicación. En el anexo V se describe detalladamente la estructura de esta interfaz.

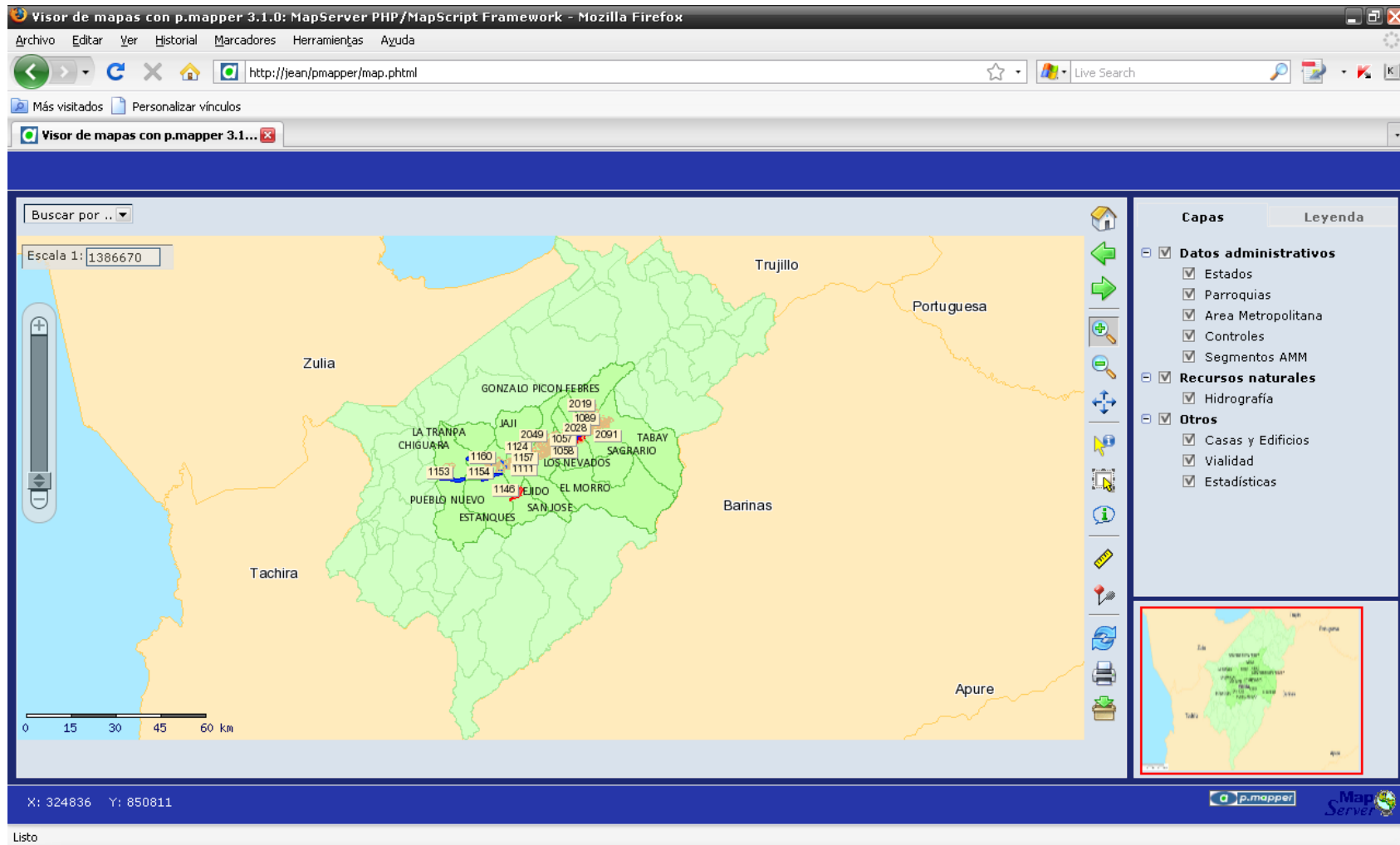


Figura. 5.16. Página principal del visor de mapas.

5.6.2 Funcionalidades del visor web de mapas

El visor cartográfico se inicia ofreciendo una vista general de todo el área de interés, a partir del cual el usuario podrá navegar a través del plano en busca de información. Posee también la herramienta de Zoom Ventana, muy demandado por los usuarios de cualquier visor SIG, con la cual se puede controlar la zona a la que se quiere acercar. Las herramientas Zoom Más y Zoom Menos sirven tanto para acercarse como para alejarse de la información espacial. Estas herramientas están también implementadas en forma de una barra, lo que facilita al usuario conocer en todo momento el nivel de zoom en el que se encuentra y cómo poder interactuar con él Espuela (2008). De igual manera cuenta con una serie de funciones propias de una aplicación SIG:

Visualización y selección de capas

Con esta herramienta se controla la visualización de las capas, pudiendo configurarse la aplicación para mostrar por defecto ciertas capas seleccionadas, así como también activar y desactivar cualquier capa cuando sea necesario, de igual manera se tiene en una pestaña aparte ubicada en la leyenda de las capas representadas.

Consultas por selección y búsquedas

Otra de las funcionalidades de cualquier SIG es permitir la búsqueda y consulta de información. En este caso, con esta herramienta, se pueden realizar búsquedas de información geográfica mediante la consulta de un atributo alfanumérico, como en el caso de estudio se puede consultar por el código de control en la capa de controles o por el nombre de la parroquia o municipio en la capa de parroquias.

Identificación de atributos de elementos del mapa

Esta es una de las funcionalidades básicas que se exigen a un SIG. Con esta herramienta se puede seleccionar el elemento que se quiere consultar y obtener toda la información alfanumérica asociada a dicho elemento.

Generación de mapas temáticos personalizados

Mediante la configuración del archivo .MAP se puede generar una representación temática de los datos alfanuméricos asociados a los objetos del mapa.

Impresión básica y avanzada

Tras cualquier análisis, estudio o consulta, resulta necesario presentar el resultado del mismo. En esta aplicación, la herramienta Imprimir permite obtener impreso el mapa en papel ó en formato PDF, incorporando además la leyenda de aquellos objetos que se visualizan en el mismo. Existe también la posibilidad de exportar el mapa en formato TIFF.

El visor de mapas le introduce un aspecto atractivo a la aplicación y facilidad al usuario para interactuar con la misma, si bien, inicialmente en los objetivos de ese trabajo no se planteó el diseño de una interfaz dinámica de usuario a fin de interactuar con los datos espaciales, es muy interesante que se desarrollen aplicaciones en esta dirección. De tal manera que este estudio sirve de base o iniciativa para la creación de nuevas aplicaciones SIG en entorno web.

Capítulo 6

Conclusiones y Recomendaciones

En este capítulo se presentan las conclusiones a las que se han llegado una vez culminado el estudio así como también una serie de recomendaciones para investigaciones posteriores que sigan esta misma línea de ideas entorno al fenómeno del déficit habitacional.

6.1 Conclusiones

En general, con esta investigación se lograron excelentes resultados gracias a la consolidación de un trabajo interdisciplinario donde no sólo se muestran las bondades de las diversas técnicas estadísticas y máquinas inteligentes, sino que además, se puede constatar la importancia y utilidad de los Sistemas de Información Geográfica al integrarlos en este tipo de investigaciones con tendencia al reconocimiento de patrones y descubrimiento de nuevo conocimiento. En este estudio la conjunción de estas técnicas permitió clasificar, cuantificar y representar el déficit de viviendas en el espacio que define el Área Metropolitana de Mérida con un alto grado de efectividad, demostrando de este modo que esta combinación es un aporte fundamental para la toma de decisiones, que en este caso particular, está vinculado a la vivienda como una variable de tipo social.

De manera más específica cabe destacar la importancia de aplicar las técnicas de análisis multivariante de datos. Estas permitieron la familiarización y tratamiento de los datos. De igual manera contribuyeron al reconocimiento de ciertos patrones y a la reducción de la dimensionalidad de las variables originales para así construir las variables indicadoras de déficit. Una de las ventajas del ACM es poder observar a través de gráficas, las relaciones que se establecen entre las distintas variables indicadoras del déficit. En los mapas perceptuales se pudo observar claramente estas relaciones. La asociación del déficit cuantitativo y el cualitativo describió con mayor información el déficit habitacional por el aporte de las variables número de hogares, tenencia, hacinamiento, condición y adecuación.

Es preciso destacar la importancia de utilizar las MVS como técnica de clasificación y predicción de datos ya que el poder de generalización de esta técnica resultó ser excelente como quedó demostrado en los resultados previos. Así, nuevas instancias de datos usados por la máquina entrenada logra excelentes pronósticos. La incorporación de las funciones *kernels*, la cual es una de las características más atractivas de las MVS interpretaron la no linealidad del fenómeno. Más que un modelo de clasificación es un modelo de pronóstico. Para las muestras prototipo la máquina podrá clasificar nuevas instancias con un porcentaje de exactitud alrededor de 99.8308 con una variación de 0.0791. En este sentido se destaca su alto nivel de pronóstico.

Por último, la incorporación del análisis espacial para el estudio del fenómeno del déficit habitacional a través del uso de un SIG, aportó mucha practicidad y elegancia al estudio. Los resultados de los análisis no son mostrados en simples tablas sino que para ello se cuenta con una interfaz constituida por un visor cartográfico, donde accediendo a una dirección web es posible interactuar con una

serie de mapas, que ofrecen información del déficit de viviendas en los sectores que conforman el Área Metropolitana de Mérida.

6.2 Recomendaciones

A continuación se presentan una serie de recomendaciones que puedan ser de utilidad y ayuda para futuros trabajos, con la finalidad de que se sigan desarrollando aplicaciones en esta dirección.

Es recomendable utilizar otros tipos de funciones kernels en el proceso de entrenamiento de las MVS, con la finalidad de tener parámetros de comparación en este ámbito de estudio respecto al desempeño de los modelos. Estos algoritmos dependen en gran medida del *kernel* que se utilice. De manera que es necesario determinar una buena técnica para ajustar los parámetros de la MVS.

Aplicar esta misma metodología a otro tipo de investigaciones de carácter social como por ejemplo a los datos de la Encuesta Nacional de Presupuesto Familiar, con el fin de realizar un análisis espacial de los patrones de consumo o de gastos de los hogares venezolanos de una determinada región.

Realizar los esfuerzos posibles para conseguir que el nivel de detalle de la representación espacial sea a nivel de las viviendas y no de los segmentos, lo que conduciría a un estudio mucho más detallado e interesante en torno a la problemática del déficit habitacional.

Referencias Bibliográficas

- Bautista, E. Guzmán, E. y Figueroa, J. (2004). Predicción de múltiples puntos de series de tiempo utilizando Support Vector Machines. Instituto Politécnico Nacional, computación y sistemas, México.
- Benavides, M. Allende, H. y Seccatore L. (2006). *Máquinas de soporte vectorial en regresión*. Reporte Técnico. Universidad Adolfo Ibáñez, Facultad de Ciencia y tecnología, Viña del mar, Chile.
- Betancourt, G. (2005). *Las Máquinas de Soporte Vectorial (SVMs)*. Scientia et Technica, Año XI, N° 27, UTP. ISSN 0122-1701, Colombia.
- Bosque, J. (1992). Sistema de información Geográfica. Ed. Rialp, Madrid, España.
- Hsu, Ch., Chang Ch. y Lin Ch. (2008). *A practical guide to Support Vector Classification*. Reporte Técnico. Department of computer Science and Information Engineering, National Taiwan University, Taiwan. Documento extraído el día 10 de abril de 2009 desde URL: <http://www.csie.ntu.edu.tw/~cjlin>.
- Chang Ch. y Lin Ch. (2008). *LIBSVM: a library for Support Vector Machine. Tutorial*. University of Taiwan.
- Cordero, J. (2007). *Sistemas de Información Geográfica orientado al geomarketing: Manual de Mapserver 4.x*. Loja, Ecuador.
- Espuela, J. (2008). *WebSIG ParcCollserola, Publicación de datos geográficos del Consorci del Parc de Collserola en un entorno web*. Proyecto de grado. Universidad Autónoma de Barcelona, Departamento de Geografía, España.

- García, A. (2006). *Cuantificación de las características físicas de la vivienda urbana en México*. Boletín de los sistemas nacionales estadísticos y de información geográfica. Vol. 2, Aguascalientes, México.
- García, M. (2000). *DEFICIT Y NECESIDAD FUTURA DE VIVIENDAS: Proceso Metodológico*. Universidad de Los Andes, Mérida, Venezuela.
- Gil, A. (2008). *Pronostico del déficit de viviendas en el Estado Mérida a través de Redes Neuronales Artificiales*. Proyecto de Grado. Universidad de Los Andes, Escuela de Ingeniería de Sistemas.
- González, V. (2005). *Mapserver y su aplicación a SIG*. Equipo de sistemas de información geográfica. Universidad Técnica Particular de Loja, Ecuador. Documento extraído el 2 julio 2009 desde URL: <http://serversig.utpl.edu.ec/sigutpl/index.php>.
- Gruson, A. (2004). *Hogar y núcleo familiar en la Encuesta de Hogares por Muestreo*. Reporte Técnico. Centro de Investigaciones en Ciencias Sociales, Universidad Católica Andrés Bello, Caracas, Venezuela.
- Gunn, S. R. (1997). *Support Vector Machines for Classification and Regression*. Reporte técnico. Image Speech and Intelligent Systems Research Group. Universidad de Southampton, Inglaterra.
- Hair, J., Anderson, R., Tatham, R. y Black, W. (1999). *Análisis Multivariante*. 5ª Edición, Prentice-Hall, Madrid, España.
- Hernández, J., Ramírez, M. y Ferri, C. (2004). *Introducción a la Minería de Datos*. Editorial Pearson Educación S.A., Madrid, España.
- INE (2008). 'Instituto Nacional de Estadística'. Sitio Web visitado el 5 de Junio de 2008 desde la URL: <http://www.ine.gov.ve/>.

- Ley del régimen Prestacional de Vivienda y Hábitat, Gaceta Oficial N° 38.182 del 9 de mayo del 2005. Artículos 12 y 13. Extraído el 25 noviembre de 2008 desde URL: <http://www.mintra.gov.ve/legal/leyesorganicas>
- Linares, A. (2004). *Déficit habitacional Venezolano en el 2004*. Reporte Técnico, Cámara Venezolana de la Construcción.
- Moro, Q. y Hurtado, A. (2006). *Introducción al diseño de experimentos para el reconocimiento de patrones*. Documento extraído el 20 febrero de 2009 desde la URL: http://www.infor.uva.es/~isaac/doctorado/Cap07_SVM.pdf.
- NCGIA (1990). 'National Center for Geographical Information and Analysis', vol.1. Santa Bárbara, Universidad de California.
- Padrón, D. (2009). *Desarrollo de Servidores de Mapas con Software Libre*, Universidad de Alcalá, España. Documento extraído el 10 julio 2009 desde URL: <http://dpadron.blogspot.com/>.
- Parra, A. (2008a). *Guía práctica de conceptos y ejercicios de Sistemas de Información Geográfica*. Universidad de Los Andes, Instituto de Fotogrametría, Mérida, Venezuela.
- Posada, N. (1999). *Modelado de datos orientado a objetos para un sistema de información geográfica*. Proyecto de grado. Universidad de las Américas-Puebla, Ingeniería en Sistemas Computacionales.
- Peña, D. (2002). *Análisis de datos multivariantes*. Editorial Mc Graw Hill. Madrid, España.
- Pérez, C. (2004). *Técnicas de Análisis Multivariante de Datos; Aplicaciones con SPSS*. Editorial Pearson Educación S.A., Madrid, España.
- Quintero, B. (2000). *El Déficit Habitacional en Bogotá: Un Análisis Espacial*. Reporte Técnico. Facultad de Geografía e Historia, Universidad de Barcelona, España.

-
- Salvador, M. (2001). *Análisis de conglomerados o cluster*. Documento extraído el 8 de enero de 2009 desde URL: <http://www.5campus.org/leccion/cluster>.
- Salvador, M. (2003). *Análisis de Correspondencia*. Documento extraído el 20 enero 2009 desde URL: <http://www.5campus.com/leccion/correspondencias>.
- Santiago, I. (2006). *Fundamentos de ArcGís*. Área de Tecnologías de Información Gubernamental Oficina de Gerencia y Presupuesto del Estado Libre Asociado de Puerto Rico, Versión 0.1, San Juan, Puerto Rico. Documento extraído el 9 de enero 2009 desde URL: <http://www.ogp.gobierno.pr>.
- Szalachman, R. (1999). *Un perfil del déficit de vivienda en Bolivia*. Reporte técnico, Comisión Económica Para América Latina y el Caribe (CEPAL), Santiago de Chile, Chile.
- Vapnik, VN. (2000). *The Nature of Statistical Learning Theory*. Springer, New York, EEUU.
- Zeiler, M. (1999). *Modeling Our World*. Environmental Systems Research Institute Inc., California, EEUU.

Anexo A

Diccionario de las variables de la EHM inherentes al déficit habitacional.

| Variables | Campo | Descripción del campo | Valores |
|--------------------------|------------------|---|---------|
| Cod.Unico | Entidad | Identifica el estado al que pertenece la observación | 12,14 |
| | Control | Identifica un segmento/sector de la muestra | 1-999 |
| | Área | Identifica las subdivisiones del sector de la muestra | 1-999 |
| | Línea | Identifica la vivienda encuestada | 1-999 |
| | Serie | Identifica a la encuesta realizada | 1-999 |
| | Semestre | Indica el semestre en que fue realizada la encuesta | 1,...25 |
| NroHogares | NroHogares | Número de hogares en una vivienda | 1,...5 |
| Tenencia | Tenencia | Condición de tenencia para cada encuestado | 1,...8 |
| Hacinamiento | Ocupantes | Número de habitantes en la vivienda | 1,...20 |
| | Nro_Dormitorios | Número de cuartos utilizados para dormir | 1,...20 |
| Condición de la Vivienda | Tipo de vivienda | Tipo de vivienda | 1,...9 |
| | Paredes | Material de las Paredes | 1,...6 |
| | Techo | Material del Techo | 1,...6 |
| | Piso | Material del Piso | 1,...4 |

| | | | |
|---------------------------|-------------------------------------|---|---------|
| Adecuación de la vivienda | Suministro de agua | A esta vivienda llega el agua por: | 1,2,3,4 |
| | Servicio de eliminación de excretas | De qué manera se realiza la eliminación de excretas | 1,2,3,4 |
| | Servicio eléctrico público | Indica si se cuenta con el servicio eléctrico | 1,2 |
| | Recolección directa de basura | Indica si se cuenta con recolección directa de basura o aseo urbano | 1,2 |
| | Contenedor de basura | Indica si se cuenta con contenedor de basura | 1,2 |

Categorías adoptadas por cada una de las variables

| Variables | Categorías |
|--------------------------------|---|
| Tipo de vivienda | Quinta (1) Casa(2) Apartamento en Edificio(3) Apartamento en casa o Quinta (4) Casa Vecindad (5) Vivienda Rustica(6) Rancho Campesino(7) Otro Tipo(8) Colectividad(9) |
| Material de las Paredes | Bloque o ladrillo frisad(Acabado) (1) Bloque o ladrillo sin frisar(No acabado) (2) Madera Aserrada(3) Adobe-Tapia-Bahareque Frisado(4) Adobe-Tapia-Bahareque sin frisar(5) Otros (Caña, palma, tablas, etc.) (6) |
| Material del Techo | Platabanda(1) Teja(2) Láminas asfálticas(3) |

| | |
|--|--|
| | <p>Fibrocemento, cemento ligero y similares(4)</p> <p>Laminas metálicas(Zinc y similares) (5)</p> <p>Asbesto y similares (6)</p> <p>Otros(Palmas, tablas y similares) (7)</p> |
| Material del Piso | <p>Mosaico, granito, vinil, cerámica, terracota, parquet, alfombra y similares (1)</p> <p>Cemento(2)</p> <p>Tierra(3)</p> <p>Otros(4)</p> |
| Suministro de Agua | <p>Acueducto (1)</p> <p>Pila pública (2)</p> <p>Camión (3)</p> <p>Otros medios (4)</p> |
| Servicio de Eliminación de Excretas | <p>Poceta a cloaca (1)</p> <p>Poceta a pozo séptico (2)</p> <p>Excusado a hoyo o letrina (3)</p> <p>No tiene poceta o Excusado (4)</p> |
| Tenencia | <p>Propia pagada totalmente (1)</p> <p>Propia pagándose (2)</p> <p>Alquilada (3)</p> <p>Alquilada parte de la vivienda (4)</p> <p>Cedida por razones de trabajo (5)</p> <p>Cedida por familiar o amigo (6)</p> <p>Tomada (7)</p> <p>Otra forma (8)</p> |

Anexo B

Gráficos de las salidas del análisis de conglomerados para la construcción de las variables condición y adecuación de la vivienda

Variable Condición de la vivienda

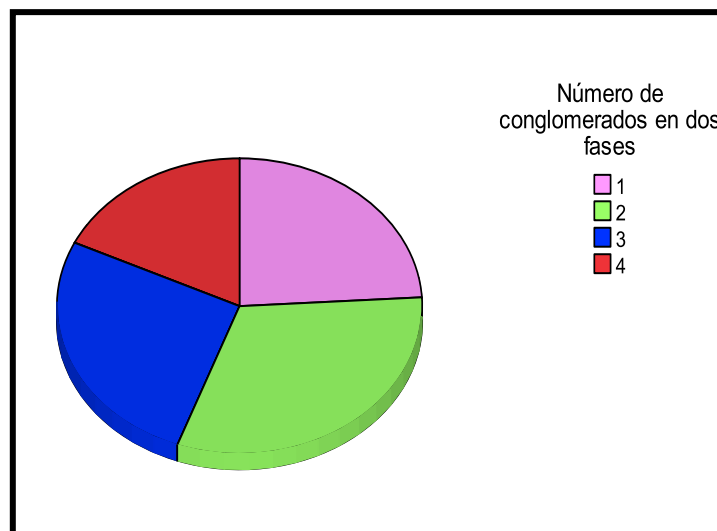


Figura b.1. Número de conglomerados para variable Condición de la vivienda.

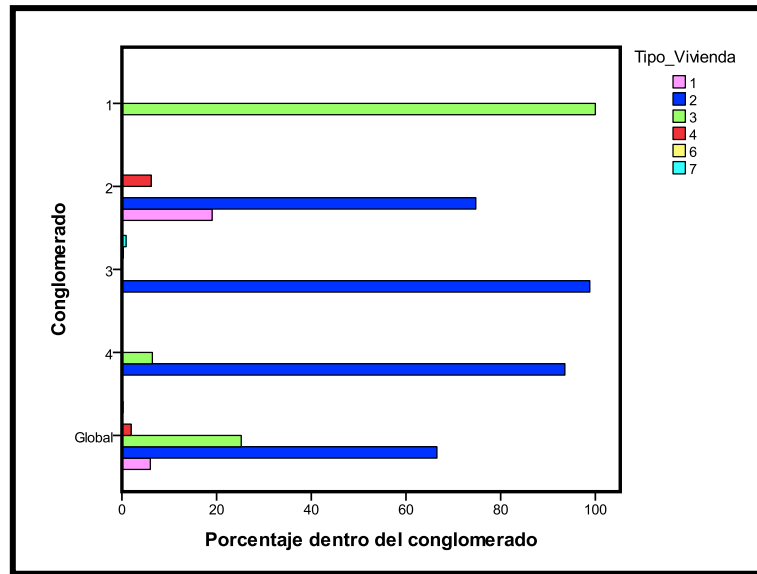


Figura b.2. Porcentaje dentro del conglomerado de la variable Tipo de vivienda.

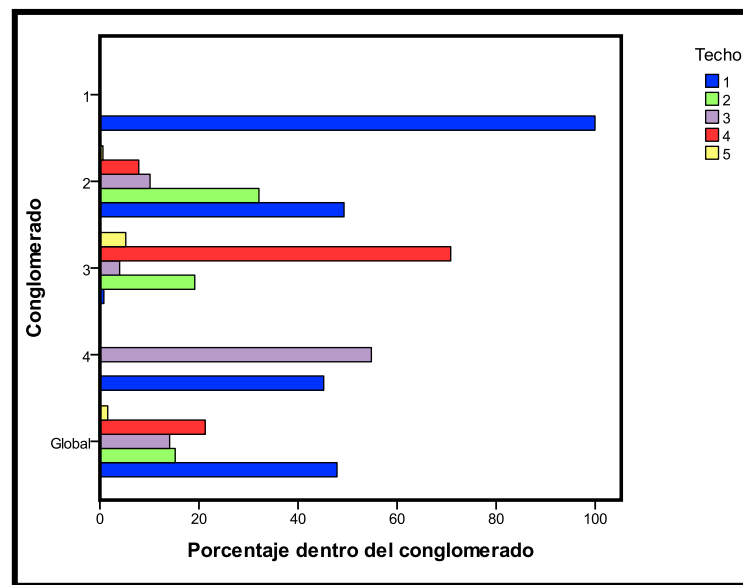


Figura b.3. Porcentaje dentro del conglomerado de la variable Techo.

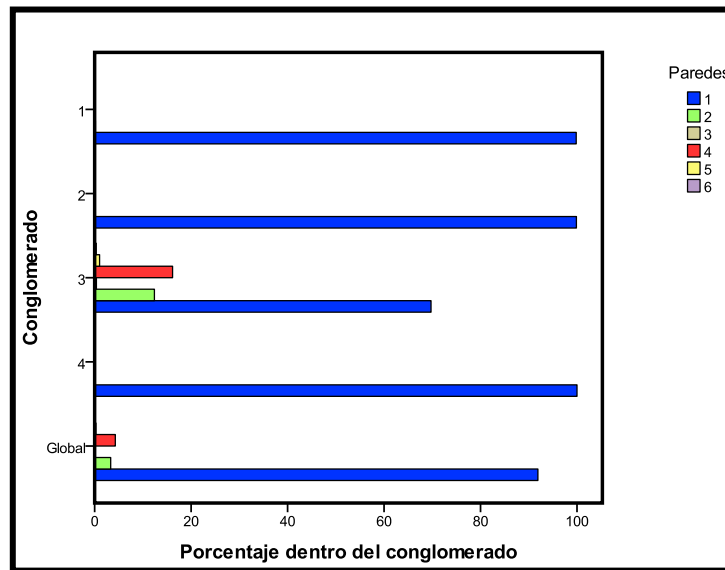


Figura b.4. Porcentaje dentro del conglomerado de la variable Paredes.

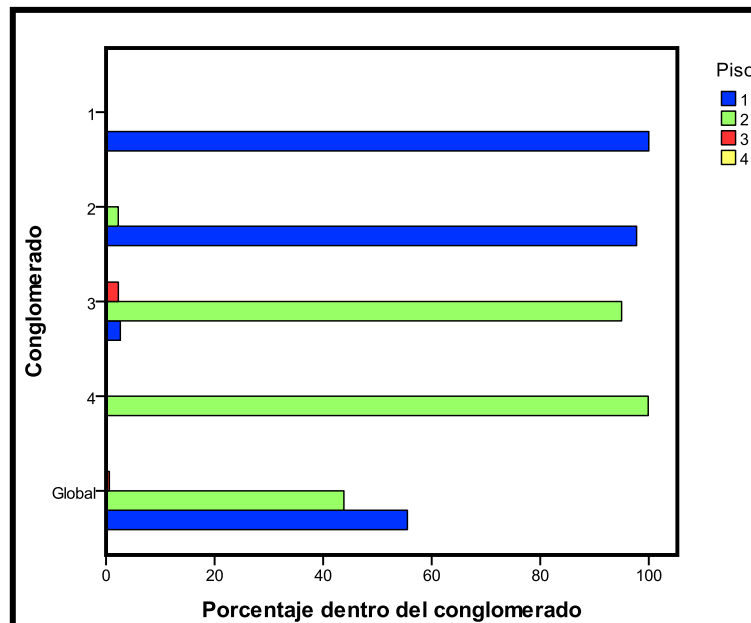


Figura b.5. Porcentaje dentro del conglomerado de la variable piso.

Variable Adecuación de la vivienda

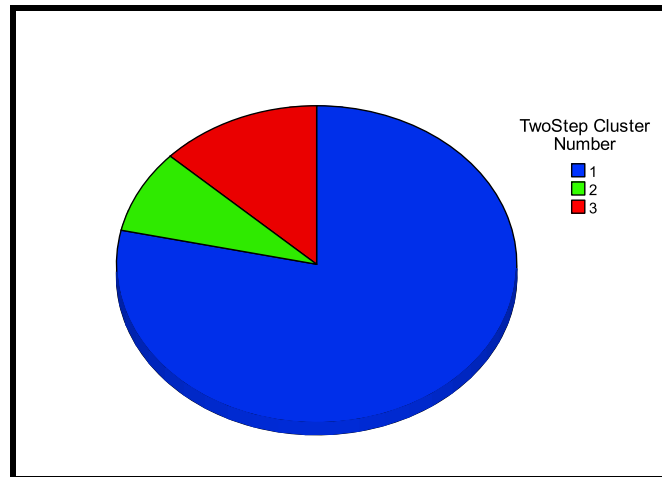


Figura b.6. Número de conglomerados para variable adecuación de la vivienda.

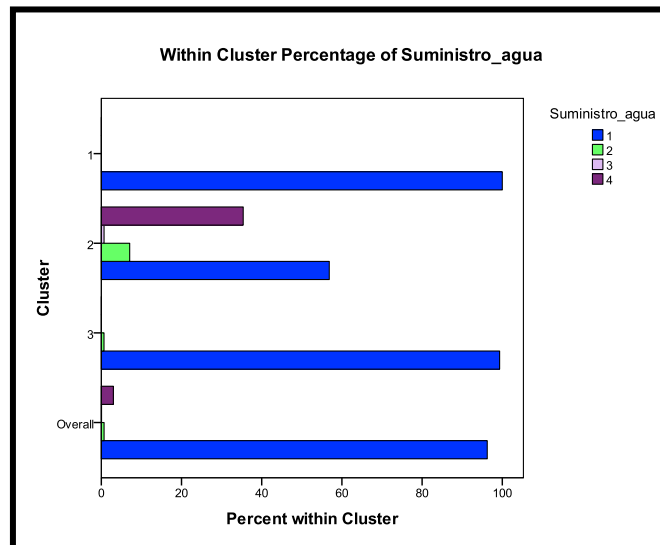


Figura b.7. Porcentaje dentro del conglomerado de la variable Suministro de agua.

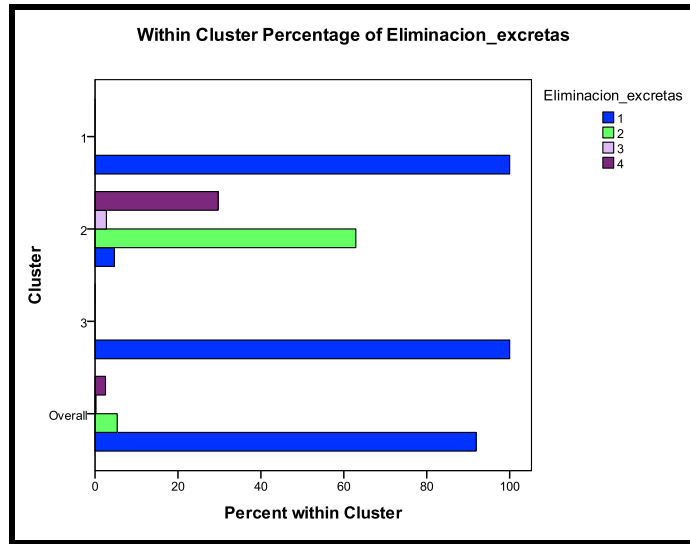


Figura b.8. Porcentaje dentro del conglomerado de la variable Eliminación de excretas.

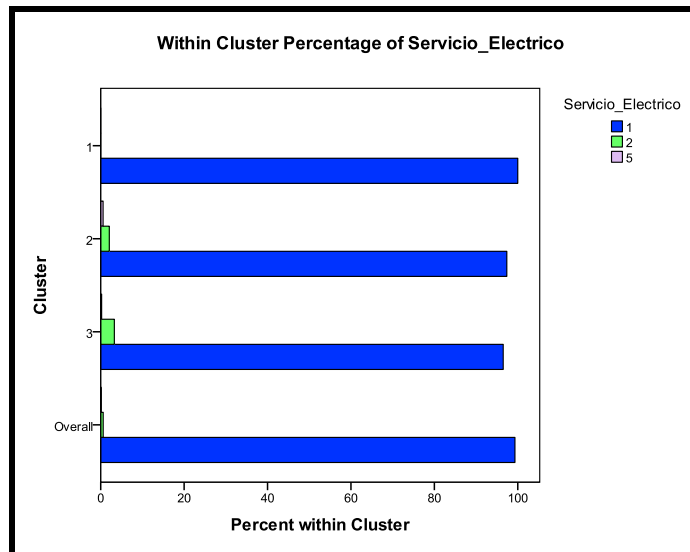


Figura b.9. Porcentaje dentro del conglomerado de la variable Servicio eléctrico.

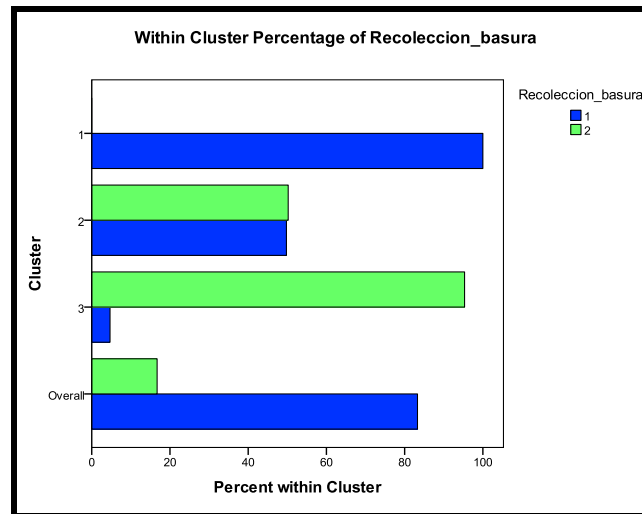


Figura b.10. Porcentaje dentro del conglomerado de la variable Recolección de basura.

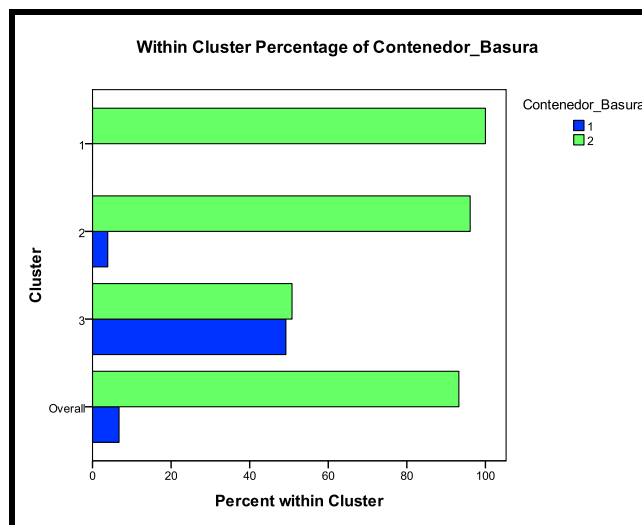


Figura b.11. Porcentaje dentro del conglomerado de la variable Contenedor de basura.

Anexo C

Modelos generados por la LIBSVM durante el proceso de clasificación, con C=35 y gamma=1.

| Semestre 1 | Descripción |
|--|---|
| svm_type c_svc | Tipo de clasificador |
| kernel_type rbf | Tipo de función kernel |
| gamma 1 | Valor gamma |
| nr_class 2 | Número de clases |
| total_sv 21 | Número total de vectores de soporte |
| label 2 1 | Etiquetas de las clases de salida |
| nr_sv 16 5 | Vectores de soporte, vectores de soporte en las fronteras |
| SV | |
| 0.3198675270371752 1:1 2:2 3:2 4:1 5:3 | |
| 0.3720377440148345 1:2 2:1 3:1 4:1 5:2 | |
| 0.9623413086027888 1:2 2:1 3:1 4:4 5:1 | |
| 0.9786598738041004 1:1 2:1 3:2 4:1 5:1 | |
| 0.9115093815382379 1:2 2:1 3:1 4:2 5:1 | |
| 0.7234499623196634 1:1 2:2 3:1 4:1 5:1 | |
| 0.2824498806051509 1:2 2:1 3:1 4:3 5:3 | |
| 0.0222530324947315 1:1 2:2 3:2 4:3 5:3 | |
| 0.9808438898247928 1:1 2:1 3:2 4:4 5:1 | |
| 0.23913142256825 1:1 2:2 3:1 4:3 5:3 | |
| 1.967750163912965 1:1 2:1 3:1 4:2 5:2 | |
| 0.692634914234237 1:1 2:2 3:1 4:2 5:1 | |
| 0.9315800026668667 1:1 2:1 3:1 4:4 5:2 | |
| 0.9665779056344316 1:1 2:2 3:1 4:4 5:1 | |
| 0.2707425397440874 1:1 2:2 3:2 4:2 5:3 | |
| 1.859902571229872 1:1 2:1 3:1 4:3 5:1 | |
| -1.371342852823306 1:1 2:1 3:1 4:1 5:3 | |
| -2.143717293711244 1:1 2:1 3:1 4:2 5:3 | |
| -1.75095146143674 1:1 2:1 3:1 4:1 5:1 | |
| -3.333005269823921 1:1 2:1 3:1 4:2 5:1 | |
| -3.882715242436973 1:1 2:1 3:1 4:4 5:1 | |

| Semestre 2 | Semestre 3 |
|--|---|
| svm_type c_svc | svm_type c_svc |
| kernel_type rbf | kernel_type rbf |
| gamma 1 | gamma 1 |
| nr_class 2 | nr_class 2 |
| total_sv 22 | total_sv 21 |
| label 2 1 | label 2 1 |
| nr_sv 17 5 | nr_sv 16 5 |
| SV | SV |
| 0.175810161492037 1:2 2:2 3:1 4:1 5:1 | 0.04545500192327775 1:2 2:2 3:1 4:3 5:1 |
| 0.7698273446381551 1:1 2:1 3:2 4:1 5:1 | 0.2247150507082715 1:1 2:1 3:2 4:3 5:3 |
| 0.2076418277159391 1:2 2:1 3:1 4:3 5:1 | 0.2040967060433376 1:1 2:2 3:1 4:3 5:3 |
| 1.057927476624309 1:1 2:2 3:1 4:1 5:3 | 1.87556427076952 1:1 2:1 3:1 4:1 5:2 |
| 0.1336589329358756 1:2 2:2 3:1 4:2 5:1 | 1.074342710895013 1:1 2:2 3:1 4:1 5:3 |
| 0.2112773469550036 1:1 2:2 3:2 4:4 5:1 | 1.078312241657183 1:1 2:1 3:2 4:4 5:1 |
| 0.9817176773776168 1:1 2:2 3:1 4:4 5:1 | 0.8132187700626483 1:1 2:1 3:2 4:1 5:1 |
| 0.5230523007200862 1:1 2:2 3:2 4:4 5:3 | 0.7957440923643113 1:1 2:2 3:1 4:1 5:1 |
| 0.130176467096118 1:1 2:1 3:2 4:3 5:3 | 1.072119821801634 1:1 2:2 3:1 4:4 5:1 |
| 1.055688799446111 1:1 2:1 3:2 4:1 5:3 | 0.7667831839323608 1:1 2:2 3:1 4:2 5:1 |
| 0.7121616844121381 1:1 2:2 3:1 4:1 5:1 | 0.2032644832131808 1:2 2:2 3:2 4:3 5:2 |
| 2.062900110445806 1:1 2:1 3:1 4:4 5:2 | 0.8632494988900766 1:1 2:1 3:1 4:3 5:3 |
| 1.201743306611369 1:1 2:1 3:1 4:2 5:2 | 0.6001729789023861 1:1 2:1 3:1 4:2 5:2 |
| 0.8670696737998174 1:1 2:1 3:1 4:3 5:3 | 0.7738974644607419 1:1 2:1 3:2 4:2 5:1 |
| 0.7156465667856992 1:1 2:2 3:1 4:2 5:1 | 2.081124521481095 1:1 2:1 3:1 4:4 5:2 |
| 0.7843833559091139 1:1 2:1 3:2 4:2 5:1 | 1.91513372711599 1:1 2:1 3:1 4:3 5:1 |
| 1.836995019962308 1:1 2:1 3:1 4:3 5:1 | -2.80343621004503 1:1 2:1 3:1 4:1 5:3 |
| -2.613552591566763 1:1 2:1 3:1 4:1 5:3 | -2.764710605543872 1:1 2:1 3:1 4:4 5:3 |
| -2.820172776901415 1:1 2:1 3:1 4:4 5:3 | -2.207978095778006 1:1 2:1 3:1 4:1 5:1 |
| -1.528414386392193 1:1 2:1 3:1 4:1 5:1 | -2.764753651019474 1:1 2:1 3:1 4:2 5:1 |
| -3.003579271326678 1:1 2:1 3:1 4:2 5:1 | -3.846315961834647 1:1 2:1 3:1 4:4 5:1 |
| -3.461959026740455 1:1 2:1 3:1 4:4 5:1 | |

| Semestre 4 | Semestre 5 |
|--|---|
| svm_type c_svc | svm_type c_svc |
| kernel_type rbf | kernel_type rbf |
| gamma 1 | gamma 1 |
| nr_class 2 | nr_class 2 |
| total_sv 21 | total_sv 24 |
| label 2 1 | label 2 1 |
| nr_sv 16 5 | nr_sv 18 6 |
| SV | SV |
| 0.3214759831606189 1:3 2:2 3:1 4:2 5:1 | 1.080763857661206 1:1 2:2 3:1 4:1 5:3 |
| 0.5392354900008125 1:1 2:2 3:1 4:1 5:3 | 1.509081748502462 1:1 2:1 3:1 4:2 5:2 |
| 0.7722275094435167 1:1 2:1 3:2 4:2 5:1 | 0.3018137901962546 1:2 2:2 3:1 4:2 5:1 |
| 0.8413013845451895 1:1 2:1 3:1 4:1 5:2 | 0.8428203375168679 1:1 2:1 3:2 4:2 5:1 |
| 0.806354445961116 1:1 2:1 3:2 4:1 5:1 | 0.0017066903784233 1:1 2:2 3:2 4:2 5:2 |
| 0.7957186747191656 1:1 2:2 3:1 4:1 5:1 | 1.574015768902747 1:1 2:1 3:1 4:1 5:2 |
| 0.1622409410531361 1:1 2:1 3:2 4:3 5:3 | 0.2003670501584243 1:2 2:2 3:1 4:3 5:3 |
| 1.074502571398661 1:1 2:1 3:2 4:4 5:1 | 0.2997308766940951 1:1 2:2 3:2 4:4 5:1 |
| 2.059108132735666 1:1 2:1 3:1 4:4 5:2 | 0.8516791569275839 1:1 2:1 3:2 4:1 5:1 |
| 1.074004691783389 1:1 2:2 3:1 4:4 5:1 | 0.8392190277247715 1:1 2:2 3:1 4:1 5:1 |
| 0.1749308582864112 1:1 2:2 3:2 4:3 5:3 | 0.4351437589311178 1:2 2:2 3:1 4:2 5:3 |
| 1.744184479716515 1:1 2:1 3:1 4:2 5:2 | 0.2997600096763455 1:1 2:1 3:2 4:3 5:3 |
| 0.1524480823606303 1:1 2:2 3:1 4:3 5:3 | 1.813032744361148 1:1 2:1 3:1 4:3 5:3 |
| 1.992875016218482 1:1 2:1 3:1 4:3 5:3 | 0.708333528062945 1:1 2:2 3:1 4:2 5:1 |
| 1.899722424565684 1:1 2:1 3:1 4:3 5:1 | 1.034400989872021 1:1 2:2 3:1 4:4 5:1 |
| 0.7667055425274024 1:1 2:2 3:1 4:2 5:1 | 1.950481191217116 1:1 2:1 3:1 4:3 5:1 |
| -3.123579264672776 1:1 2:1 3:1 4:4 5:3 | 2.099068787758692 1:1 2:1 3:1 4:4 5:2 |
| -3.198377359632088 1:1 2:1 3:1 4:2 5:3 | 0.03593975111200375 1:1 2:2 3:2 4:3 5:2 |
| -1.881273514212961 1:1 2:1 3:1 4:1 5:1 | -2.502065426328919 1:1 2:1 3:1 4:2 5:3 |
| -3.133807685913018 1:1 2:1 3:1 4:2 5:1 | -1.849970100063437 1:1 2:1 3:1 4:1 5:3 |
| -3.839998417695593 1:1 2:1 3:1 4:4 5:1 | -2.073980739877838 1:1 2:1 3:1 4:1 5:1 |
| | -2.987313532235512 1:1 2:1 3:1 4:4 5:3 |
| | -3.055194972649552 1:1 2:1 3:1 4:2 5:1 |
| | -3.408834294498967 1:1 2:1 3:1 4:4 5:1 |

| Semestre 6 | Semestre 7 |
|---|---|
| svm_type c_svc | svm_type c_svc |
| kernel_type rbf | kernel_type rbf |
| gamma 1 | gamma 1 |
| nr_class 2 | nr_class 2 |
| total_sv 24 | total_sv 23 |
| label 1 2 | label 1 2 |
| nr_sv 6 18 | nr_sv 6 17 |
| SV | SV |
| 1.399949409701203 1:1 2:1 3:1 4:1 5:3 | 1.612905712883849 1:1 2:1 3:1 4:1 5:1 |
| 2.71979124152173 1:1 2:1 3:1 4:2 5:3 | 1.394194151455336 1:1 2:1 3:1 4:1 5:3 |
| 1.579368379519085 1:1 2:1 3:1 4:1 5:1 | 3.325587945202127 1:1 2:1 3:1 4:2 5:3 |
| 3.474617601998719 1:1 2:1 3:1 4:4 5:3 | 3.847506352555094 1:1 2:1 3:1 4:4 5:3 |
| 3.626068396529052 1:1 2:1 3:1 4:2 5:1 | 3.600130749935639 1:1 2:1 3:1 4:2 5:1 |
| 4.215442589329681 1:1 2:1 3:1 4:4 5:1 | 3.847754739178762 1:1 2:1 3:1 4:4 5:1 |
| -1.009007765842148 1:2 2:1 3:1 4:4 5:1 | -0.7038242267619899 1:2 2:1 3:1 4:1 5:1 |
| -1.003793378905879 1:2 2:1 3:1 4:2 5:1 | -0.7056142884290914 1:2 2:1 3:1 4:2 5:1 |
| -0.0414449092884096 1:2 2:2 3:1 4:1 5:1 | -0.9308127853640877 1:1 2:1 3:2 4:2 5:1 |
| -0.1686090050417121 1:1 2:2 3:2 4:1 5:3 | -0.8901950372545083 1:1 2:1 3:2 4:2 5:3 |
| -0.9468326632295331 1:1 2:2 3:1 4:1 5:3 | -0.7003271824797289 1:1 2:2 3:1 4:1 5:1 |
| -1.005301244692048 1:1 2:2 3:1 4:4 5:3 | -0.2991878297280763 1:2 2:1 3:2 4:1 5:3 |
| -0.026995770634072 1:1 2:2 3:2 4:3 5:3 | -0.6736041484733799 1:1 2:2 3:1 4:2 5:3 |
| -0.1537849852275993 1:1 2:1 3:2 4:3 5:3 | -0.6730617723376423 1:1 2:2 3:1 4:2 5:1 |
| -0.7584890091895609 1:1 2:1 3:2 4:1 5:1 | -0.930647467204076 1:1 2:1 3:2 4:4 5:1 |
| -0.7252588461226851 1:1 2:2 3:1 4:1 5:1 | -0.9311760727613063 1:1 2:1 3:2 4:4 5:3 |
| -2.022243532756359 1:1 2:1 3:1 4:2 5:2 | -0.6995301032854262 1:1 2:2 3:1 4:1 5:3 |
| -0.7287480059605259 1:1 2:1 3:2 4:2 5:1 | -0.9357206374871744 1:1 2:2 3:1 4:4 5:3 |
| -1.012991652113581 1:1 2:1 3:2 4:4 5:1 | -1.989934955954013 1:1 2:1 3:1 4:4 5:2 |
| -0.995672664920095 1:1 2:2 3:1 4:4 5:1 | -0.9357324322099748 1:1 2:2 3:1 4:4 5:1 |
| -1.86191594888467 1:1 2:1 3:1 4:3 5:1 | -1.99080236267562 1:1 2:1 3:1 4:2 5:2 |
| -0.73013754015308 1:1 2:2 3:1 4:2 5:1 | -1.819017805949845 1:1 2:1 3:1 4:3 5:1 |
| -1.801944477878915 1:1 2:1 3:1 4:3 5:3 | -1.818890542854865 1:1 2:1 3:1 4:3 5:3 |
| -2.022066217758586 1:1 2:1 3:1 4:4 5:2 | |

| Semestre 8 | Semestre 9 |
|---|---|
| svm_type c_svc | svm_type c_svc |
| kernel_type rbf | kernel_type rbf |
| gamma 1 | gamma 1 |
| nr_class 2 | nr_class 2 |
| total_sv 23 | total_sv 27 |
| label 1 2 | label 2 1 |
| nr_sv 6 17 | nr_sv 21 6 |
| SV | SV |
| 1.663238104013633 1:1 2:1 3:1 4:1 5:3 | 0.05072770369524138 1:2 2:2 3:1 4:2 5:3 |
| 1.56398528523995 1:1 2:1 3:1 4:1 5:1 | 1.978165296440689 1:1 2:1 3:1 4:2 5:2 |
| 3.266397284843816 1:1 2:1 3:1 4:2 5:1 | 0.05879176514271225 1:2 2:2 3:1 4:1 5:3 |
| 3.0997045753259 1:1 2:1 3:1 4:4 5:1 | 0.9219199357762163 1:1 2:1 3:2 4:4 5:3 |
| 2.993097667209908 1:1 2:1 3:1 4:2 5:3 | 0.0948175591750739 1:3 2:2 3:2 4:3 5:3 |
| 3.105728880532004 1:1 2:1 3:1 4:4 5:3 | 0.9213902599846089 1:2 2:1 3:1 4:4 5:1 |
| -1.029603946394619 1:1 2:2 3:1 4:1 5:3 | 0.08095720909020765 1:2 2:1 3:2 4:2 5:1 |
| -0.759934246220614 1:1 2:1 3:2 4:1 5:3 | 0.680586071750501 1:1 2:2 3:1 4:1 5:1 |
| -0.1978057752154971 1:2 2:1 3:2 4:1 5:1 | 0.9191295622265873 1:2 2:1 3:1 4:4 5:3 |
| -0.0884563906667551 1:1 2:1 3:2 4:4 5:2 | 0.07627664065552996 1:2 2:1 3:2 4:1 5:3 |
| -0.1978310543870026 1:2 2:1 3:1 4:3 5:3 | 0.9042540352015822 1:1 2:1 3:2 4:1 5:3 |
| -0.7322004591092069 1:1 2:1 3:2 4:2 5:3 | 0.6585079821930346 1:1 2:2 3:1 4:1 5:3 |
| -0.6880588012739811 1:1 2:1 3:2 4:1 5:1 | 1.977320737526922 1:1 2:1 3:1 4:4 5:2 |
| -0.7560932087402943 1:1 2:2 3:1 4:1 5:1 | 0.6555393885248122 1:1 2:2 3:1 4:2 5:1 |
| -1.804079251298794 1:1 2:1 3:1 4:3 5:3 | 0.8860986737215453 1:1 2:1 3:2 4:2 5:1 |
| -0.7314452351754458 1:1 2:1 3:2 4:2 5:1 | 0.9052327073125699 1:1 2:1 3:2 4:4 5:1 |
| -0.7447666658490036 1:1 2:2 3:1 4:2 5:1 | 1.801973295659521 1:1 2:1 3:1 4:3 5:1 |
| -0.9844886568455334 1:1 2:1 3:2 4:4 5:1 | 0.9094352141531135 1:1 2:2 3:1 4:4 5:1 |
| -1.016102527516134 1:1 2:2 3:1 4:4 5:1 | 0.9096356750710696 1:1 2:2 3:1 4:4 5:3 |
| -1.875937005580094 1:1 2:1 3:1 4:3 5:1 | 0.6366604754090397 1:1 2:2 3:1 4:2 5:3 |
| -1.03023607428486 1:1 2:2 3:1 4:4 5:3 | 1.801710764336129 1:1 2:1 3:1 4:3 5:3 |
| -2.070218756693158 1:1 2:1 3:1 4:2 5:2 | -1.374342574059825 1:1 2:1 3:1 4:1 5:1 |
| -0.9848937419142176 1:1 2:1 3:2 4:4 5:3 | -1.717220100180697 1:1 2:1 3:1 4:1 5:3 |
| | -3.005855902101215 1:1 2:1 3:1 4:2 5:3 |
| | -4.191954783083519 1:1 2:1 3:1 4:4 5:1 |
| | -3.342661445299994 1:1 2:1 3:1 4:2 5:1 |
| | -4.197096148321457 1:1 2:1 3:1 4:4 5:3 |

| Semestre 10 | Semestre 11 |
|--|---|
| svm_type c_svc | svm_type c_svc |
| kernel_type rbf | kernel_type rbf |
| gamma 1 | gamma 1 |
| nr_class 2 | nr_class 2 |
| total_sv 26 | total_sv 24 |
| label 2 1 | label 2 1 |
| nr_sv 20 6 | nr_sv 18 6 |
| SV | SV |
| 0.9291176307412816 1:1 2:1 3:2 4:1 5:3 | 0.745544296360616 1:1 2:1 3:2 4:1 5:1 |
| 0.0918833799923149 1:2 2:2 3:1 4:1 5:1 | 1.006670095249323 1:2 2:1 3:1 4:2 5:1 |
| 0.968826792694511 1:2 2:1 3:1 4:2 5:3 | 0.06490303951439104 1:2 2:2 3:2 4:4 5:3 |
| 0.6760305864266086 1:1 2:2 3:1 4:2 5:3 | 0.1494937171263079 1:2 2:1 3:2 4:2 5:3 |
| 0.9645714926493759 1:2 2:1 3:1 4:4 5:1 | 0.9631493365214647 1:1 2:1 3:2 4:4 5:3 |
| 0.7034102785638835 1:1 2:2 3:1 4:1 5:3 | 0.03749540895376208 1:2 2:2 3:2 4:3 5:2 |
| 0.131628726389321 1:3 2:2 3:2 4:3 5:1 | 0.6975734842180634 1:1 2:1 3:2 4:2 5:1 |
| 0.2599317912716703 1:1 2:1 3:2 4:2 5:2 | 0.9771763915380214 1:1 2:1 3:2 4:4 5:1 |
| 0.6704176083797604 1:1 2:2 3:1 4:1 5:1 | 0.968314013214583 1:1 2:2 3:1 4:4 5:3 |
| 0.9346849034191662 1:1 2:1 3:2 4:4 5:1 | 2.048599463853621 1:1 2:1 3:1 4:4 5:2 |
| 0.0136365410401087 1:3 2:1 3:2 4:4 5:1 | 1.848280095744259 1:1 2:1 3:1 4:3 5:3 |
| 2.027648777910029 1:1 2:1 3:1 4:4 5:2 | 0.9172327462804939 1:1 2:1 3:2 4:2 5:3 |
| 0.8504427827645902 1:1 2:1 3:2 4:2 5:1 | 0.7028231634132955 1:1 2:2 3:1 4:2 5:3 |
| 0.0652153953799512 1:2 2:2 3:1 4:2 5:1 | 1.848122995649529 1:1 2:1 3:1 4:3 5:1 |
| 0.9390825564838671 1:1 2:2 3:1 4:4 5:1 | 0.9771652616154874 1:1 2:2 3:1 4:4 5:1 |
| 0.6516709593831603 1:1 2:2 3:1 4:2 5:1 | 0.7023361271971792 1:1 2:2 3:1 4:2 5:1 |
| 1.822266494557656 1:1 2:1 3:1 4:3 5:1 | 0.7323780382401364 1:1 2:2 3:1 4:1 5:3 |
| 0.94042827878827 1:1 2:2 3:1 4:4 5:3 | 0.7326272457071791 1:1 2:2 3:1 4:1 5:1 |
| 0.9509373662017558 1:1 2:1 3:2 4:4 5:3 | -3.848820366842251 1:1 2:1 3:1 4:4 5:3 |
| 1.823215256668826 1:1 2:1 3:1 4:3 5:3 | -2.618158524926301 1:1 2:1 3:1 4:2 5:3 |
| -1.350406467474621 1:1 2:1 3:1 4:1 5:1 | -3.85414045241265 1:1 2:1 3:1 4:4 5:1 |
| -1.691629591806896 1:1 2:1 3:1 4:1 5:3 | -1.318440032459814 1:1 2:1 3:1 4:1 5:3 |
| -4.214403910372681 1:1 2:1 3:1 4:4 5:1 | -2.887409648244747 1:1 2:1 3:1 4:2 5:1 |
| -2.624420739974211 1:1 2:1 3:1 4:2 5:1 | -1.59291589551195 1:1 2:1 3:1 4:1 5:1 |
| -3.866031715026503 1:1 2:1 3:1 4:4 5:3 | |
| -2.668155175051195 1:1 2:1 3:1 4:2 5:3 | |

| Semestre 12 | Semestre 13 |
|--|---|
| svm_type c_svc | svm_type c_svc |
| kernel_type rbf | kernel_type rbf |
| gamma 1 | gamma 1 |
| nr_class 2 | nr_class 2 |
| total_sv 26 | total_sv 25 |
| label 1 2 | label 1 2 |
| nr_sv 6 20 | nr_sv 6 19 |
| SV | SV |
| 2.99821960076972 1:1 2:1 3:1 4:2 5:3 | 3.48724013089763 1:1 2:1 3:1 4:4 5:3 |
| 1.301358379718208 1:1 2:1 3:1 4:1 5:3 | 2.631984426418589 1:1 2:1 3:1 4:2 5:3 |
| 3.480321479040788 1:1 2:1 3:1 4:4 5:1 | 1.67922816952529 1:1 2:1 3:1 4:1 5:3 |
| 2.996379184919938 1:1 2:1 3:1 4:2 5:1 | 4.22940304742375 1:1 2:1 3:1 4:4 5:1 |
| 1.302342021075352 1:1 2:1 3:1 4:1 5:1 | 2.618709314721825 1:1 2:1 3:1 4:2 5:1 |
| 3.48449805640297 1:1 2:1 3:1 4:4 5:3 | 1.307559502610248 1:1 2:1 3:1 4:1 5:1 |
| -1.010146230229025 1:2 2:1 3:1 4:4 5:3 | -0.0291710671620717 1:2 2:2 3:2 4:2 5:1 |
| -1.00887276192478 1:1 2:1 3:2 4:4 5:1 | -1.017763949826565 1:2 2:1 3:1 4:4 5:1 |
| -0.101310496786511 1:2 2:2 3:2 4:2 5:1 | -0.1464835035192336 1:2 2:2 3:1 4:2 5:1 |
| -0.227069943263793 1:4 2:1 3:2 4:4 5:1 | -1.020385070222082 1:2 2:1 3:1 4:2 5:3 |
| -2.06043275854052 1:1 2:1 3:1 4:2 5:2 | -1.011762982659605 1:1 2:1 3:2 4:1 5:3 |
| -0.997383224823114 1:1 2:2 3:1 4:4 5:3 | -0.0258660018055099 1:1 2:2 3:2 4:1 5:1 |
| -1.010306519183384 1:1 2:1 3:2 4:4 5:3 | -2.053469004419921 1:1 2:1 3:1 4:4 5:2 |
| -0.129939142711903 1:1 2:2 3:2 4:1 5:1 | -0.1379062813853132 1:1 2:1 3:2 4:3 5:3 |
| -0.174209670534174 1:1 2:2 3:2 4:1 5:3 | -0.0444087082814369 1:2 2:2 3:2 4:4 5:2 |
| -0.718125801979068 1:1 2:2 3:1 4:2 5:3 | -0.7130549224069214 1:1 2:2 3:1 4:2 5:3 |
| -0.699207777317791 1:1 2:2 3:1 4:1 5:1 | -0.9874205971568839 1:1 2:2 3:1 4:4 5:3 |
| -0.681915878681155 1:1 2:2 3:1 4:1 5:3 | -0.7407923103304545 1:1 2:2 3:1 4:1 5:3 |
| -0.996111489620397 1:1 2:2 3:1 4:4 5:1 | -0.6542526357601732 1:1 2:2 3:1 4:2 5:1 |
| -1.003581219410178 1:2 2:1 3:1 4:4 5:1 | -0.9878141468630558 1:1 2:1 3:2 4:2 5:1 |
| -0.123103583276358 1:2 2:2 3:1 4:2 5:1 | -0.9878784764461254 1:1 2:2 3:1 4:4 5:1 |
| -0.626735730136752 1:1 2:2 3:1 4:2 5:1 | -0.7314706868088479 1:1 2:2 3:1 4:1 5:1 |
| -0.186589733884833 1:3 2:2 3:2 4:3 5:1 | -1.806823067868966 1:1 2:1 3:1 4:3 5:3 |
| -0.083410818999727 1:1 2:2 3:2 4:2 5:1 | -1.857350402435148 1:1 2:1 3:1 4:3 5:1 |
| -1.861396831327652 1:1 2:1 3:1 4:3 5:1 | -1.000050776239019 1:1 2:1 3:2 4:4 5:1 |
| -1.863269109295855 1:1 2:1 3:1 4:3 5:3 | |

Anexo D

Información atributiva necesaria para la representación espacial

| Campo | Descripción |
|------------------|--|
| Entidad | Código de entidad federal. |
| Municipio | Municipio perteneciente al Área Metropolitana donde se levanta le encuesta. |
| Parroquia | Parroquia perteneciente al Área Metropolitana donde se levanta le encuesta. |
| Localidad | Centro poblado perteneciente al Área Metropolitana donde se levanta le encuesta. |
| Control | Identificador del segmento encuestado. |
| Segmento | Área geográfica que abarca entre 200 y 250 viviendas (área urbana). |
| Sector | Área geográfica que abarca entre 180 y 200 viviendas (área rural). |

Archivo del código de configuración del visor de mapas

```
#===== Inicio del Archivo Mapa=====
```

```
MAP
EXTENT 179068.120191 848614.674201
329863.666075 1017383.835149
UNITS meters
SIZE 600 500

#IMAGECOLOR 230 255 255
IMAGECOLOR 172 234 255
SHAPEPATH ".../demodata/Layers_Tesis"
SYMBOLSET ".../etc/symbols.sym"
FONTSET ".../etc/fonts.txt"

RESOLUTION 96
IMAGETYPE png
INTERLACE OFF
CONFIG "PROJ_LIB" "/ms4w/proj/nad/"
```

Características generales:

Extent: coordenadas x y max y min, que definen la extensión del mapa.

Shapepath: directorio local de datos espaciales.

Units: unidades de medida.

Resolution: resolución en % de la calidad de la imagen del mapa

```
PROJECTION
  "init=epsg:23030"
END
```

Projection: Se define la proyección del mapa.

```
OUTPUTFORMAT
NAME png
DRIVER "GD/PNG"
MIMETYPE "image/png"
IMAGEMODE RGB
FORMATOPTION INTERLACE=OFF
TRANSPARENT OFF
END
```

Output format: Formato de salida de la imagen.

```
#===== Definición de la Interfaz Web=====
```

```
WEB
  TEMPLATE "map.html"
  IMAGEPATH "/ms4w/tmp/ms_tmp/"
  IMAGEURL "/ms_tmp/"
```

WEB: se define el lugar donde se almacenarán temporalmente las imágenes generadas por MapServer.

```
METADATA
```

```
  # "MAPFILE_ENCODING" "ISO-8859-1"
```

```
END # Metadata
```

```
END # Web
```

```
#=====Definición del mapa de referencia=====
```

```
REFERENCE
```

```
  EXTENT 179068.120191 848614.674201
        329863.666075 1017383.835149
  IMAGE "../..//images/referencel.gif"
  SIZE 200 150
  COLOR -1 -1 -1
  OUTLINECOLOR 255 0 0
  STATUS on
  END # Reference
```

Reference: Parámetros de la imagen del mapa de referencia.

```
#===== Definición de la Barra de escala=====
```

```
SCALEBAR
```

```
  STATUS embed
  TRANSPARENT on
  INTERVALS 4
  SIZE 200 3
  UNITS kilometers
  COLOR 250 250 250
  OUTLINECOLOR 0 0 0
  BACKGROUNDCOLOR 100 100 100
  STYLE 0
  POSTLABELCACHE true
  LABEL
    COLOR 0 0 90
    #OUTLINECOLOR 200 200 200
    SIZE small
  END # Label
END # Reference
```

Scalebar: Se corresponde con la escala gráfica, se definen los intervalos, unidad de medida, y aspectos de diseño visual

```

# Simbolos usados en Pmapper
#
Symbol
  Name 'circle'
  Type ELLIPSE
  Filled TRUE

Points
  1 1
  END
END

```

Symbol: Se definen los símbolos que se utilizarán para representar los objetos.

```

#===== Inicio de la sección de capas =====#

LAYER
  NAME "Estados"
  STATUS default
  TYPE polygon
  DATA "Estados"

LABELITEM "Nb_estado"

METADATA

  'wms_srs' 'EPSG:23030 EPSG:4326'
  "DESCRIPTION" "Estados"
  "RESULT_FIELDS" "Nb_estado,Nb_capital"
  "RESULT_HEADERS" "Estado,Capital"
  #"RESULT_HYPERLINK" "Nb_estado, Nb_capital"
END

CLASS
  STYLE
    COLOR 255 235 189
    OUTLINECOLOR 255 206 93

  END

  TEMPLATE void

  LABEL

    POSITION Auto
    COLOR 0 0 0
    #BACKGROUNDCOLOR 255 255 255
    BACKGROUNDSHADOWCOLOR 150 150 150
    BACKGROUNDSHADOWSIZE 4 4

```

Definición de la representación de la capa en el mapa.


```
        OUTLINECOLOR 255 255 255
        ANTIALIAS TRUE
        TYPE truetype
        FONT arial
        SIZE 10
    END
END
LAYER
    NAME "Parroquias"
    STATUS default
    TYPE polygon
    DATA "Parroquias"

    METADATA
        'wms_srs' 'EPSG:23030 EPSG:4326'
        "DESCRIPTION" "Parroquias"
        "RESULT_FIELDS" "MUNICIPIO,PARROQUIA"
        "RESULT_HEADERS" "Municipio,Parroquia"
    END

    CLASS
        STYLE
            #COLOR 226 255 215
            COLOR 209 255 196
            OUTLINECOLOR 170 232 153
        END

        TEMPLATE void

        LABEL
            TYPE TRUETYPE
            FONT vera
            SIZE 6
            PARTIALS FALSE
            COLOR 0 0 0
        END
    END
END

LAYER
    NAME "AMM"
    STATUS default
    TYPE polygon
    DATA "AreaMM"
    LABELITEM "PARROQUIA"

    METADATA
```

```
'wms_srs' 'EPSG:23030 EPSG:4326'
"DESCRIPTION" "Area Metropolitana"
"RESULT_FIELDS" "PARROQUIA"
"RESULT_HEADERS" "PARROQUIA"
#"RESULT_HYPERLINK" "Nb_estado, Nb_capital"

END

CLASS
  STYLE
    #COLOR 226 255 215
    COLOR 194 255 165
    OUTLINECOLOR 99 197 71
  END
  LABEL
    TYPE TRUETYPE
    FONT CALIBRI
    SIZE 8
    PARTIALS FALSE
    COLOR 0 0 0
    POSITION auto
  END
END

END

LAYER

NAME "Controles"

METADATA
'wms_srs' 'EPSG:23030 EPSG:4326'
"DESCRIPTION" "Controles"
"RESULT_FIELDS" "Control,LOCALIDAD,COD_SEG"
"RESULT_HEADERS" "Cod. Control,Centro Poblado,Segmento"
"RESULT_HYPERLINK" "Indicador1|| Nro Hogares,
Indicador2|| Tenencia, Indicador3|| Hacinamiento,
Indicador4|| Condicion, Indicador5|| Adecuacion,
Salida|| Salida"

END

STATUS default
TYPE POLYGON
DATA "Controles"
LABELITEM "Control"

CLASS
```

```
NAME "Campo Elias"
EXPRESSION ([COD_MU] = 06)
STYLE
  COLOR 253 84 82
  OUTLINECOLOR 255 0 0
END

TEMPLATE void

LABEL
  POSITION Auto
  PARTIALS FALSE
  COLOR 0 0 0
  TYPE TRUETYPE
  FONT arial
  SIZE 7
  BACKGROUNDCOLOR 255 244 202
  BACKGROUNDSHADOWCOLOR 150 150 150

END

END

CLASS
  NAME "Libertador"
  EXPRESSION ([COD_MU] = 12)
  STYLE
    COLOR 255 200 10
    OUTLINECOLOR 252 139 0
  END

  TEMPLATE void

  LABEL
    POSITION Auto
    SIZE SMALL
    PARTIALS FALSE
    COLOR 0 0 0
    TYPE TRUETYPE

    FONT arial
    SIZE 7
    BACKGROUNDCOLOR 255 244 202
    BACKGROUNDSHADOWCOLOR 150 150 150
  END

END

CLASS
  NAME "Sucre"
```

```
        EXPRESSION ([COD_MU] = 20)
        STYLE
            COLOR 0 80 255
            OUTLINECOLOR 0 0 200
        END

    TEMPLATE void
    LABEL
        POSITION Auto
        SIZE SMALL
        PARTIALS FALSE
        COLOR 0 0 0
        TYPE TRUETYPE
        FONT arial
        SIZE 7
        BACKGROUNDCOLOR 255 244 202
        BACKGROUNDSHADOWCOLOR 150 150 150

    END
END

CLASS
    NAME "Santos Marquina"
    EXPRESSION ([COD_MU] = 19)
    STYLE
        COLOR 191 55 167
        OUTLINECOLOR 133 41 117
    END

    TEMPLATE void

    LABEL

        POSITION Auto
        SIZE SMALL
        PARTIALS FALSE
        COLOR 0 0 0
        TYPE TRUETYPE
        FONT arial
        SIZE 7
        BACKGROUNDCOLOR 255 244 202
        BACKGROUNDSHADOWCOLOR 150 150 150

    END
END
END

LAYER
    NAME "Segmentos-AMM"
```

```
METADATA
    "wms_srs" "EPSG:23030 EPSG:4326"
    "DESCRIPTION" "Segmentos AMM"

END

STATUS DEFAULT
TYPE LINE
DATA "Segmentos-AMM"

CLASS
    STYLE
        COLOR 221 189 128
        #OUTLINECOLOR 0 0 0
    END
    LABEL
        SIZE SMALL

        PARTIALS FALSE
        COLOR 0 0 0
    END
END

LAYER

NAME "Vialidad"
METADATA
    "wms_srs" "EPSG:23030 EPSG:4326"
END

STATUS on
TYPE LINE
DATA "vialidad"

CLASS
    STYLE
        COLOR 255 0 0
    END
    LABEL
        SIZE SMALL
        PARTIALS FALSE
        COLOR 0 0 0
    END
END
```

```
LAYER

  NAME "Estadísticas"
  METADATA
    'wms_srs' 'EPSG:23030 EPSG:4326'
    "DESCRIPTION" "Estadísticas"
    "RESULT_FIELDS" "Indicador1,Indicador2,Indicador3,
      Indicador4,Indicador5, Salida"
    "RESULT_HEADERS" "Indicador1, Indicador2,
      Indicador 3,Indicador4,Indicador5, Salida"
    "RESULT_HYPERLINK" "Indicador1||
      Nro Hogares, Indicador2|| Tenencia,
      Indicador3|| Hacinamiento, Indicador4||
      Condicion, Indicador5|| Adecuacion,
      Salida|| Salida"
  END

  STATUS on
  TRANSPARENCY 0
  TYPE polygon
  DATA "Estadísticas"

  CLASS
    STYLE
      COLOR 255 255 255
    END

  TEMPLATE void

    LABEL
      SIZE SMALL
      PARTIALS FALSE
      COLOR 0 0 0
    END
  END
END

END #Fin del archivo Mapa
```

Anexo E

Interfaz gráfica del visor web

Se procederá a describir cada uno de los elementos que conforman la interfaz gráfica del visor web de mapas, con el propósito de ilustrar de manera más detallada el funcionamiento y utilidad de la misma, para esto se dispone de la figura e.1 donde numeran cada una de las herramientas a fin de que el lector tenga una mejor comprensión.

1.- **La ventana del mapa**, es la parte principal de la vista, y como su nombre indica está dedicada a mostrar el mapa resultante de las capas de datos activadas por defecto o el resultado de las búsquedas hechas en los distintos niveles.

2.- **Las coordenadas**, representan la posición del cursor en el sistema de referencia, en este caso UTM huso 19 N, correspondiente al área geográfica Venezuela.

3.- **La escala de visualización**. En esta sección existen dos herramientas relacionadas con la vista: una en la que se muestra la escala numérica de la vista actual y otra es una barra con un selector que permite aumentar o disminuir el nivel de la vista (zoom).

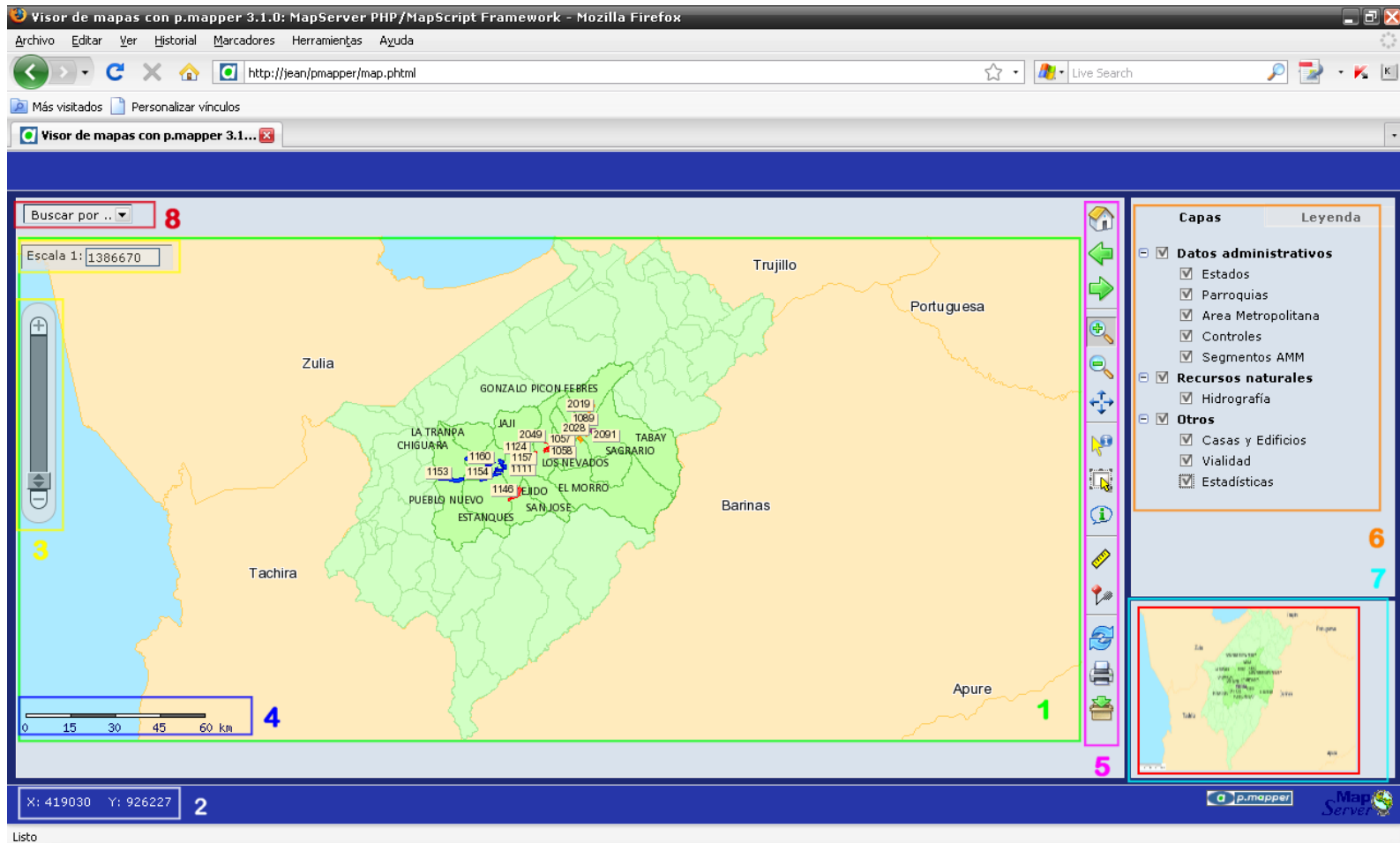


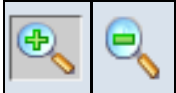











Figura e.1. Interfaz gráfica del visor web.

4.- **La Escala gráfica.** Es generada automáticamente por la aplicación, de acuerdo a la escala de visualización del mapa. De esta manera al realizar un zoom, no sólo se actualizará la vista del mapa, sino también la escala gráfica. El tipo de escala, unidades y estilo son definidos en el archivo de configuración (.map).

5.- **La barra de herramientas,** esta es la que permite interactuar con la aplicación. Pulsando sobre cualquier icono se selecciona la herramienta y puede operar sobre la ventana del mapa por medio del ratón. La activación de algunas herramientas posibilita la visualización de otros menús o ventanas donde se pueden seguir seleccionando opciones. En la tabla siguiente se describen cada una de estas herramientas.

| Herramienta | Descripción |
|---|--|
|  | Permite la visualización de toda la extensión del área de visión. |
|  | Permite avanzar o retroceder en la pantalla de la ventana del mapa principal. |
|  | <p>Permite hacer zoom más o menos.</p> <p>Después de haber seleccionado la herramienta zoom, se puede seleccionar un área para agrandar desplazando el ratón mientras se mantiene pulsado el botón derecho del mismo.</p> <p>Para definir la escala y poder acceder directamente a un valor determinado numérico, se puede pulsar sobre la ventana de la escala numérica (arriba derecha de la ventana del mapa principal)</p> |

| | |
|---|---|
| | <p>e introducir el valor numérico y automáticamente el mapa adoptará dicha escala.</p> <p>También al pulsar sobre la ventana de la escala se pueden seleccionar cualquiera de las escalas predefinidas que aparecen.</p> <p>La escala se puede modificar también desplazando la barra de zoom (centro derecha de la ventana del mapa principal)</p> <p>Si se dispone de un ratón con para avance, mediante el movimiento de la rueda se puede aumentar o disminuir el zoom.</p> |
|  | <p>Permite desplazar el mapa principal. También se puede desplazar el mapa, pulsando sobre el mapa de referencia.</p> |
|  | <p>Seleccionado y haciendo clic sobre un objeto del mapa nos devuelve los atributos alfanuméricos asociados a la posición geográfica donde se haga clic, abriendo una ventana dinámica con los atributos disponibles.</p> |
|  | <p>Tras seleccionar esta herramienta se abre la ventana de opción para elegir sobre que capa se desea actuar. Una vez elegida, se puede crear un área de selección de objeto moviendo el ratón mientras se mantiene pulsado el botón derecho y todos los objetos que se encuentren en esta área se seleccionaran. Se mostrará como en el caso anterior los datos alfanuméricos asociados a los objetos seleccionados en una ventana dinámica.</p> |
|  | <p>Herramienta MapTip. Mientras permanece pulsada se puede seleccionar sobre la capa a actuar (abajo derecha del mapa) y al pasar el cursor sobre los objetos de la capa nos mostraran los atributos asociados en una ventana dinámica en la esquina superior derecha de la ventana del mapa.</p> |

| | |
|---|---|
|  | <p>Herramienta para realizar medidas lineales o cálculo de áreas. Al estar activada la herramienta se visualizan dos ventanas en el extremo inferior izquierdo del mapa donde se pueden visualizar las distancias líneas a lo largo de los puntos marcados con el cursor y en caso que se realice una figura cerrada nos dará también el área del polígono. Las unidades de medida en este caso están en km.</p> <p>El icono del puntero mientras está activa esta herramienta será un icono representando una regla.</p> |
|  | <p>Activando esta opción podemos pulsar sobre un punto del mapa y se abrirá una ventana que nos permite introducir un comentario que quedará asociado a ese punto geográfico.</p> |
|  | <p>Seleccionando esta opción se desactivan las selecciones existentes, volviendo a recargarse el mapa sin objetos seleccionados.</p> |
|  | <p>Permite imprimir el mapa que está desplegado en la Ventana principal. Al pulsar esta opción se abre la ventana de impresión, donde podemos configurar las opciones de impresión.</p> |
|  | <p>Mediante esta opción podemos descargar la imagen que aparece en la ventana del mapa. Se pueden elegir varias resoluciones de descarga (a mayor resolución elegida mayor tiempo de descarga).</p> |

6.- **La Tabla de Contenidos y Leyenda**, muestra las capas de información geográfica que componen la aplicación, permitiendo activar o desactivar la visualización del contenido de las mismas, adicionalmente dispone de una pestaña donde puede ser

visualizada la leyenda de los datos representados, como si ilustra en la figura siguiente.



Figura e.2. Detalle de la tabla de contenidos y leyenda.

Información descargada en el visor se reagrupó en tres categorías:

- a) Datos Administrativos (Estados, Parroquias, Área Metropolitana, Controles, Segmentos Área Metropolitana).
- b) Recursos naturales (Hidrografía).
- c) Otros (Casa y Edificios, Vialidad, Estadísticas).

Las capas se activan a través de botones de activación. No es necesario seleccionar “redibujar” pues lo hace en automático tras activar o desactivar alguna opción. Existe la posibilidad de realizar una selección múltiple de las capas de información para su visualización. La superposición de las capas se realizará con el orden en que se hayan definido las mismas en el archivo de configuración del servidor de mapas. La aplicación basada en *MapServer* genera automáticamente la simbología con la que

se representa cada capa de información, basándose en la configuración de estilos que se definió en el archivo “.map” de configuración.

7.- **Mapa interactivo de referencia.** Abajo a la derecha aparece una pequeña imagen general de la zona en la que se puede ver la posición relativa de la vista general del mapa. La zona visible del mapa está marcada por un rectángulo de bordes rojo, que es sustituido por una cruz cuando el área de visualización esta a una escala muy pequeña. También es posible realizando un clic en un determinado sector del mapa de referencia, generar automáticamente el mapa de dicha zona.

8.- **Herramienta de selección y búsqueda,** mediante un menú podemos seleccionar el tipo de objeto a buscar, e ir progresivamente definiendo la búsqueda. Una vez elegida la opción permite abrir una ventana donde se mostrarán los resultados coincidentes con la búsqueda, y la vista del objeto seleccionado en la ventana del mapa.

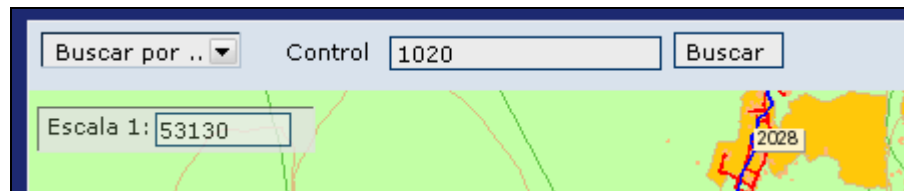


Figura e.3. Detalle de la herramienta de búsqueda.

Tras seleccionar un objeto se abrirá dinámicamente una ventana en la que aparecerán todos los objetos cuyos atributos coincidan con los parámetros seleccionados ver la figura siguiente.

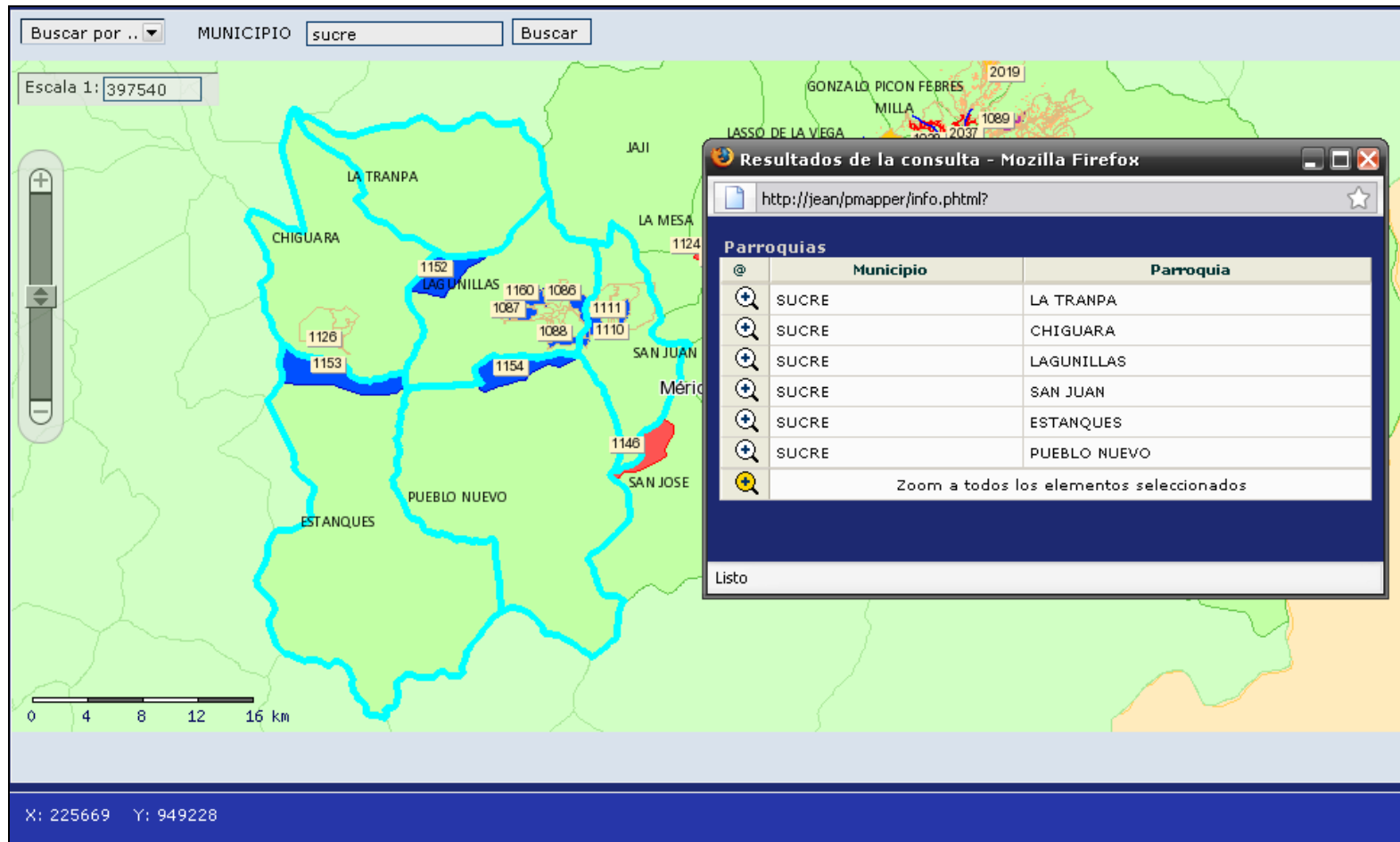


Figura e.4. Selección de los elementos tras la búsqueda realizada.

En caso de que varios objetos coincidan con los parámetros de búsqueda, se puede seleccionar uno en concreto pulsando sobre su icono de lupa. El resultado será que la imagen del mapa se modificará para mostrar ese objeto con el zoom predeterminado en el centro de la ventana del mapa.

A fin de que los resultados de las consultas sean más flexibles, la aplicación ofrece la posibilidad de crear hipervínculos para un campo seleccionado. De modo que es posible crear hipervínculos sobre los campos de una capa y mostrar alguna información complementaria como; imágenes, tablas, direcciones web, etc. Para el caso particular de esta aplicación, las gráficas con información estadística referente los indicadores de déficit de cada segmento, fueron generadas en la aplicación SIG creada en ArcGís Desktop 9.2, y son mostradas en el visor web a través de hipervínculos, es decir, pulsando sobre seleccionados campos estos desplegarán las gráficas con las estadísticas de cada segmento, como se ilustra en las figuras e.7 y e.8.

Resultados de la consulta - Mozilla Firefox

http://jean/pmapper/info.phtml?PHPSESSID=43417d7dfe32940f3917d8e4a315fb8

| | Estado | Capital |
|---|--------|---------|
| 🔍 | Mérida | Mérida |

Parroquias

| | Municipio | Parroquia |
|---|-----------|-----------|
| 🔍 | SUCRE | CHIGUARA |

Controles

| | Cod. Control | Centro Poblado | Segmento |
|---|--------------|----------------|----------|
| 🔍 | 1126 | Chiguara | 004 |

Estadísticas

| | Indicador 1 | Indicador 2 | Indicador 3 | Indicador 4 | Indicador 5 | Salida |
|---|-------------|-------------|--------------|-------------|-------------|--------|
| 🔍 | Nro Hogares | Tenencia | Hacinamiento | Condicion | Adecuacion | Salida |

javascript:openHyperlink('Estadísticas', 'Indicador2', '1126')

Figura e.5. Tabla de Atributos del segmento 1126.

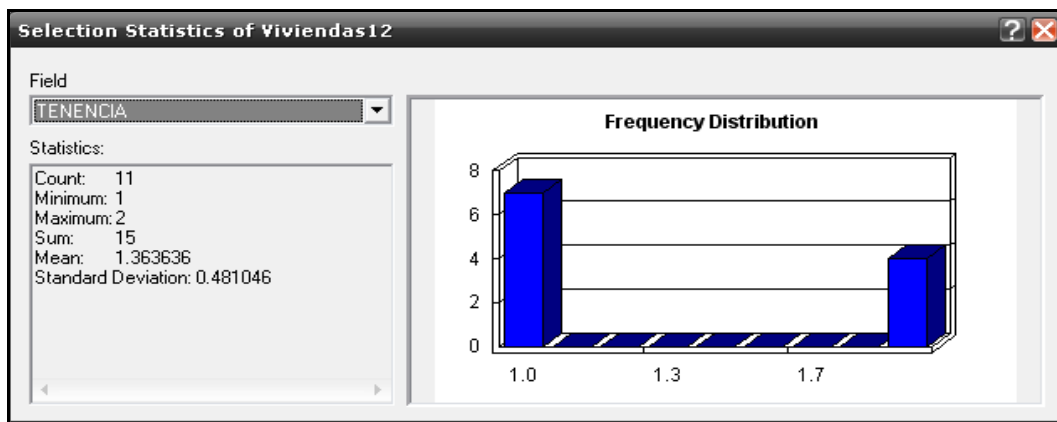


Figura e.6. Gráfica de las estadísticas del indicador Tenencia.

En la figura e.5 se puede observar que en la capa llamada estadísticas cada uno de sus campos posee un hipervínculo, cuya función es desplegar cada una de las estadísticas referentes a cada campo a través de gráficas, en la figura e.6 se ilustra las estadísticas del segmento 1126 con información referente a la tenencia de la vivienda.