



# UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

CENTRO DE INVESTIGACIÓN EN TECNOLOGÍAS DE INFORMACIÓN  
Y SISTEMAS

MAESTRÍA  
EN  
CIENCIAS COMPUTACIONALES

"HERRAMIENTA PARA LA ESTRATIFICACIÓN DE MUNICIPIOS EN  
ZONAS DE RIESGO PARA LA SALUD"

TESIS

Que para obtener el grado  
de Maestro en Ciencias Computacionales

Presenta

Lic. en Comp. Marco Antonio López Caviedes

Asesores  
Dr. Guillermo Sánchez Díaz  
M.S.P. Juan Ruiz Casañas

## RESUMEN

El programa "Atención Integral" de los Servicios de Salud de Hidalgo (SSH) está enfocado a reducir los niveles de enfermedad de las diferentes localidades que integran al Estado de Hidalgo, con un presupuesto autorizado limitado. Para realizar este trabajo, se determinaron cuáles de los 84 municipios que conforman el Estado de Hidalgo presentan los mayores índices de rezago en atención a la salud, carencia de recursos y marginalidad. Obteniendo una estructuración (agrupamiento) en cuatro zonas de riesgo.

Actualmente, los expertos que trabajan con el programa "Atención Integral" no cuentan con una herramienta computacional válida que permita la verificación y comparación de la primera estructuración generada con las nuevas estructuraciones. Además, requieren de dar seguimiento a dichas estructuraciones al paso de los años.

En este trabajo de tesis se propone una herramienta computacional que permita realizar estructuraciones de municipios en forma automatizada, para cada uno de los diferentes grupos de edad en el programa "Atención Integral". Brindándole al usuario la opción de obtener estratificaciones libres o restringidas a un cierto número de grupos. Esta herramienta también proporciona módulos para realizar un análisis visual y tabular de los resultados obtenidos en las diferentes estructuraciones generadas. Así como el de llevar un control de las estratificaciones generadas anteriormente para un posible análisis futuro de los resultados. Además, de verificar si los municipios del Estado de Hidalgo solo pueden estratificarse en 4 zonas de riesgo. Así como el de comparar esta estratificación en 4 zonas de riesgo de los SSH con las estratificaciones que se obtengan utilizando un algoritmo restringido a 4 clases.

La herramienta computacional propuesta se basa en técnicas del Reconocimiento Lógico Combinatorio de Patrones (RLCP), debido a las necesidades de los expertos y la naturaleza de los datos en estudio. Permitiendo manejar diversas formas en la presentación de resultados, en forma gráfica y tabular, así como brindar información acerca de qué indicadores influyen más en la obtención del promedio de riesgo total de cada uno de los grupos formados. Todo esto para el apoyo en la toma de decisiones. Por la gran importancia que tiene el factor salud en la población y en el país, esta herramienta fue diseñada de tal forma que pueda ser utilizada por otras instituciones de salud en México. Donde realicen estratificaciones de municipios en zonas de riesgo. Además, esta herramienta reducirá considerablemente el tiempo que actualmente los SSH requieren para obtener dicha estratificación. Así mismo, como el reducir los errores en el manejo de los datos y el número de personas necesarias que se encargan en cada fase de la estratificación actual de municipios.

# Contenido

<b>1</b>	<b>INTRODUCCIÓN</b>	<b>1</b>
1.1	Antecedentes	1
1.2	Planteamiento del problema	1
1.3	Objetivos	2
1.4	Metodología de trabajo	2
1.5	Justificación	4
1.6	Alcances y limitaciones	4
1.7	Estructura del trabajo	5
<b>2</b>	<b>MODELACIÓN MATEMÁTICA PARA LA ESTRATIFICACIÓN DE MUNICIPIOS</b>	<b>6</b>
2.1	Problemas de Reconocimiento de Patrones	6
2.2	Clasificación no supervisada	6
2.2.1	Descripciones de objetos	7
2.2.2	Criterios de comparación o analogía entre variables	7
2.2.3	Funciones de semejanza (entre objetos)	7
2.2.4	Criterios agrupacionales	8
2.3	Diseño y definición de los criterios de comparación	9
2.4	Obtención de la aportación informacional de los atributos o indicadores	13
2.5	Diseño y definición de la función de semejanza	13
2.6	Obtención del promedio de riesgo final por grupo	15
<b>3</b>	<b>TRABAJOS RELACIONADOS</b>	<b>19</b>
3.1	Procedimiento para obtener estratificaciones de municipios de los SSH....	19
3.2	Agrupamiento de datos	21
3.2.1	Algoritmos de agrupamiento en partición	21
3.2.2	Algoritmos de agrupamiento jerárquicos	22
3.2.3	Métodos y algoritmos basados en semejanzas	22
3.3	Sistemas que incorporan herramientas de Reconocimiento de Patrones	23
3.3.1	Sistema PROGNOSIS	23
3.3.2	Sistema GINKO	24
3.3.3	Sistema MCLUST	24
3.3.4	Sistema CLUTO	25
3.3.5	Sistema BANG	25
3.3.6	Programa Kminflu	26
3.3.7	Programa CLUSTER	26

<b>4</b>	<b>ANÁLISIS Y DISEÑO DEL SISTEMA DE AGRUPAMIENTO DE DATOS ( SAD )</b>	<b>28</b>
4.1	Antecedentes	28
4.1.1	Programa "Atención Integra"	28
4.1.2	Necesidad de la estratificación de municipios por los SSH	30
4.2	Análisis del SAD	30
4.2.1	Definición de requisitos funcionales	30
4.2.2	Definición de la base de datos	32
4.2.3	Especificaciones técnicas	37
4.3	Diseño del SAD	38
4.3.1	Análisis y carga de los datos	40
4.3.2	Preprocesamiento de los datos	42
4.4	Beneficios	43
<b>5</b>	<b>DESARROLLO DEL SAD</b>	<b>44</b>
5.1	Programación e implementación de los criterios agrupacionales	44
5.2	Programación del criterio propuesto conjuntos compactos con vecindad	47
5.3	Presentación de resultados de forma gráfica	49
5.4	Análisis gráfico de estratificaciones nuevas y anteriores	52
5.5	Análisis tabular de estratificaciones nuevas y anteriores	52
5.6	Consulta de estratificaciones almacenadas	55
<b>6</b>	<b>APLICACIÓN DEL SAD AL GRUPO MENOR DE 5 AÑOS</b>	<b>57</b>
6.1	Análisis del grupo de edad menor de 5 años	58
6.2	Estratificaciones obtenidas aplicando el SAD	61
6.2.1	Estratificación generada utilizando componentes conexas	62
6.2.2	Estratificación utilizando $\epsilon$ -means para uso de funciones de similaridad	66
6.2.3	Estratificación generada usando conjuntos compactos	70
6.2.4	Estratificación obtenida utilizando conjuntos compactos con vecindad	73
6.3	Análisis entre estratificaciones	78
6.3.1	Comparación de la estratificación actual y la obtenida con el criterio agrupacional componentes conexas	79
6.3.2	Comparación de la estratificación actual y la obtenida utilizando k-means para uso de funciones de similaridad	81
6.3.3	Comparación de la estratificación actual y la obtenida utilizando conjuntos compactos con vecindad	83
6.3.4	Comparación de la estratificación utilizando componentes conexas y la obtenida utilizando $\epsilon$ -means para funciones de similaridad	84
6.4	Resultados entre estratificaciones	87
	<b>CONCLUSIONES Y TRABAJO FUTURO</b>	<b>89</b>
	Conclusiones	89
	Trabajo futuro	90
	Publicaciones generadas	90
	<b>BIBLIOGRAFÍA</b>	<b>91</b>
	<b>ANEXOS</b>	

A	Estratificación de municipios en zonas de riesgo para la "Atención Integral".....	95
B	Indicadores de salud por grupo de edad .....	97
C	Ejemplo del procedimiento de estratificación actual de los SSH.....	99
D	Mortalidad por municipio para el grupo de edad del adolescente.....	101
E	Estratificación final de los SSH para el grupo menor de 5 años .....	103
F	Procedimiento de ponderación de los indicadores para el grupo de edad del adolescente de los SSH .....	105
	<b>GLOSARIO DE TÉRMINOS</b>	<b>106</b>

# Índice de figuras

1	Tipo de dato del indicador	9
2	Captura de criterios de comparación para cada indicador	10
3	Casos para la evaluación de aportación de riesgo de los indicadores	16
4	Influencia de los indicadores para el cálculo del riesgo	17
5	Reporte del análisis de aportaciones de riesgo por indicador	18
6	Estructura general del programa Atención Integral	29
7	Modelo Conceptual del SAD	31
8	Modelo relacional de la base de datos	37
9	Diagrama de flujo de datos del SAD	38
10	Diagrama a bloques del módulo de estratificación	39
11	Diagrama detallado a bloques del módulo de estratificación	39
12	Subconjunto de la base de datos en Excel para el grupo menor de 5 años	40
13	Acceso a los datos de los SSH	41
14	Módulo de alta para la base de datos en Access	41
15	Captura de fórmulas para el preprocesamiento de datos para cada indicador	42
16	Pantalla de agrupamientos utilizando los criterios agrupacionales	45
17	Análisis del grupo formado	46
18	Reporte del análisis del grupo y sus aportaciones de riesgo	46
19	Obtención de conjuntos compactos utilizando conjuntos compactos con vecindad	48
20	Obtención de componentes conexas utilizando conjuntos compactos con vecindad	48
21	Captura de coordenadas del mapa del Estado de Hidalgo	49
22	Apariencia del mapa cuando ha sido capturado por completo	50
23	Representación de la estratificación en el mapa de forma automática	50
24	Paleta de colores según las necesidades del experto	51
25	Representación de la estratificación en el mapa de forma manual	51
26	Análisis gráfico entre diferentes estratificaciones	52
27	Análisis tabular entre un par de estratificaciones cualesquiera	53
28	Número de municipios por grupo y estratificación	54
29	Reporte del análisis tabular de estratificaciones	54
30	Consulta de la estratificación que realizó los SSH en 1999 para el grupo de edad del menor de 5 años	55
31	Parte del reporte que generan la consultas	56
32	Mapa de la estratificación obtenida por los SSH en 1999 para el grupo de edad menor de 5 años	56
33	Resultados gráficos para el grupo de edad del menor de 5 años según	
65	componentes conexas	
34	Resultados gráficos para el grupo de edad del menor de 5 utilizando &-means	69
35	Resultados gráficos para el grupo de edad del menor de 5 años aplicando el	72
36	criterio conjuntos compactos	
37	Resultados gráficos para el grupo de edad del menor de 5 años aplicando el	77

	nuevo criterio conjuntos compactos con vecindad	
38	Análisis comparativo entre la estratificación actual (SSH) y la obtenida	
	37	
	utilizando componentes conexas	
39	Análisis comparativo entre la estratificación actual (SSH) y la obtenida	
	81	
	utilizando k-means para uso de funciones de similaridad	
40	Análisis comparativo entre la estratificación actual (SSH) y la obtenida	
	83	
	utilizando conjuntos compactos con vecindad	
41	Análisis comparativo entre la estratificación generada usando componentes conexas y la obtenida utilizando A:-means para funciones de similaridad	84
42	Análisis gráfico entre estratificaciones para el mismo grupo de edad	
	86	

# Índice de tablas

A	Mortalidad por municipio para el grupo de edad del adolescente	101
B	Mortalidad por municipio para el grupo de edad del adolescente continuación	102
C	Procedimiento de ponderación de los indicadores para el grupo de edad del adolescente de los SSH	105
1	Relación de municipios en coordenadas de píxeles	32
2	Relación de descripciones de municipios	33
3	Relación de fórmulas para cada indicador	33
4	Relación de criterios de comparación por indicador	33
5	Relación de estratificaciones almacenadas de municipios	34
6	Lista de tasas para el indicador Mortalidad por IRAS en menores de 1 año	34
7	Estratificación generada para el grupo menor de 5 años	35
8	Municipios pertenecientes al grupo 1	35
9	Lista de indicadores en función de la aportación de riesgo	36
10	Comparación entre dos estratificaciones	36
11	Datos para los indicadores del 1 - 5	57
12	Datos para los indicadores del 6-16	58
13	Análisis por indicador	59
14	Tasas, índices, porcentajes y casos de cada indicador para el grupo de edad "Menor de 5 años"	60
15	Subconjunto del reporte de tasas para el indicador Mortalidad por IRAS en < 1 año.	61
16	Reportes para el grupo 1 aplicando componentes conexas	63
17	Reportes para el grupo 2 aplicando componentes conexas	64
18	Reportes para el grupo 3 aplicando componentes conexas	64
19	Reportes para el grupo 4 aplicando componentes conexas	65
20	Reportes para el grupo 1 usando A:-means para funciones de similaridad	67
21	Reportes para el grupo 2 usando A:-means para funciones de similaridad	68
22	Reportes para el grupo 3 usando A:-means para funciones de similaridad	68
23	Reportes para el grupo 4 usando A;-means para funciones de similaridad	69
24	Reportes para el grupo 1 usando el criterio conjuntos compactos	71
25	Reportes para el grupo 2 usando el criterio conjuntos compactos	72
26	Reportes para el grupo 1 aplicando el nuevo criterio conjuntos compactos con vecindad	74
27	Reportes para el grupo 2 aplicando el nuevo criterio conjuntos compactos con vecindad	75
28	Reportes para el grupo 3 aplicando el nuevo criterio conjuntos compactos con vecindad	75
29	Reportes para el grupo 4 aplicando el nuevo criterio conjuntos compactos con vecindad	76
30	Reportes para el grupo 5 aplicando el nuevo criterio conjuntos compactos con vecindad	76
31	Parte del reporte "Cambio de las aportaciones de riesgo por municipio" entre	80

- la estratificación actual contra la obtenida usando componentes conexas
- 32 Parte del reporte "Cambio de las aportaciones de riesgo por municipio" entre  
82
- la estratificación actual contra la obtenida usando A>means
- 33 Parte del reporte "Cambio de las aportaciones de riesgo por municipio" entre  
85
- la estratificación obtenida usando componentes conexas contra A>means
- 34 Resultados de forma tabular de los diferentes resultados  
87
- 35 Correspondencia general entre métodos en función del número de municipios  
87
- pertenecientes a cada grupo
- 36 Correspondencia entre número de municipios del grupo 1 de los SSH contra  
88
- componentes conexas, k-means para funciones de similaridad y conjuntos compactos con vecindad
- 37 Correspondencia entre número de municipios del grupo 1 (de mayor riesgo)  
88
- para componentes conexas, A:-means para funciones de similaridad y conjuntos compactos con vecindad

# Capítulo 1

## INTRODUCCIÓN

### **1.1.-Antecedentes**

El programa "Atención Integral" de los Servicios de Salud de Hidalgo (SSH) está enfocado a reducir los niveles de enfermedad de las diferentes localidades que integran al Estado de Hidalgo, con un presupuesto autorizado limitado. Para realizar este trabajo, se determinaron cuáles de los 84 municipios que conforman el Estado de Hidalgo presentan los mayores índices de rezago en atención a la salud, carencia de recursos y marginalidad. Obteniendo una estructuración (agrupamiento) en cuatro zonas de riesgo. Haciendo uso de métodos estadísticos como la distribución por cuartiles para determinar que municipios pertenecen a cada grupo de riesgo. Obteniendo dichas estratificaciones en función de las ponderaciones de riesgo calculadas para cada municipio.

### **1.2.- Planteamiento del problema**

Actualmente, los expertos que trabajan con el programa "Atención Integral" no cuentan con una herramienta computacional válida que permita obtener estratificaciones de municipios en zonas de riesgo de forma automatizada. Así como la verificación y comparación de la primera estructuración generada con las nuevas estructuraciones. Además, requieren de dar seguimiento a dichas estructuraciones al paso de los años. Dicha estructuración de municipios que realizaron los SSH fue generada de forma manual, sin considerar la naturaleza de los datos (atributos cuantitativos, cualitativos y mezclados). También, se observan errores en algunos de sus datos (tasas). Tampoco existe conocimiento alguno para decidir cuando 2 municipios pertenecen a un mismo grupo. Además, de que los SSH no cuentan con una herramienta computacional que les permita obtener de manera válida y sistemática dichas estratificaciones de municipios.

## **1.3- Objetivos**

### **Objetivo General**

El objetivo de esta tesis es proporcionar al experto una herramienta computacional que permita generar estructuraciones (agrupamientos) de municipios en forma automatizada, para cada uno de los grupos de edad en el programa "Atención Integral".

### **Objetivos Específicos.**

- Diseñar y desarrollar una herramienta computacional (prototipo) que permita obtener diferentes estratificaciones de municipios en zonas de riesgo para el Estado de Hidalgo.
- Hacer uso de técnicas del RLCP [RuJ+ 2002a] y programar algoritmos para el agrupamiento libre y restringido aplicados al problema en estudio.
- Diseñar y desarrollar una base de datos en Access que permita almacenar las estratificaciones generadas.
- Obtener un promedio de riesgo de salud para cada grupo obtenido que permita etiquetar los grupos de mayor a menor riesgo.
- Presentar los resultados obtenidos de forma gráfica y tabular, así como módulos que permitan realizar un análisis básico entre estratificaciones.
- Verificar si los municipios del Estado de Hidalgo solo pueden estar agrupados en 4 estratificaciones de riesgo.

## **1.4- Metodología de trabajo**

La metodología adoptada para alcanzar los objetivos planteados fue la siguiente:

Inicialmente se realizó un estudio detallado del método actual de estratificación de municipios en zonas de riesgo de los SSH. Con el objeto de observar e identificar las deficiencias de este método, así como el de proponer una solución al problema de estratificación utilizando técnicas del Reconocimiento Lógico Combinatorio de Patrones [RuJ+ 1999a].

Una vez analizado el método actual, se realizó un análisis de los datos del problema aquí expuesto. Esto con la finalidad de conocer a detalle los valores con los que se describen los 84 municipios del Estado de Hidalgo en cuanto a salud se refiere. Además, de conocer el ámbito en el que se desarrolla este trabajo de tesis. Así como el determinar la forma en que el usuario podrá acceder a dichos datos para poder ser utilizados por la herramienta computacional. Es importante mencionar que los datos necesitan ser procesados para obtener tasas, índices, porcentajes o casos y por tal motivo, fue necesario el determinar la forma y tiempo en que estos datos deberán pasar por cierto preprocesamiento para posteriormente poder ser utilizados por la herramienta.

El siguiente paso fue el realizar pruebas experimentales a los datos aplicando criterios de comparación existentes en la bibliografía [RuJ+ 1999a, RuJ+ 2002a]. Con el objetivo de obtener una matriz de semejanza la cual contiene el grado de semejanza entre los objetos del problema en estudio. Para ello fue necesario el diseñar una función de semejanza la cual determina el grado en que se asemejan dos descripciones de objetos, y que esta compuesta por dos funciones de semejanza parciales [AIE+ 1997, RuJ+ 1999a] . Una para los indicadores numéricos y otra mas para los indicadores Aivalente.

Posteriormente se aplicaron los criterios agrupación ales componentes conexas, conjuntos compactos [MaF+ 2000, RuJ+ 1995b] y A>means para uso de funciones de similaridad [GaJ+ 1999], para obtener estratificaciones de municipios en función de la matriz de semejanza de los objetos. Debido a los resultados obtenidos, fue necesario el diseñar e implementar nuevos criterios de comparación que permitieran obtener una mejor semejanza entre las descripciones que conforman a los objetos en estudio.

De forma paralela se analizó y diseño la forma de calcular la aportación información al para cada uno de los indicadores o variables de los municipios. Para obtener dicho peso se utilizó un método estadístico (coeficiente de variación) debido a los diferentes dominios en los que se representan los datos para los indicadores numéricos. Es de suma importancia mencionar que únicamente se obtiene la aportación información al de los indicadores numéricos debido a la experiencia del experto y la naturaleza de los datos para cada indicador.

Una vez que se obtuvieron los agrupamientos de municipios fue necesario el etiquetar los grupos obtenidos en zonas de riesgo de salud. Para ello, se calcularon las aportaciones de riesgo para cada uno de los municipios. En completa cooperación con el experto se especifico para cada uno de los indicadores la forma en que se evaluará la aportación de riesgo. Se obtiene un promedio de aportaciones de riesgo entre los municipios pertenecientes a cada grupo, el cual representará el riesgo de salud final del grupo. De forma paralela también se obtiene una lista de indicadores en función de la influencia que tuvieron al calcular el riesgo de salud final del grupo. Esto con el objetivo de proporcionar al experto información importante para la toma de decisiones.

En función de los requerimientos y necesidades de los expertos y el usuario se diseñaron tres formas básicas para mostrar los resultados. La primera es el mapa del Estado de Hidalgo coloreado en función de los grupos obtenidos, con la posibilidad de utilizar colores de forma automática o eligiéndolos de una paleta de colores. La segunda opción, son tablas presentadas en pantalla con toda la información referente a la estratificación obtenida. Y la tercera son los reportes en papel.

De forma paralela se realizaron diferentes tipos de pruebas y experimentos, los cuales sirvieron para considerar nuevos módulos de la herramienta. Con el objetivo de proporcionar la mayor información posible a los SSH para que se apoyen en la toma de decisiones además de agilizar el procedimiento de estratificación de municipios.

## 1.5- Justificación

Algunas áreas como las Geociencias, las Ciencias Biológicas, la Medicina, la Sociología, la Pedagogía, entre otras, presentan algunas características comunes, como las siguientes: los objetos (de cualquier naturaleza) en estudio se describen en términos de rasgos o variables cuantitativas y cualitativas simultáneamente; algunas descripciones de objetos pueden estar incompletas; y se emplea la *analogía* entre las descripciones como base metodológica para las decisiones. Es decir, aquellas descripciones de objetos que *más se parezcan* deben en general estar en la misma categoría de objetos, tener el mismo diagnóstico o clasificación.

Estas ciencias se denominan Zonas del Conocimiento Poco Formalizadas (o simplemente Ciencias Poco Formalizadas).

El Reconocimiento Lógico Combinatorio de Patrones proporciona herramientas matemáticas para la modelación matemática de problemas relacionados con las ciencias poco formalizadas de manera natural, sin tener que forzar la realidad.

Bajo este enfoque, se han resuelto diversos problemas en torno al ámbito de la Medicina. Entre estos, se pueden mencionar los siguientes: Electrodiagnóstico médico utilizando las herramientas de Reconocimiento de Patrones [LoS+ 1997], una nueva metodología para diagnóstico diferencial de enfermedades [OrM+ 1996], definición de una función de semejanza para pacientes con fisuras de paladar [OrM+ 1998], selección de atributos para clasificación de pacientes con Uveítis [MaJ+ 2000], etc.

Entonces, realizando un estudio del problema a resolver, la naturaleza de los datos, su complejidad y necesidades, se concluyó que las técnicas de agrupamiento paramétricas desarrolladas bajo el RLCP son adecuadas para ser aplicadas y dar una solución al problema de estratificación de municipios de riesgo en el Estado de Hidalgo.

Existen otras técnicas desarrolladas en otros enfoques, las cuales pudieran ser aplicadas para dar otra solución para problema planteado. Sin embargo, esto se encuentra fuera de los alcances de este trabajo de tesis.

Por la gran importancia que tiene el factor salud en la población y en el país, esta herramienta esta diseñada de tal forma que pueda ser utilizada por otras Instituciones de Salud en México. Donde se lleven acabo estratificaciones de municipios en zonas de riesgo. Además, esta herramienta reducirá considerablemente el tiempo que actualmente los SSH requieren para obtener dicha estratificación así mismo, reducir los errores en el manejo de los datos. Así como el de reducir el número de personas necesarias que se encargan en cada fase de la estratificación actual de municipios.

## 1.6- Alcances y limitaciones

El alcance de este trabajo está determinado por la complejidad del mismo. En esta tesis se plantea una herramienta computacional para la estratificación de municipios en zonas de riesgo para el programa "Atención Integral" de los SSH. Utilizando los datos del grupo de edad menor de 5 años. Esto no significa que la herramienta este diseñada específicamente para este grupo de edad. Utilizando algoritmos de clasificación no supervisada, para ofrecer

al experto en salud una solución al problema de estructuración de municipios, ofreciendo una interfaz y manejo amigables al experto en salud.

Los resultados logrados son los siguientes: el desarrollo e implementación de un módulo de interfaz del usuario, que permite cargar, almacenar y procesar un grupo de edad específico; la implementación de un módulo de agrupamiento, que permite utilizar cuatro algoritmos de clasificación no supervisada para estructurar los grupos de edad mencionados; la implementación de un módulo de análisis de resultados generados por los diferentes algoritmos de agrupamiento; así como el diseño e implementación de un nuevo criterio de agrupamiento, basado en las necesidades de los expertos en salud.

El módulo de interfaz del usuario permite manejar diversos criterios para comparar las variables de los diferentes objetos. Aunque tiene un número limitado de formas de comparar estas variables, debido a la diversidad existente y diferentes necesidades que cada usuario tiene, comprendiendo un área de investigación que por si sola se considera compleja. Por lo tanto, este trabajo se delimita al manejo de ocho criterios de comparación entre variables, los cuales fueron obtenidos trabajando en conjunto con el experto en salud.

Además, el módulo de agrupamiento hace uso de cuatro algoritmos de clasificación no supervisada, tres para agrupamiento libre y uno para agrupamiento restringido. No poniendo restricciones en el manejo del número de objetos y variables a manejar. Sin embargo, este módulo no incorpora más algoritmos de agrupamiento, ya que esto rebasa el alcance de este trabajo.

Por otra parte, el módulo de análisis de resultados auxilia al experto de dos formas: por medio de la visualización de uno o varios mapas coloreados de acuerdo a los agrupamientos generados; además de unas tablas que le ofrecen la información de que municipios cambiaron en su clasificación de riesgo. Este módulo esta delimitado a la generación de resultados en solamente estas dos formas, debido a las necesidades plasmadas por el experto en salud.

## **1.7.- Estructura del trabajo**

Este trabajo de tesis está organizado de la siguiente manera: En el capítulo 2 se presenta la modelación matemática para la estratificación de municipios, así como los conceptos básicos necesarios para la discusión de los algoritmos y métodos de agrupamiento propuestos. En el capítulo 3 se presenta un estudio sobre los trabajos relacionados. El análisis y diseño de la herramienta computacional muestra en el capítulo 4. En el capítulo 5 se desarrolla la herramienta computacional así como la implementación del nuevo criterio agrupacional conjuntos compactos con vecindad. Los experimentos efectuados con datos reales de los SSH aplicados a la herramienta computacional son mostrados en el capítulo 6. por último se presentan las conclusiones y trabajo futuro.

# Capítulo 2

## MODELACIÓN MATEMÁTICA PARA LA ESTRATIFICACIÓN DE MUNICIPIOS

Con el objetivo de tener mayor claridad sobre el marco en el que se desarrolla este trabajo de tesis, en este capítulo se presentan los conceptos básicos y definiciones que se manejan a lo largo del documento. Así como la modelación matemática del problema para la estratificación de municipios en zonas de riesgo para el Estado de Hidalgo.

### **2.1.- Problemas de Reconocimiento de Patrones**

En Reconocimiento de Patrones existen esencialmente cuatro familias de problemas [RuJ+ 1995a], [RuJ+ 1995b] las cuales son: Clasificación Supervisada, Selección de Variables, Clasificación Parcialmente Supervisada y Clasificación no Supervisada. Sin embargo, este trabajo de tesis estará enfocado principalmente en la Clasificación no Supervisada, debido a la naturaleza del problema aquí planteado.

### **2.2.- Clasificación no supervisada**

En esta familia se tiene un universo de objetos, pero no se conoce cómo se agrupan (no existen clases definidas), y este es precisamente el objetivo principal de dicha clasificación. El problema radica en encontrar cómo se estructura el universo de objetos mencionado anteriormente. De dicha estructuración, en términos generales puede dividirse en dos clases: restringida y libre. En la primera, se determina a priori el número de clases que se deben generar, en la segunda, se desconoce el número de clases en la que se estructuran los objetos.

Resolver un problema de clasificación no supervisada [RuJ+ 1995b], consiste en hallar la estructura interna de un conjunto de descripciones de objetos en el espacio de representación inicial, la determinación de una función de analogía ( semejanza) en términos de la cual se definen las relaciones entre las diferentes descripciones de los objetos y de un cierto criterio de agrupamiento [RuJ+ 1995a], y así obtener un cierto número de grupos de

objetos en función de sus descripciones.

Uno de los objetivos de este trabajo es verificar si los municipios del Estado de Hidalgo solo pueden estratificarse en 4 zonas de riesgo. Así como el de comparar esta estratificación en 4 zonas de riesgo de los SSH contra las estratificaciones que se obtengan utilizando un algoritmo restringido a 4 clases. Por esta razón, este trabajo de tesis está enfocado en la familia de clasificación no supervisada.

### 2.2.1.- Descripciones de objetos

En un problema de clasificación no supervisada [RuJ+ 1995b], la descripción de un objeto, se entiende como un n-uplo de valores cuantitativos y cualitativos, definidos en un espacio de representación inicial denotado por  $M_1 \times \dots \times M_n$  (i.e.  $O = (x_1(o), x_2(o), \dots, x_n(o)) \in M_1 \times \dots \times M_n$ ), donde  $x_i(o) \in M_i$ ,  $n$  es el número total de rasgos;  $M_i$  es el conjunto de valores admisibles del rasgo  $x_i$ . Los valores que toman los rasgos pueden ser de cualquier naturaleza (i.e. atributos cualitativos y cuantitativos). Por lo tanto, las descripciones de los objetos pueden estar conformadas por rasgos cuantitativos y cualitativos simultáneamente. Además de no considerar operaciones algebraicas, lógicas, o distancias definidas de forma a priori. Estas operaciones y medidas pueden estar presentes, pero no son necesarias. También, se denotará al conjunto de rasgos por  $R$ , y al conjunto de descripciones de objetos por  $MI$ .

### 2.2.2.- Criterios de comparación o analogía entre variables

Analogía o similaridad es el parecido que tienen entre sí dos objetos en dependencia de los rasgos que los describen. [RuJ+ 1999a, RuJ+ 2002a, RuJ+ 2002b].

**Definición 1** [RuJ+ 1999a]. Un criterio de comparación entre valores de una variable es una función  $C_i: M_i \times M_i \rightarrow L_i$  que retorna la semejanza o parecido que guardan dos valores de un mismo rasgo, siendo  $L_i$  un conjunto ordenado

El criterio de comparación  $Q$ ,  $i=1, \dots, n$  ofrece una evaluación del grado de semejanza entre dos valores cualesquiera de una variable.

### 2.2.3.- Funciones de semejanza (entre objetos)

**Definición 2** [A1E+ 1997, RuJ+ 1999a, RuJ+ 2002b]. Una función de semejanza (parcial)  $\beta: MI \times MI \rightarrow L$  es una evaluación del grado en que se asemejan dos descripciones de objetos de  $MI$ , tal que las relaciones de semejanza entre algún par de objetos son preservadas cuando se considera el total de la semejanza entre estos objetos, y el máximo valor de semejanza se alcanza cuando se considera la misma parte del mismo objeto para cualquier subconjunto no vacío de  $R$ .

**Definición 3** [MaF+ 2000, SaG+ 2003]. Un umbral de semejanza es un valor denotado por  $\beta_0 \in L$ , el cual permite controlar que tan similares deben ser un par de objetos para ser considerados semejantes.

Dependiendo de la cercanía o similaridad deseada, un valor apropiado de este umbral debe ser seleccionado por el usuario o experto, y entonces se podrán obtener diferentes estructuraciones de los datos.

**Definición 4** [MaF+ 2000, SaG+ 2003]. Se dirá que dos objetos  $O_i, O_j \in M$  son semejantes (i.e.  $\beta_0$  semejantes) si la semejanza entre ellos es mayor o igual que el valor del umbral de semejanza. Dicho de otra forma,  $O_i, O_j \in M$  son  $\beta_0$  semejantes si  $\beta(O_i, O_j) \geq \beta_0$  -Si un objeto  $O_i \in MI$  no es semejante con ningún objeto  $O_j \in MI$ , entonces se dirá que  $O_i$  es un objeto aislado (i.e.  $\wedge$ -aislado).

#### 2.2.4.- Criterios agrupacionales

Un criterio agrupacional define cuando un objeto va a pertenecer a un grupo, o porqué dos objetos pertenecen a una misma agrupación.

**Definición 5 (componentes conexas)** [MaF+ 2000, RuJ+ 1995b]. Una componente conexa  $CX$  es un conjunto no vacío que cumple con las siguientes condiciones:

- Para cualquier par de elementos  $O_i, O_j$  de  $CX$ , existe una sucesión de elementos (camino) los cuales pertenecen a  $CX$ , que empieza en  $O_i$ , y termina en  $O_j$ , tales que cada objeto es  $\beta_0$ -semejante con el siguiente.
- No existe ningún elemento  $O_j$  fuera de  $CX$  que sea  $\beta_0$ -semejante con algún elemento perteneciente a  $CX$ .
- Todo elemento  $\beta_0$ -aislado es una componente conexa.

**Definición 6 (conjuntos compactos)** [MaF+ 2000, RuJ+ 1995b]. Un conjunto compacto  $CP$  es un conjunto no vacío, el cual cumple con lo siguiente:

- Todo elemento  $O_j$  de  $CP$ , tiene en  $CP$  al elemento que más se le parece, y que además  $\beta_0$  es semejante con él. Además, no existe fuera del conjunto  $CP$  ningún elemento  $O_j$  cuyo elemento más parecido y  $\wedge$ -semejante con  $O_j$ , este dentro de  $CP$ .
- Para cualquier par de elementos  $O_i, O_j$  de  $CP$  existe una sucesión de elementos los cuales son los más semejantes y pertenecen a  $CP$ , que empieza en  $O_i$ , y termina en  $O_j$ , tales que cada objeto es  $\beta_0$  semejante con el siguiente.
- Todo elemento  $\beta_0$ -aislado es un conjunto compacto.

**Definición 7 (conjuntos compactos con vecindad)** [LoM+ 2004]. Un conjunto compacto con vecindad  $CV$  es un conjunto no vacío que cumple las siguientes condiciones:

- Todo elemento  $O_j$  de  $CV$ , tiene en  $CV$  al elemento más parecido con él, así como los elementos que se encuentran dentro de la vecindad del máximo y que son  $\wedge$ -semejantes con él. Además, no existe fuera del conjunto  $CV$  ningún elemento  $O_j$  cuyos elementos más parecidos y  $\%$ -semejantes con  $O_j$  estén dentro de la vecindad del máximo que está en  $CV$ . Entendiéndose por vecindad del máximo la diferencia entre el valor máximo y un valor  $asL$ .
- Para cualquier par de elementos  $O_i, O_j$  de  $CV$  existe una sucesión de elementos los

- cuales se encuentran en la vecindad del máximo y pertenecen a  $CP$ , que empieza en  $O_i$ , y termina en  $O_j$ , tales que cada objeto es  $\wedge$ -semejante con el siguiente.
- c) Todo elemento  $\wedge$ -aislado es un conjunto compacto con vecindad.

*Definición 8 (algoritmo k-means para funciones de similitud) [GaJ+ 1999].* El algoritmo de agrupamiento restringido k-means para funciones de similitud es una extensión del algoritmo clásico k-means, el cual permite manejar descripciones de objetos mezcladas (cuantitativas y cualitativas), así como funciones de similitud para la comparación entre objetos. Este algoritmo se basa en unos objetos que son los que más se parecen a los objetos restantes de un mismo grupo, que los denomina Holotipos, para representar los centro ideas (i.e. los centros en el algoritmo k-means clásico) de cada grupo. Estos Holotipos son usados para efectuar el reagrupamiento de los objetos restantes en cada iteración, hasta realizar un número de iteraciones determinado.

### 2.3.- Diseño y definición de los criterios de comparación

Sin duda una parte fundamental de este trabajo de tesis fue la de diseñar los criterios de comparación adecuados para este tipo de problema. Cabe mencionar que dichos criterios fueron creados y supervisados con ayuda de los expertos de los SSH, ya que estos criterios son la base para obtener estratificaciones de municipios apegados a la realidad y considerando las necesidades de los expertos.

Se propusieron y evaluaron algunos criterios de comparación existentes en la bibliografía para determinar cuándo dos municipios en un mismo atributo son semejantes, tanto para atributos numéricos como para atributos k-variantes. Aunque la mayoría de ellos resultaban ser demasiado restrictivos o demasiado flexibles para este problema en especial. Por lo que se tuvo que profundizar más en el análisis de la naturaleza de los datos y así poder crear criterios adecuados a las necesidades de cada indicador, donde se requiere conocer el tipo de atributo o variable, esto se observa en la figura 1.

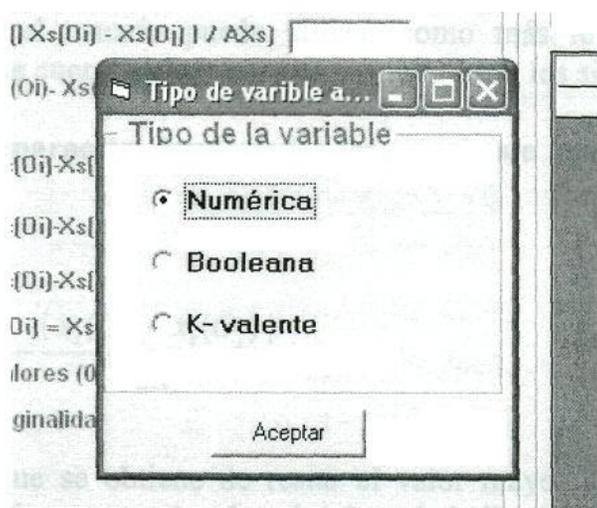


Figura 1. Tipo de dato del indicador

Esto no quiere decir que la herramienta no acepte cualquier tipo de criterio, de hecho se incorporaron en la herramienta algunos criterios típicos de la literatura, ver figura 2. Con la

finalidad de que si se incorpora algún otro tipo de atributo el experto cuente con criterios de comparación adicionales para poder utilizar dicho indicador en una nueva estratificación, o dar soluciones a problemas similares que no requieran de criterios elaborados.

Es importante mencionar que cada indicador  $x_s$ ,  $s = 1, \dots, n$ , necesita un criterio de comparación para cuando cualquier par de objetos  $O_i, O_j$  se estén comparando en ese indicador.

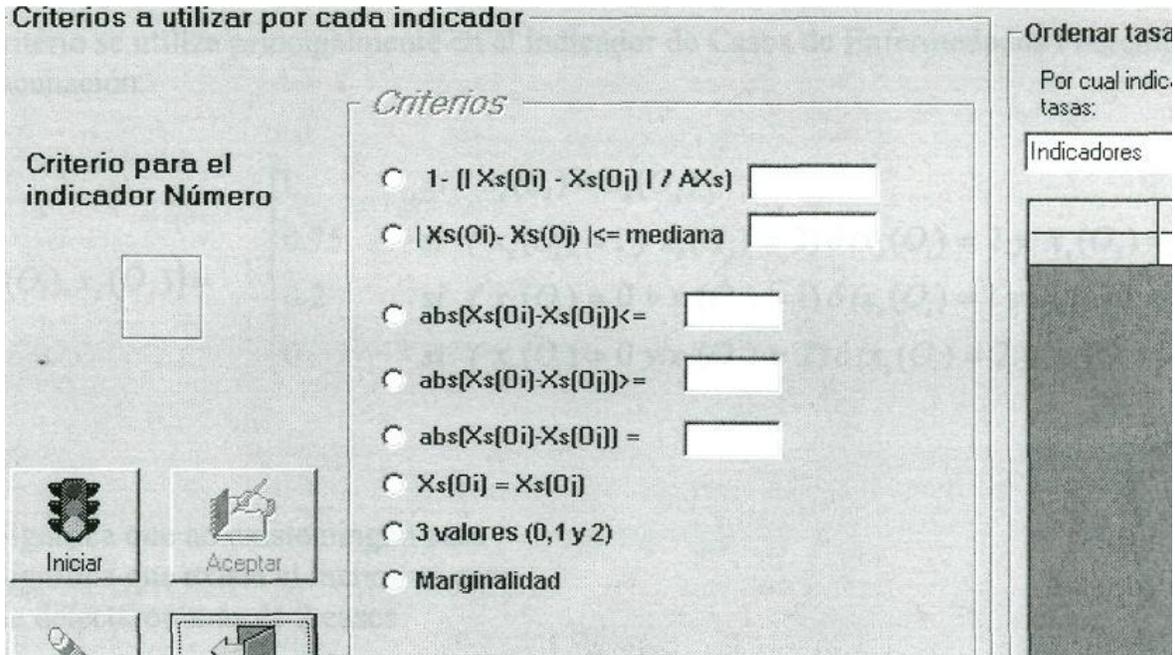


Figura 2. Captura de criterios de comparación para cada indicador.

En esta figura, se visualizan los 8 criterios de comparación que maneja esta herramienta computacional, y que el usuario puede utilizar como más le parezca. Los criterios de comparación con los que cuenta la herramienta propuesta son los siguientes:

**Los criterios de comparación implementados para este problema específico son los siguientes:**

a)

$$C_s(x_s(O_i), x_s(O_j)) = 1 - \frac{|x_s(O_i) - x_s(O_j)|}{\Delta x_s}$$

Donde:

$\Delta x_s$ : Es un factor que se obtiene de restar el valor mayor al valor menor de las tasas (porcentajes, índices o casos) ordenadas de cada indicador numérico.

b)

$$C_s(x_s(O_i), x_s(O_j)) = \begin{cases} 1 & \text{Si } |x_s(O_i) - x_s(O_j)| \leq \text{mediana} \\ & \end{cases}$$

0 en otro caso

Donde:

*mediana*: Es la mediana de los datos para cada indicador.

c)

Este criterio se utiliza principalmente en el indicador de Casos de Enfermedades Prevenibles por Vacunación.

$$C_s(x_s(O_i), x_s(O_j)) = \begin{cases} 1 & \text{si } x_s(O_i), x_s(O_j) \\ 0.75 & \text{si } (x_s(O_i), x_s(O_j)) = 2 \text{ } \wedge \text{ } (x_s(O_i)) = 2 \text{ y } x_s(O_j) = 1 \\ 0.2 & \text{si } (x_s(O_i), x_s(O_j)) = 1 \text{ } \wedge \text{ } (x_s(O_i)) = 1 \text{ y } x_s(O_j) = 0 \\ 0 & \text{si } (x_s(O_i), x_s(O_j)) = 2 \text{ } \wedge \text{ } (x_s(O_i)) = 2 \text{ y } x_s(O_j) = 0 \end{cases}$$

Donde:

- 0 : Significa que no existe ningún caso
- 1 : Significa que existe al menos un caso
- 2 : Se detectaron más de 2 casos

d)

Este criterio se utiliza principalmente en el indicador de marginalidad

$$C_s(x_s(O_i), x_s(O_j)) = \begin{cases} 1 & \text{si } |x_s(O_i) - x_s(O_j)| = 0 \\ 0.75 & \text{si } |x_s(O_i) - x_s(O_j)| = 1 \\ 0.5 & \text{si } |x_s(O_i) - x_s(O_j)| = 2 \\ 0.25 & \text{si } |x_s(O_i) - x_s(O_j)| = 3 \\ 0 & \text{si } |x_s(O_i) - x_s(O_j)| = 4 \end{cases}$$

**Los criterios de comparación utilizados existentes en la bibliografía son los siguientes:**

e)

$$C_s(x_s(O_i), x_s(O_j)) = \begin{cases} 1 & \text{si } |x_s(O_i) - x_s(O_j)| \leq VAR \\ 0 & \text{en otro caso} \end{cases}$$

Donde:

VAR: Es un valor proporcionado por el usuario.

f)

$$C_s(x_s(O_i), x_s(O_j)) = \begin{cases} 1 & \text{si } |x_s(O_i) - x_s(O_j)| \geq VAR \\ 0 & \text{en otro caso} \end{cases}$$

Donde:

VAR: Es un valor proporcionado por el usuario.

g)

$$C_s(x_s(O_i), x_s(O_j)) = \begin{cases} 1 & \text{si } |x_s(O_i) - x_s(O_j)| = VAR \\ 0 & \text{en otro caso} \end{cases}$$

Donde:

VAR: Es un valor proporcionado por el usuario.

h)

Este es el criterio de comparación más sencillo y además el más restrictivo, generalmente se usa en atributos booleanos.

$$C_s(x_s(O_i), x_s(O_j)) = \begin{cases} 1 & \text{si } x_s(O_i) = x_s(O_j) \\ 0 & \text{en otro caso} \end{cases}$$

## 2.4.- Obtención de la aportación informacional de los atributos o indicadores

Para obtener la aportación informacional de cada indicador se utilizó un método estadístico, para calcular el coeficiente de variación y utilizarlo como el peso informacional del indicador o atributo. Debido a los diferentes dominios en los que se presentan los datos de los indicadores numéricos, se utiliza el coeficiente de variación.

Varianza de los datos

$$S^2 = \frac{1}{t} \sum_{i=1}^t (v_i - \bar{x})^2$$

Donde

$t$ : Es el total de la población finita de los datos

$\bar{x}$ : Es el promedio de los datos

$v_i$ : Valor  $i$ , donde  $t = 1, 2, \dots, t$

Posteriormente se obtiene la Desviación Estándar  $S$ , la cual se obtiene de la raíz cuadrada de la varianza de los datos. El coeficiente de variación obtiene la dispersión de los datos en función del promedio de los datos. En este caso en particular, el hecho de que los datos se encuentren más alejados del promedio, significa que no se tiene control por parte de los SSH en ese indicador, puesto que en cada municipio se realizan acciones de forma diferente, por lo que a mayor dispersión mayor será la atención a ese indicador.

El coeficiente de variación de determina de la siguiente manera:

$$P_i = \frac{\text{desviacion es tan dar}}{\text{media}} = \frac{S}{\bar{x}}$$

Por lo tanto,  $P_i$  es la aportación informacional del indicador  $i$ , donde  $i = 1, 2, \dots, \text{total}$  de indicadores.

## 2.5.- Diseño y definición de la función de semejanza

En este problema en especial, debido a la naturaleza de los atributos (mezclados), para determinar la semejanza entre un par de objetos, fue necesario el implementar una función de semejanza total compuesta de dos *funciones de semejanza parcial*, una para los atributos numéricos y otra para los atributos Avalentes. Se denomina *función de semejanza parcial* a cualquier restricción de una función de semejanza a un subconjunto de  $R$  [A1E+ 1997].

**La función de semejanza parcial para los atributos cuantitativos es la siguiente:**

$$\beta_{numérica}(O_i, O_j) = \frac{\left[ \frac{\sum_{i=1}^n P_i \varphi_i}{n} \right]}{\sum_{i=1}^n P_i}$$

Donde:

$i = 1, 2, \dots, n$ , donde  $n$  es el total de atributos o indicadores numéricos

$P_i$ : Es la aportación informacional del atributo  $i$ .

$\varphi_i$ : Es el valor que arroja el criterio de comparación del atributo o indicador  $i$ .

**La función de semejanza parcial para los atributos cualitativos es la siguiente:**

En este tipo de atributos, no se calcula la aportación informacional, i.e. su aportación es el valor tal cual que arroja el criterio de comparación. Debido a que en este caso no existe mucha dispersión de los datos.

$$\beta_{k\text{-valente}}(O_i, O_j) = \frac{\left[ \frac{\sum_{i=1}^k P_i \varphi_i}{k} \right]}{k}$$

Por lo tanto, la función queda de la siguiente manera:

$$\beta_{k\text{-valente}}(O_i, O_j) = \frac{\left[ \frac{\sum_{i=1}^k P_i \varphi_i}{k} \right]}{k^2}$$

Donde:

$i = 1, 2, \dots, k$ , donde  $k$  es el total de atributos o indicadores k-valentes.

$\varphi_i$ : Es el valor que arroja el criterio de comparación del atributo o indicador  $i$ .

Finalmente, la función de semejanza total entre un par de objetos ( $\beta_{TO}(O_i, O_j)$ ) considerando todos sus atributos se define de la siguiente manera:

$$\beta_{TO}(O_i, O_j) = \lfloor \beta_{numérica}(O_i, O_j) \rfloor + \lfloor \beta_{k\text{-valente}}(O_i, O_j) \rfloor$$

## 2.6.- Obtención del promedio de riesgo final por grupo

En este trabajo de tesis también se desarrolló un módulo que etiqueta los grupos obtenidos en zonas de riesgo. Esto debido a que los agrupamientos generados no tendrían razón alguna para el experto si no se sabe cuál de estos grupos de municipios presenta mayor riesgo de salud en el Estado. Para dar una solución a este problema se realizó el siguiente procedimiento:

- a) Se obtuvo la aportación de riesgo de salud de cada uno de los 84 municipios del Estado de Hidalgo, especificando para cada indicador la forma en que se evaluará para calcular la aportación de riesgo por indicador.
- b) Se obtiene un promedio de aportaciones de riesgo de los municipios que pertenecen a un grupo específico, el cual representará el promedio de riesgo final del grupo.
- c) Se ordenan de forma descendente los grupos en función de su promedio de riesgo final, y así el experto tendrá conocimiento acerca de cuál es el grupo que contiene a la mayoría de municipios con mayores aportaciones de riesgo. Además de apoyar al experto en la toma de decisiones.

Se proponen dos casos de evaluación de la aportación de riesgo para los indicadores, debido a que en algunos indicadores un valor alto no significa que tenga mayor riesgo. Estos casos son los siguientes:

**Caso 1: A mayor valor en el dato implica mayor aportación de riesgo para el indicador.**

$$Ap = \frac{V(x_i(O_m))}{Max(x_i)}$$

Donde:

$Ap$ : Aportación de riesgo para el caso 1

$V(x_i(O_m))$  : Es el valor (tasa, índice, casos, etc) que tiene el municipio  $O_m$ ,  $m = 1, \dots, h$  donde  $h$  es el total de objetos, en el indicador  $x_i$ ,  $i = 1, \dots, n$ , donde  $n$  es el total de indicadores.

$Max(x_i)$  : Es el mayor valor (tasa, índice, casos, etc) que toma el indicador  $i$ .

**Caso 2: A mayor valor en el dato implica menor aportación de riesgo para el indicador.**

$$Ap = 1 - \frac{V(x_i(O_m))}{Max(x_i)}$$

Donde:

$A_p$  : Aportación de riesgo para el caso 2.

$V(x_j(O_m))$  : Es el valor (tasa, índice, casos, etc) que tiene el municipio  $O_m$ ,  $m = 1, \dots, h$ , donde  $h$  es el total de objetos, en el indicador  $x_j$ ,  $j = 1, \dots, n$ , donde  $n$  es el total de indicadores.

$Max(x_j)$  : Es el mayor valor (tasa, índice, casos, etc) que toma el indicador  $j$ .

La forma en que se capturan estos casos para cada indicador se observa en la figura 3.

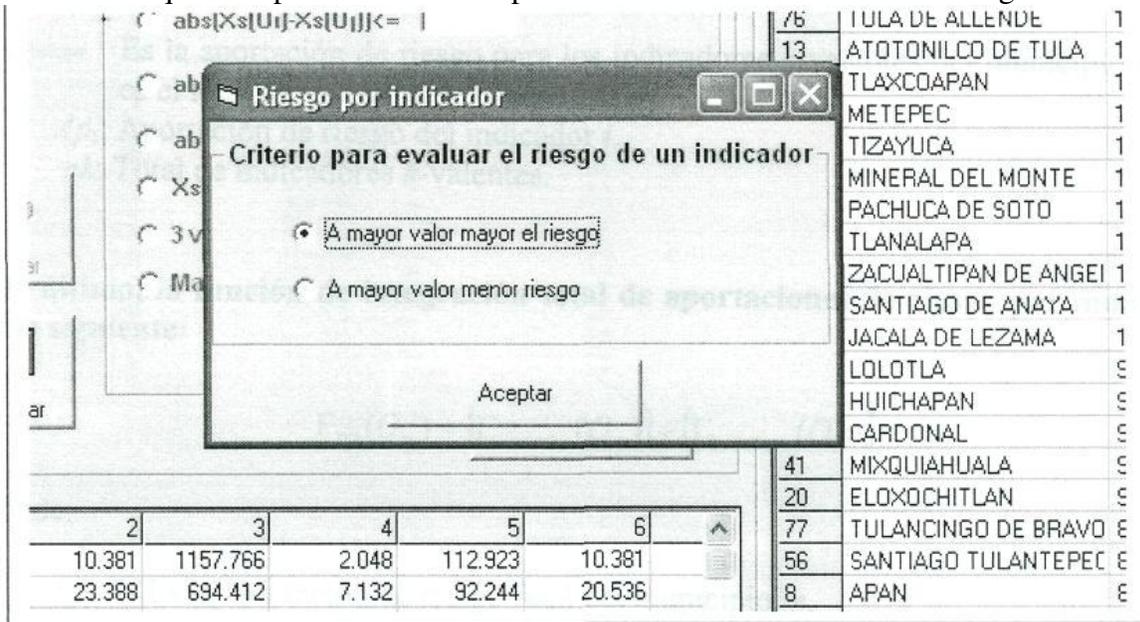


Figura 3. Casos para la evaluación de aportación de riesgo de los indicadores.

Finalmente se aplica una función de integración de los indicadores para obtener la aportación de riesgo final por municipio, sin perder en cuenta la naturaleza de los atributos o indicadores, así como su aportación informacional. Por lo que se aplican también dos funciones parciales de integración de riesgo.

**La función parcial de integración de aportaciones de riesgo para los atributos numéricos es la siguiente:**

$$\Gamma_{numérica}(O_m) = \frac{\sum_{i=1}^{nm} P_i A_{p_i}}{\sum_{i=1}^{nm} P_i}$$

Donde:

$\Gamma_{Numérica}$ : Es la aportación de riesgo para los indicadores numéricos del municipio  $m$ .

$A_{p_i}$ : Aportación de riesgo del indicador  $i$ .

$P_i$ : Aportación informacional del indicador  $i$ .

$nm$ : Total de indicadores numéricos.

**De igual forma, la función parcial de integración de aportaciones de riesgo para los**

**atributos k-valentes esta definida de la siguiente forma:**

$$\Gamma_{k\text{-valente}}(O_m) = \frac{\sum_{i=1}^{nk} P_i A p_i}{k}$$

Donde:

$\Gamma_{k\text{-valente}}$  :Es la aportación de riesgo para los indicadores k-valentes del municipio  $m$ , donde  $k$  es el total de indicadores k-valentes.

$A p_i$ : Aportación de riesgo del indicador  $i$ .

$nk$ : Total de indicadores &-valentes.

**Por último, la función de integración total de aportaciones de riesgo por municipio esta es la siguiente:**

$$\Gamma_{TO}(O_m) = [\Gamma_{numérica}(O_m)] + [\Gamma_{k\text{-valente}}(O_m)]$$

Donde:

$\Gamma_{TO}(O_m)$  : Es la aportación de riesgo final del municipio  $m$ .

Finalmente, se le proporciona al experto un listado ordenado de forma descendente en función a la aportación de riesgo de cada indicador como se puede ver en la figura 4, para poder conocer qué indicadores influyeron más en el riesgo total del grupo. Y así los SSH podrán tener más herramientas para la toma de decisiones.

Indicador	Deteccion de	Casos de EPV	Marginalidad	Morb. EDAS	Morb. IRAS	Morb. II
Prom. riesgo	1.0	0.78	0.52	0.2	0.19	

Figura 4. Influencia de los indicadores para el cálculo del riesgo.

De igual manera, se puede obtener el reporte de este análisis, el cual se ilustra en la figura 5. Este reporte muestra a todos los indicadores ordenados de forma descendente en función de la aportación de riesgo para cada uno de los indicadores y así observar cuales de ellos influyeron más para obtener la aportación de riesgo de cada grupo.

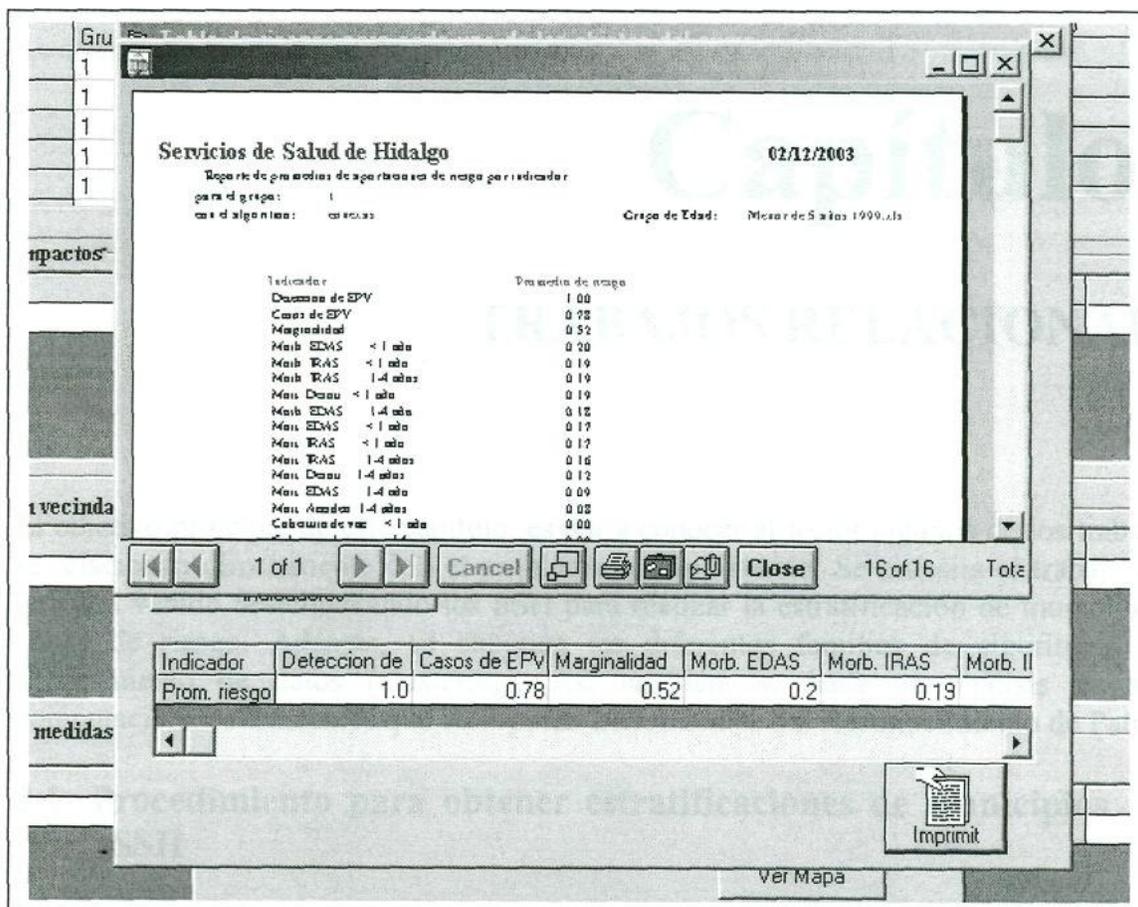


Figura 5. Reporte del análisis de aportaciones de riesgo por indicador.

# Capítulo 3

## TRABAJOS RELACIONADOS

El objetivo principal de este capítulo, es dar a conocer al lector algunos de los trabajos que se relacionan directamente con el problema aquí planteado. Se muestra el trabajo original que han venido desempeñando los SSH para realizar la estratificación de municipios en 4 zonas de riesgo. Además, se exponen las diferentes familias de algoritmos para el agrupamiento de datos (clustering). Así también, se hace un análisis de sistemas computacionales existentes que incorporan herramientas del Reconocimiento de Patrones.

### 3.1- Procedimiento para obtener estratificaciones de municipios de los SSH

Como se mostró anteriormente, la estratificación de municipios se realiza por grupos de edad como se observa en el Anexo A. Para lograr la estratificación de municipios en zonas de riesgo los expertos realizaron el siguiente procedimiento:

1.- Obtener las tasas, porcentajes, número de casos o marginalidad para todos y cada uno de los indicadores por municipio que conforman al grupo de edad correspondiente.

Para obtener las tasas en los indicadores de morbilidad y mortalidad se utilizó la siguiente fórmula:

$$\frac{\text{No. Casos o No. Defunciones}}{\text{Poblacion (indicador)}} \times k \text{ Habitantes}$$

Donde:

$k$  Habitantes : Es una constante definida por un número de habitantes y que los expertos determinan el valor que puede tomar.

**Nota:** Es importante mencionar que esta fórmula para obtener la tasa es la que más se utiliza en los indicadores que conforman los grupos de edad, pero no para todos los indicadores. En algunos indicadores como (casos EPV, coberturas de vacunación y la marginalidad) no se les aplica fórmula alguna, es decir el valor que tienen es el que se utiliza para los cálculos.

2.- Ordenar de forma descendente las tasas, casos, porcentajes o marginalidad para todos los municipios en cada uno de los indicadores como se muestra en las tablas A y B del Anexo D.

3.- Aplicar la distribución por cuartiles a dichas ordenaciones por medio de:

$$Q_0 \frac{nm}{100} \times 25$$

$$Q_1 \frac{nm}{100} \times 50$$

$$Q_2 \frac{nm}{100} \times 75$$

$$Q_3 \frac{nm}{100} \times 100$$

Donde:

$Q_i$  = Cantidad de municipios dentro de la zona de riesgo  $i$ , para  $i=0,1,2,3$   
 3= Muy alto, 2 = Alto, 1 = Medio, 0= Bajo

$nm$  = Total de municipios

4.- Se asigna ( $pond_j$ ) =  $i$  (zona de riesgo) para cada indicador  $j$  de cada municipio.

5.- Sumar las ponderaciones para cada municipio de todos los indicadores del grupo de edad correspondiente

$$Ponderacion\ Final = \sum_{k=1}^{nm} pond_j \text{ donde } (pond_j) \text{ es el valor de ponderacion que tiene el municipio } k \text{ en el indicador } j$$

6.- Ordenar las sumas de ponderaciones finales de todos los municipios de forma descendente.

7.- Aplicar nuevamente la distribución por cuartiles a la lista ordenada de ponderaciones con las formulas anteriores, además de etiquetar las ponderaciones para zonas de riesgo de la siguiente manera:

3= Muy alta, 2= Alta, 1 = Medio y 0= Baja

Un ejemplo de este procedimiento se puede observar en el Anexo C.

Este procedimiento realiza la estratificación de municipios para cada uno de los indicadores en función del ordenamiento de las tasas, aplicando la distribución por cuartiles, donde prácticamente cada 21 municipios pertenecen a una zona de riesgo. Se observa que, dependiendo del valor de la tasa del límite del cuartil que toman los municipios con el mismo valor de la tasa, para entonces juntarlos y lograr grupos con un número de municipios diferente, por ejemplo (grupo 1 =27 municipios, grupos2=19 municipios, etc.).

Además,-no se considera que la semejanza entre las tasas de una misma zona de riesgo varía considerablemente, como para pertenecer a un mismo grupo. Al obtener la estratificación final, se realiza una suma final de ponderaciones por indicadores, se aplica nuevamente la distribución por cuartiles y se ordenan nuevamente por orden descendente. Sin embargo, dos municipios en uno o más indicadores pueden diferir en gran medida. Por tener el mismo valor en la suma final de ponderaciones hace que pertenezcan al mismo grupo.

## **3.2- Agrupamiento de Datos**

Actualmente, se han desarrollado algoritmos de agrupamiento para bases de datos. Sin embargo, no todos pueden manejar datos mezclados. El agrupamiento de datos es una técnica importante para el análisis de datos exploratorio, y se ha estudiado por varios años. Se han desarrollado técnicas en muchos dominios prácticos tales como la estadística [JaA+ 1996, DuB+ 1980, DuR+ 2000], en aprendizaje automático (machine learning) [CaJ 1992] y en bases de datos [NgR+ 1994], entre otros. Generalmente se manejan dos familias de algoritmos de agrupamiento [EsL 1977], [KaL+ 1990]: en partición y jerárquicos.

### **3.2.1- Algoritmos de agrupamiento en partición.**

Los algoritmos en partición [EsL 1977], asumen que el número  $k$  de grupos a formar es conocido, el objetivo es distribuir los objetos en los diferentes grupos, de tal forma que la distancia intra grupos sea minimizada. Y al mismo tiempo, la divergencia entre cualesquiera dos grupos sea maximizada. Estos algoritmos inician con una partición inicial y van redistribuyendo los objetos de un grupo a otro tratando de optimizar la calidad del grupo. Estos algoritmos que operan en espacios métricos presentan las siguientes limitaciones: Pueden aplicarse en problemas donde existen objetos descritos por atributos cuantitativos o cualitativos exclusivamente, pero no mezclados. Y no manejan ausencia de información. No garantizan encontrar un óptimo global pero si local. Tampoco se puede saber si se halló la mejor solución, y con diferentes ordenamientos en los objetos de entrada se pueden generar diferentes soluciones. También es necesario especificar el número de grupos que formarán el agrupamiento. Estos algoritmos pueden presentar problemas cuando se procesan grupos con formas arbitrarias.

### **3.2.2.- Algoritmos de agrupamiento jerárquicos.**

El objetivo de estos algoritmos [DuR+ 2000, EsL 1977] es unir grupos para crear uno nuevo

(métodos aglomerativos), o bien dividiendo grupos para crear dos nuevos (métodos divisivos). Estos procesos son factibles de realizar si las uniones o divisiones se desarrollan sucesivamente, y una distancia/ semejanza entre grupos en la conglomeración es minimizada/maximizada, convergiendo hacia un mayor nivel de generalización. O una distancia/ semejanza entre grupos en el proceso de división es maximizada/minimizada tendiendo hacia un mayor nivel de particularización.

Sin embargo, los algoritmos jerárquicos presentan las siguientes limitaciones: Los resultados que brindan no constituyen precisamente una clasificación. Con muchos objetos, el dendograma puede volverse complejo y el usuario puede interpretarlo de manera incorrecta. Los resultados obtenidos dependen del orden de entrada de los datos. Además, estos algoritmos requieren de un criterio de parada para finalizar su ejecución. En el mejor de los casos, la complejidad de estos algoritmos es exponencial. También, hay otras familias de algoritmos, las cuales pueden contener propiedades de las familias mencionadas anteriormente, pero que difieren en su formalización y forma de trabajo. Entre estas técnicas, se pueden mencionar los métodos y algoritmos basados en semejanzas.

Por todo esto, surgen los métodos y algoritmos basados en semejanzas capaces de utilizar datos numéricos, cualitativos y mezclados, aún con ausencia de información en sus atributos o variables. En resumen, este trabajo de tesis está enfocado en estos métodos y algoritmos debido a la naturaleza de los datos y a las necesidades de los expertos.

### **3.2.3.-Métodos y algoritmos basados en semejanzas.**

Estos métodos no se basan en el enfoque estadístico o sintáctico estructural. Utilizan medidas arbitrarias de semejanza entre objetos, algunos de ellos utilizan un umbral de semejanza que indica cuando dos objetos son semejantes y que pueden pertenecer a un mismo grupo, pero otros algoritmos no lo usan, generando particiones (componentes conexas y conjuntos compactos) o cubrimientos (conjuntos fuertemente compactos y conjuntos completos maximales) en agrupamientos libres (sin un número determinado de grupos). Los métodos basados en la teoría de grafos, tratan a los objetos como vértices, y a la similaridad entre pares de objetos como las aristas de un grafo.

Los métodos usados en el RLCP suponen la determinación de un espacio de representación inicial de los objetos (no necesariamente métrico), así como la función de semejanza  $P$  en términos de la cual se definen las relaciones entre las diferentes descripciones de los objetos en estudio, y de un cierto criterio de agrupamiento. Al igual que los métodos basados en la teoría de grafos [AuJ+ 1970], consideran que se tiene un umbral de semejanza llamado  $\beta_0$ , y que el objeto  $O_i$  es  $\beta_0$  semejante con el objeto  $O_j$  si  $\beta(O_i, O_j) \geq \beta_0$ . Estos métodos pueden aplicarse en problemas donde los objetos están descritos por variables cuantitativas y cualitativas mezcladas, y pueden presentar ausencia de información.

En estos métodos generalmente, se define una matriz cuadrada de  $n \times n$  denominada matriz de semejanza o similaridad, la cual contiene las comparaciones entre los pares de objetos del conjunto a agrupar. Sin embargo existen algoritmos que no requieren del cálculo de la matriz de semejanza  $\mathcal{E}$ -means para funciones de similaridad [GaJ+ 1999].

En particular, algunos métodos definidos en el RLCP que generan cubrimientos como los conjuntos fuertemente compactos y conjuntos completos maximales, los resultados obtenidos pueden ser difíciles de manipular para los expertos, por la forma tan restrictiva de agrupar los objetos. Lo anterior se debe a que pueden generar muchos grupos con un solo objeto, siendo de poca utilidad dicha agrupación para ellos. Por esta razón, dichos criterios agrupacionales no son manejados en este trabajo de tesis.

### 3.3.- Sistemas que incorporan herramientas de Reconocimiento de Patrones

Actualmente, se han desarrollado sistemas computacionales que incorporan herramientas del Reconocimiento de Patrones. Sin embargo, ninguna de las existentes satisfacen por completo las necesidades de los expertos como son (Ambiente Windows, lectura de datos en formatos de Excel, algoritmos del RLCP, resultados de forma gráfica y tabular, entre otros). Ejemplos de estos sistemas son: PROGNOSIS [RuJ+ 1992], GINKO [DeM+ 2003], MCLUST [FrC+ 1999], CLUTO [ZhY+ 2001], BANG [ScE+ 1996], KMINFLU [MiG+ 1995], CLUSTER [CLUSTER]. Pero también existen sistemas que incorporan Reconocimiento de Patrones para minería de datos, por mencionarlos: INLEN [KaK+ 1991], Clementine [ISCLE], MineSet (SGI) [SGRAPH], y Explora [K1W 1996], entre otros.

#### 3.3.1.- Sistema PROGNOSIS

El sistema PROGNOSIS [RuJ+ 1992], es un sistema para expertos en Geociencias que proporciona a los especialistas de estas ramas los instrumentos indispensables para la solución de problemas de clasificación con aprendizaje, con aprendizaje parcial y sin aprendizaje, trabaja con variables cualitativas y cuantitativas que admiten ausencia de información.

Uno de los aspectos novedosos de este sistema es que de acuerdo a la experiencia de sus creadores, y de los especialistas en estas áreas han determinado un conjunto de criterios de comparación y funciones de semejanza entre objetos. Aunque esto no quiere decir que el sistema esté restringido a un número determinado de criterios de comparación y funciones de semejanza. Sin embargo este sistema tiene las siguientes limitaciones:

- La interpretación del dendograma depende de la experiencia del usuario para poder realizarla.
- No cuenta con una módulo para preprocesamiento de los datos.
- No cuenta con un módulo para etiquetar los grupos y determinar cuál es el grupo de mayor riesgo.

#### 3.3.2.- Sistema GINKO

GINKO [DeM+ 2003] es un sistema que incorpora varias herramientas de análisis multivariante hacia usuarios no expertos en estadística, a través de una interfaz de usuario sencilla. Incorpora medidas de similaridad y distancia poco comunes en los paquetes estadísticos estándar, y que son frecuentemente utilizadas en ámbitos de aplicación, como

la ecología. También permite transformar matrices de similaridad en disimilaridad y viceversa. Es una aplicación orientada a la representación y clasificación de individuos a partir de datos multivariantes. Las principales técnicas que contempla son: algoritmos jerárquicos y de partición ( $k$ -means y fuzzy  $A$ -means). Análisis discriminante lineal, cuadrático y basado en distancias. En cuanto a técnicas de clustering, Ginko permite aplicar 3 modelos distintos de agrupamiento: clustering jerárquico aglomerativo, algoritmos partitivos y clustering no partitivo. Sin embargo se realizan las siguientes observaciones de GINKO:

- Presenta las mismas limitaciones que el Prognosis.
- Su enfoque principalmente es en la clasificación supervisada.
- A pesar de manejar medidas de similaridad, este programa esta enfocado en la clasificación en distancias para valores numéricos.

### **3.3.3.- Sistema MCLUST**

El sistema MCLUST [FrC+ 1999], es un software para agrupamiento y análisis discriminante realizado en Fortran. El sistema incorpora algoritmos de agrupamiento jerárquicos aglomerativos. Mclust incluye funciones para combinar agrupamiento jerárquico. Este sistema puede ser usado en sistemas operativos UNIX y Windows. Utiliza únicamente datos numéricos como rasgos para cada objeto, los cuales son representados como puntos en un espacio de representación. También Mclust puede ser utilizado para el agrupamiento de grandes conjuntos de datos, pero aún no se ha comprobado la eficiencia del sistema para estos. De igual forma, en este software, los resultados o agrupamientos no se presentan al usuario de forma bien definida, en otras palabras, que objetos pertenecen a cada grupo. Además, no se proporciona alguna información relevante en cuanto a los grupos obtenidos. Sin embargo las observaciones de este sistema son las siguientes:

- Utiliza únicamente datos numéricos en las descripciones de los objetos.
- No cuenta con una forma gráfica de presentación de resultados adecuada.
- Implementa algoritmos jerárquicos.

### 3.3.4.- Sistema CLUTO

CLUTO [ZhY+ 2001], es un sistema para el agrupamiento de conjuntos de datos de alta y baja dimensión, así como el análisis de las características de los agrupamientos que este genere. Cluto es adecuado para organizar los agrupamientos de datos generados en diversas áreas de aplicación incluyendo recuperación de información, WEB, GIS, ciencia y Biología. Cluto implementa tres diferentes tipo de algoritmos de agrupamiento que pueden operar juntos directamente sobre el espacio de representación de las características de los objetos o sobre la matriz de semejanza de los objetos. Los algoritmos que implementa CLUTO son algoritmos basados en partición, aglomerativos y partición de grafos. Los algoritmos basados en partición y aglomerativos, son factibles para encontrar grupos ante todo globulares

Se observa que este sistema cuenta con la facultad de utilizar medidas de similaridad y disimilaridad, puede utilizar matriz de semejanza o utilizar las descripciones de los objetos, pero al igual que los demás sistemas aquí planteados, para obtener agrupamientos libres solo utilizan algoritmos jerárquicos aglomerativos, esto trae consigo mostrar resultados en dendogramas únicamente. También. Principalmente el de proporcionar al usuario un ambiente amigable y que los resultados pueden ser visibles de forma gráfica como lo es el mapa del Estado de Hidalgo en donde se distinguen perfectamente los grupos con sus municipios correspondientes. Las observaciones a este software son:

- No cuenta con algún tipo de almacenamiento de resultados para poder ser utilizados posteriormente.
- Utiliza diferentes funciones de similaridad y distancia: Distancia euclidiana, coseno, coeficiente de correlación, extensión de Jaccard.
- Hace uso de métodos aglomerativos como: Conexión simple, conexión completa, UPGMA.
- Sus resultados son presentados también en un dendograma.

### 3.3.5.- Sistema BANG

El sistema BANG [ScE+ 1996], presenta un nuevo análisis jerárquico para conjuntos grandes de datos. Utilizando el método de agrupamiento de BANG y algoritmos de agrupamiento jerárquico de rejillas. Este método utiliza una rejilla multidimensional en la cual, cada objeto es representado por un punto. Además, estos objetos buscan ser agrupados en blocks rectangulares, la estructuración de los objetos esta en función de encontrar el centroide de estos blocks y unirlos a través de un algoritmo de agrupamiento de vecindad topológica. Este algoritmo BANG fue implementado sobre el sistema operativo UNIX usando X11 para proveer al usuario un ambiente agradable y fácil de usar. Las plataformas en las que el sistema BANG puede ejecutarse son Linux, Sun y Workstations HP. Este sistema ofrece lo siguiente: Permite mantener un control sobre el conjunto de datos a analizar.

Muestra la representación gráfica de los agolpamientos de la estructura del BANG y el

estado del proceso de agrupamiento. Así como también, muestra, cuál es el centro de cada grupo, un dendograma en el que se muestra una historia de la unión de los objetos.

Por otro lado, este sistema utiliza presenta las siguientes limitaciones:

- Únicamente utiliza medidas de disimilaridad como es la distancia euclidiana.
- El objetivo principal de este sistema es el de obtener un método que genere de forma más rápida los agrupamientos que otros algoritmos.
- Una vez más, este sistema no es capaz de mostrar alguna información relevante de los grupos formados.
- Este sistema no es capaz de etiquetar los grupos formados para conocer cuál es el más significativo en función de un riesgo calculado.

### **3.3.6.- Programa KMINFLU**

El programa KMINFLU [MiG+ 1995], como su nombre lo indica, solamente implementa el algoritmo &-means para datos numéricos. Se ejecuta en modo MS-DOS, este programa es completamente dependiente de los valores para los parámetros siguiente: se debe especificar correctamente el tamaño del conjunto de datos, el cual esta limitado a solo 550 objetos. Otro valor necesario es el número de variables, las cuales también están limitadas a 30. Además, se necesita obviamente el número de clases o grupos a formar. El programa obtiene los resultados y son almacenados en un archivo externo al programa con extensión. DAT. Esto es, para poder visualizar los resultados generados, es necesario editar dicho archivo e interpretar los resultados. Otro dato necesario es el de especificar al programa cuales serán las semillas iniciales para comenzar el agrupamiento. Una de las capacidades de este programa es que lee el conjunto de datos desde un archivo externo al programa. En resumen, sus limitaciones son las siguientes:

- Únicamente utiliza el algoritmo A:-means para datos numéricos.
- Utiliza datos numéricos.
- Su ejecución es en modo MS-DOS.
- No presenta resultados de forma gráfica.
- Es demasiado restrictivo en cuanto al número de objetos y variables.
- Requiere que el usuario introduzca el número de variables.

### **3.3.7.- Programa CLUSTER**

El programa CLUSTER 3.0 [CLUSTER] provee al usuario una interfaz gráfica para acceder a rutinas de agrupamiento. Este programa es usado por Windows, Mac Os X, y Linux/Unix. Cluster 3.0 es una versión mejorada de Cluster La principal mejora consiste en implementar el algoritmo A>means el cual ahora incluye múltiples pruebas para obtener la mejor solución en cuanto a agrupamiento de datos. Esta rutina incluye representación de los datos en geometrías rectangulares en dos dimensiones. Además, la distancia euclidiana y la de manhattan (city-block) son incluidas en esta nueva versión de Cluster. Sin embargo muestra las siguientes limitaciones:

- Los resultados también son presentados en un dendograma
- Usa datos numéricos

- Solo utiliza funciones de distancia.

Existen otros ambientes computacionales que incorporan herramientas del Reconocimiento de Patrones. Pero cuyo desempeño no es mejor que los mencionados anteriormente, o simplemente se enfocan en áreas de análisis de imágenes y señales. Algunos de estos sistemas son por mencionar: MADIP, EXPARAM.

Por esta razón, se considera necesario el desarrollo de una nueva herramienta computacional hecha a la medida de las necesidades de los usuarios y de las necesidades para la estratificación de municipios en Zonas de riesgo para los SSH. Que incorpore técnicas y algoritmos del RLCP. Desarrollada en un ambiente gráfico con herramientas totalmente diseñadas en completo acuerdo con el experto para que el pueda tomar decisiones en función de las estructuraciones observadas.

# Capítulo 4

## ANÁLISIS Y DISEÑO DE LA HERRAMIENTA (SAD)

En el presente capítulo se describen las fases de análisis y diseño de la herramienta computacional SAD (*Sistema de Agrupamiento de Datos*).

### 4.1.-Antecedentes

En materia de atención a la salud, en los últimos años la mayoría de los países incluyendo el nuestro, han establecido el compromiso de adoptar la estrategia de Atención Primaria de la Salud (APS). Suscrito en Alma - Ata, Rusia 1978 [SSA+ 1994], adaptado a las condiciones de cada país.

El Modelo de Atención a la Salud para Población Abierta (MASPA) fundamenta sus acciones en esta estrategia, entendida como: la asistencia sanitaria esencial, basada en métodos y tecnologías prácticas científicamente fundadas y socialmente aceptadas, puesta al alcance de todos los individuos y de todas las familias a un costo que el país pueda soportar en todas y cada una de sus etapas de desarrollo, cumpliendo con responsabilidad y participación [SSA+ 1999, PNS+ 2001, FMS+ 1994].

#### 4.1.1- Programa "Atención Integral"

Los Servicios de Salud de Hidalgo (SSH) se vieron en la necesidad de aplicar, adecuar y desarrollar el MASPA [SSA+ 1994] en el Estado de Hidalgo, buscando adecuarse a las necesidades y problemas de cada uno de los 84 municipios que conforman al Estado ubicando los servicios lo más cerca posible del lugar donde se demanden, trayendo con ello el surgimiento del programa "Atención Integral". Este programa incorpora acciones como: promoción, prevención, diagnóstico, tratamiento y rehabilitación; considerando que la salud y la enfermedad son fenómenos dinámicos determinados por factores físicos, biológicos, psicológicos y sociales, por lo que la atención es dirigida al individuo, a la familia y a la comunidad [SSA+ 1999, FMS+ 1994].

El principal objetivo de este programa es lograr la óptima organización y el funcionamiento

de los servicios estatales de salud para asegurar a la población la disponibilidad y acceso a servicios de salud de calidad. Con plena satisfacción de los usuarios y de los prestadores de servicios. La estructura general de este programa se puede observar en la figura 6 [KuS 1989].



Figura 6. Estructura general del programa " Atención Integral".

El grupo de edad del menor de 5 años comprende a los niños y niñas de 0 a 4 años de edad. El grupo escolar se enfoca en los niños y niñas entre los 5 y 14 años. El grupo del adolescente contempla a los hombres y mujeres entre los 10 y 19 años. El siguiente grupo el de la mujer solo se enfoca en las mujeres de 20 a 45 años. Así como el grupo de edad del adulto comprende a los hombres de 20 a 45 años. Finalmente, el grupo del adulto mayor a los hombres y mujeres mayores de 45 años. Los indicadores que conforman cada grupo se pueden visualizar a detalle en el Anexo B. A través de la promoción de la salud, el modelo, no se limita a reparar los daños [KuS+ 1993], se pretende que las estrategias que se impulsan permitan identificar riesgos a la salud e intervenir antes de que exista el daño.

Este programa tiene un módulo de "Estratificación de Municipios en zonas de riesgo" las cuales son: (Muy alto, Alto, Medio y Bajo). Esta estratificación de municipios se realiza para los 6 grupos de edad (Menor de 5 años, Escolar, Adolescente, De la mujer, Del adulto y Del adulto mayor) [SSH 2000] como se ilustran en el Anexo A, donde cada grupo está descrito por un número fijo de indicadores de daños a la salud [CON 1993] [SiN 1993a] los cuales pueden observarse en el Anexo B.

Es de suma importancia mencionar que gracias a la información proporcionada por los SSH se utilizan datos confidenciales del grupo de edad del menor de 5 años, el cual es nuestro caso de estudio. Es importante decir que por esta razón en esta tesis no se muestran todos los datos originales con fin de agradecer la confianza que nos han otorgado y que dichos datos no puedan ser utilizados para otros fines que no sean los permitidos en este trabajo.

#### 4.1.2.- Necesidad de la estratificación de municipios por los SSH

Normalmente las instituciones estatales de salud no cuentan con los recursos necesarios para atender todos y cada uno de los problemas en cada uno de los municipios que integran al Estado de Hidalgo [SiN 1993b]. Así también, se ha comprobado que la atención unidireccional de problemas de salud lleva consigo dispendio de recursos y bajo logro de objetivos [SSH 2000]. Por estas razones, los SSH han instrumentado una metodología de estratificación de municipios en zonas de riesgo basada en la "Distribución por Cuartiles" [SSH 2000], en función de un conjunto de indicadores de daños a la salud para cada uno de los grupos de edad, sociodemográficos, factores de riesgo y factores protectores [SiN 1993a]. La meta principal de dicha estratificación consiste en integrar estos indicadores por grupo de edad, para todos y cada uno de los municipios, y así apoyar en la toma de decisiones para orientar estrategias, actividades y recursos [SiN 1993b].

## 4.2.- Análisis del SAD

Los expertos requieren de una herramienta completamente diseñada en función de sus necesidades y de las de los SSH. Por ello es importante el considerar y evaluar los requerimientos necesarios para proporcionar una herramienta confiable y fácil de utilizar. Muchos de los sistemas existentes que incorporan técnicas del Reconocimiento de Patrones que más adelante se mencionan, no cubren al 100 por ciento las necesidades y requerimientos de los usuarios de los SSH. Por esta razón, es necesario el diseñar y construir una herramienta que cumpla con todas sus exigencias. El modelo conceptual de la herramienta SAD se puede visualizar en la figura 7.

### 4.2.1- Definición de requisitos funcionales

Los requisitos funcionales de los expertos en salud así como lo que esperan de la herramienta son los siguientes:

- La herramienta debe ser completamente bajo ambiente gráfico (específicamente bajo Windows).
- Cargar y almacenar archivos utilizando diálogos (los ofrecidos por Windows).
- La base de datos fuente debe estar en una hoja de cálculo en Excel, sin límite de municipios e indicadores.
- Así mismo, la herramienta debe leer los datos y etiquetas directamente de esta hoja de cálculo.
- La herramienta debe obtener de forma automática las tasas, índices y porcentajes necesarios para la estratificación.
- Utilizar una base de datos para llevar un control de las estratificaciones de municipios generadas, y así poder realizar consultas en ella.
- Mostrar el mapa del Estado de Hidalgo coloreado en función de los grupos obtenidos.

- La forma en que se elegirán los colores para el mapa deben ser de forma manual o de forma automática.
- Poder utilizar cualquier mapa para los diferentes Estados de la República Mexicana.
- Obtener estratificaciones de municipios de manera automatizada (agrupamientos).
- Consultar estratificaciones ya almacenadas en la base de datos.
- Poder realizar una sencilla comparación entre estratificaciones.
- Mostrar resultados en tablas, reportes y mapas.
- Mostrar que indicadores influyeron más para el promedio de riesgo final del grupo.
- Que la herramienta pueda ser utilizada para cualquier grupo de edad.

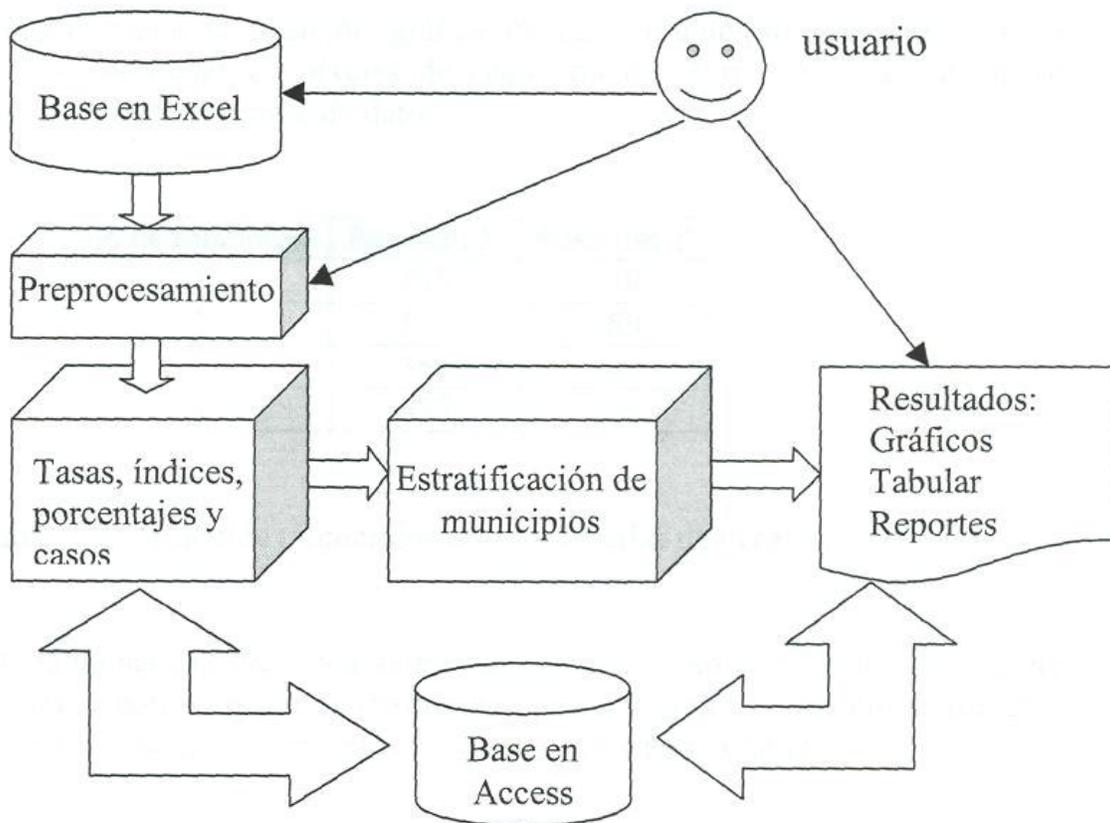


Figura 7. Modelo conceptual del SAD

Es importante mencionar que para el desarrollo tanto del análisis como del diseño del SAD fue necesaria una constante interacción con el experto en salud. Esta interacción comenzó desde la definición del problema, las necesidades de los expertos en salud, el manejo de los datos de los expertos, así como la definición de las formas para comparar los municipios. También hubo una constante interacción al momento de generar los resultados tanto parciales como finales y presentárselos a los expertos para su valoración y validación,

concluyendo con el aprendizaje y manejo de la herramienta SAD por el experto en salud.

#### 4.2.2- Definición de la base de datos

Una de las principales necesidades de los SSH, es el de disponer prácticamente cuando se requiera de las estratificaciones generadas en otros periodos. Así como el de almacenar las nuevas estratificaciones obtenidas por la herramienta. De igual forma, es necesario el disponer de los mapas que se obtuvieron en su momento. La herramienta también necesita esta base para apoyarse en los diferentes procesos que conlleva la realización de la estratificación de municipios. Como son: almacenar las coordenadas de los píxeles de los municipios del Estado de Hidalgo en el mapa para poder entregar al usuario dicho mapa coloreado en función de los grupos y municipios correspondientes a cada grupo. Otra de las necesidades es el de capturar en la base los criterios de comparación y fórmulas para el preprocesamiento de información de cada uno de los indicadores. Así como las descripciones de los municipios.

Para que el sistema conozca la posición gráfica de cada municipio (ver tabla 1), es necesario capturar las coordenadas en píxeles de cada uno de ellos. A continuación se muestra la tabla *coordenadas* de la base de datos.

No de municipio	Posición X	Posición Y
1	498	389
2	577	389
3	356	349
4	533	330

Tabla 1. Relación de municipios en coordenadas de píxeles.

Se observa que esta tabla hace referencia a un número de municipio, por lo cual se requiere de una tabla *municipios* para conocer la descripción del municipio y ser posteriormente utilizada en cualquier proceso de la herramienta esto se observa en la tabla 2.

No de municipio	Descripción
1	ACATLAN
2	ACAXOCHITLAN
3	ACTOPAN
4	AGUA BLANCA DE ITURBIDE

Tabla 2. Relación de descripciones de municipios.

En el procedimiento de preprocesamiento de datos, es necesario capturar la fórmula correspondiente para cada indicador (ver tabla 3). Y así, la herramienta conocerá que tipo de dato obtendrá, ya sea índices, tasas o porcentajes en la *tabla formulas*.

Indicador	Fórmula	Constante	Población
1	1	1000	3
2	1	10000	1
3	1	1000	3

Tabla 3. Relación de fórmulas para cada indicador.

Para capturar los criterios de comparación a cada indicador (ver tabla 4), así como almacenar la aportación informativa de cada uno, y el caso para evaluar el riesgo de salud para cada indicador es necesaria una tabla denominada *indicador*.

Indicador	Criterio	Valor Usuario	Factor	Aportación informativa	Tipo de Variable	Caso de riesgo	Mayor valor indicador
1	3	0	1907.232	0.519	num	1	2049.193
2	3	0	49.505	0.694	num	1	49.505
3	3	0	3841.944	0.475	num	1	4164.12
4	3	0	16.4	1.177	num	1	16.4

Tabla 4. Relación de criterios de comparación por indicador.

Es necesario almacenar las estratificaciones generadas por la herramienta (ver tabla 5), con el objetivo de poder realizar consultas posteriores cuando el usuario las requiera. Por este motivo, la tabla *municipios grupos* permite llevar un control de estas estratificaciones de municipios.

No de Municipio	Grupo al que pertenece	Archivo	Fecha de creación	Color asignado	Riesgo calculado	Algoritmo utilizado
1	1	Menor de 5 años 2000.xls	31/01/2004	157	0.75	comp2
1	1	Menor de 5 años 2000.xls	31/01/2004	255	0.75	conex
1	1	Menor de 5 años 2000.xls	31/01/2004	255	0.75	kmean
1	4	Menor de 5 años 1999.xls	12/10/2004	65535	0	SSH
2	1	Menor de 5 años 1999.xls	12/10/2004	238	0	SSH
2	2	Menor de 5 años 2000.xls	31/01/2004	255	0.85	comp2
2	2	Menor de 5 años 2000.xls	31/01/2004	16761088	0.85	conex
2	2	Menor de 5 años 2000.xls	31/01/2004	16761088	0.85	kmean

Tabla 5. Relación de estratificaciones almacenadas de municipios.

Además, la herramienta proporcionará reportes impresos, por lo tanto, existen en la base de datos tablas auxiliares para la generación de dichos reportes pero que no son necesarios en los procesos de estratificación que realiza la herramienta (ver tablas 6,7,8,9 y 10). Estas tablas son las siguientes: *imprimir indicador*, *reporte de estratificación almacenada*, *reporte de objetos por grupo*, *reporte de riesgo por indicador* y *reporte de comparación entre estratificaciones*.

La tabla *imprimir indicador* auxiliar para obtener el reporte de tasas, índices, porcentajes o casos por indicador tiene la siguiente estructura.

No. Municipio	Descripción	Tasa	Grupo de edad	Indicador
24	HUASCA DE OCAMPO	49.505	Menor de 5 años 1999.xls	Mort.IRAS < 1 año
12	ATOTONILCO EL GRANDE	28.834	Menor de 5 años 1999.xls	Mort.IRAS < 1 año
5	AJACUBA	27.473	Menor de 5 años 1999.xls	Mort.IRAS < 1 año
59	TOCOZAUTLA	26.786	Menor de 5 años 1999.xls	Mort.IRAS < 1 año

Tabla 6. Lista de tasas para el indicador Mortalidad por IRAS en menores de 1 año.

La tabla *reporte estratificación almacenada* se utiliza para poder generar el reporte de alguna estratificación previamente almacenada en la base de datos. Su estructura es la siguiente:

Grupo de edad	Algoritmo	Fecha	No Municipio	Descripción	Grupo	Riesgo
Menor de 5 años 1999.xls	SSH	12/10/2003	1	ACATLAN	4	0
Menor de 5 años 1999.xls	SSH	12/10/2003	2	ACAXOCHITLAN	1	0
Menor de 5 años 1999.xls	SSH	12/10/2003	3	ACTOPAN	2	0
Menor de 5 años 1999.xls	SSH	12/10/2003	4	AGUA BLANCA DE ITURBIDE	2	0

Tabla 7. Estratificación generada para el grupo menor de 5 años.

De igual forma, la tabla *reporte objetos grupo* se utiliza para generar el reporte con toda la información relevante de cada uno de los grupos de la estratificación.

No. Municipio	Descripción	Grupo	Riesgo	Grupo de edad	Algoritmo	Promedio de riesgo	Total de objetos
34	LOLOTLA	1	1	Menor de 5 años 1999.xls	conexas	0.76	38
60	TENANGO DE DORIA	1	0.94	Menor de 5 años 1999.xls	conexas	0.76	38
40	LA MISION	1	0.93	Menor de 5 años 1999.xls	conexas	0.76	38
73	TLANCHINOL	1	0.92	Menor de 5 años 1999.xls	conexas	0.76	38

Tabla 8. Municipios pertenecientes al grupo 1.

Una de las aportaciones de esta herramienta, es el de ofrecer a los expertos un listado de la influencia de los indicadores para el cálculo del promedio de riesgo final del grupo. La tabla *reporte riesgo* permite generar el reporte correspondiente.

Indicador	Descripción	Riesgo indicador	Grupo de edad	Grupo	Algoritmo
1	Detección de EPV	1	Menor de 5 años 1999.xls	1	conexas
10	Mort. IRAS < 1 años	0.17	Menor de 5 años 1999.xls	1	conexas
11	Mort. IRAS 1 - 4 años	0.16	Menor de 5 años 1999.xls	1	conexas
12	Mort. Desnu. 1 – 4 años	0.12	Menor de 5 años 1999.xls	1	conexas
13	Mort. EDAS 1 – 4 años	0.09	Menor de 5 años 1999.xls	1	conexas
14	Mort. Acciden. 1 – 4 años	0.08	Menor de 5 años 1999.xls	1	conexas
15	Cobertura de vac. < 1 año	0	Menor de 5 años 1999.xls	1	conexas

Tabla 9. Lista de indicadores en función de la aportación de riesgo.

Por último, también es necesaria una tabla que almacene la comparación entre dos estratificaciones en el módulo de análisis para generar el reporte correspondiente, esta tabla se llama *reporte 1*.

No municipio	Descripción	Riesgo1	Riesgo2	status	Periodo 1	Periodo 2
1	ACATLAN	0	0.75	Subió	Menor de 5 años 1999.xls	Menor de 5 años 2000.xls
2	ACAXOCHITLAN	0	0.85	Subió	Menor de 5 años 1999.xls	Menor de 5 años 2000.xls
3	ACTOPAN	0	0.48	Subió	Menor de 5 años 1999.xls	Menor de 5 años 2000.xls
4	AGUA BLANCA DE ITURBIDE	0	0.91	Subió	Menor de 5 años 1999.xls	Menor de 5 años 2000.xls
5	AJACUBA	0	0.38	Subió	Menor de 5 años 1999.xls	Menor de 5 años 2000.xls

Tabla 10 . Comparación entre dos estratificaciones.

El modelo relacional de la base de datos que se puede visualizar en la figura 8 se utiliza para definir la forma lógica de la base, especificando un conjunto de tablas, algunas de las cuales tienen claves comunes. Este modelo permite definir las relaciones entre los datos elementales sin considerar la organización física de la base de datos.

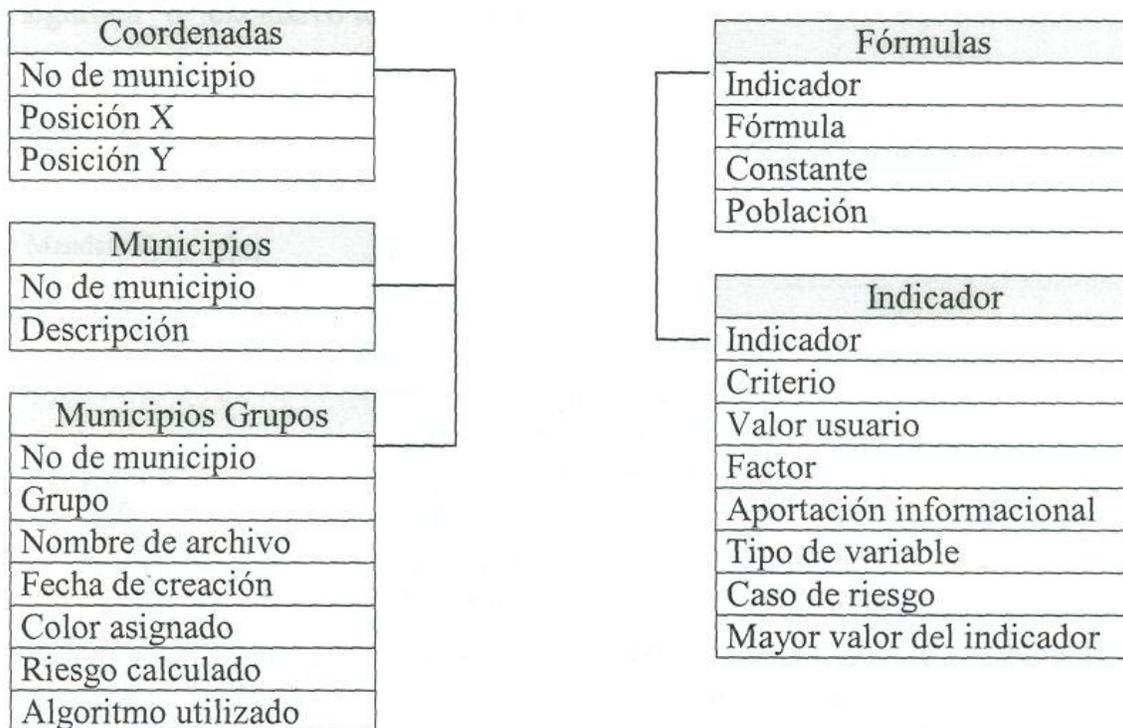


Figura 8. Modelo relacional de la base de datos.

#### 4.2.3- Especificaciones técnicas

Los requisitos técnicos para esta herramienta en función del hardware y software son los siguientes:

- Windows 98, Me y XP
- 64 Mb en RAM o superior
- 266 Mhz de velocidad o superior

### 4.3.- Diseño del SAD

En esta sección se exponen los modelos diseñados, que ayudan a la implementación de la herramienta. Es importante mencionar que los modelos son diseñados partiendo de los elementos establecidos en la fase de análisis. El siguiente diagrama de flujo (ver figura 9) documenta cómo los datos de entrada se transforman en datos de salida. El símbolo © significa "o" exclusivo lógico (XOR).

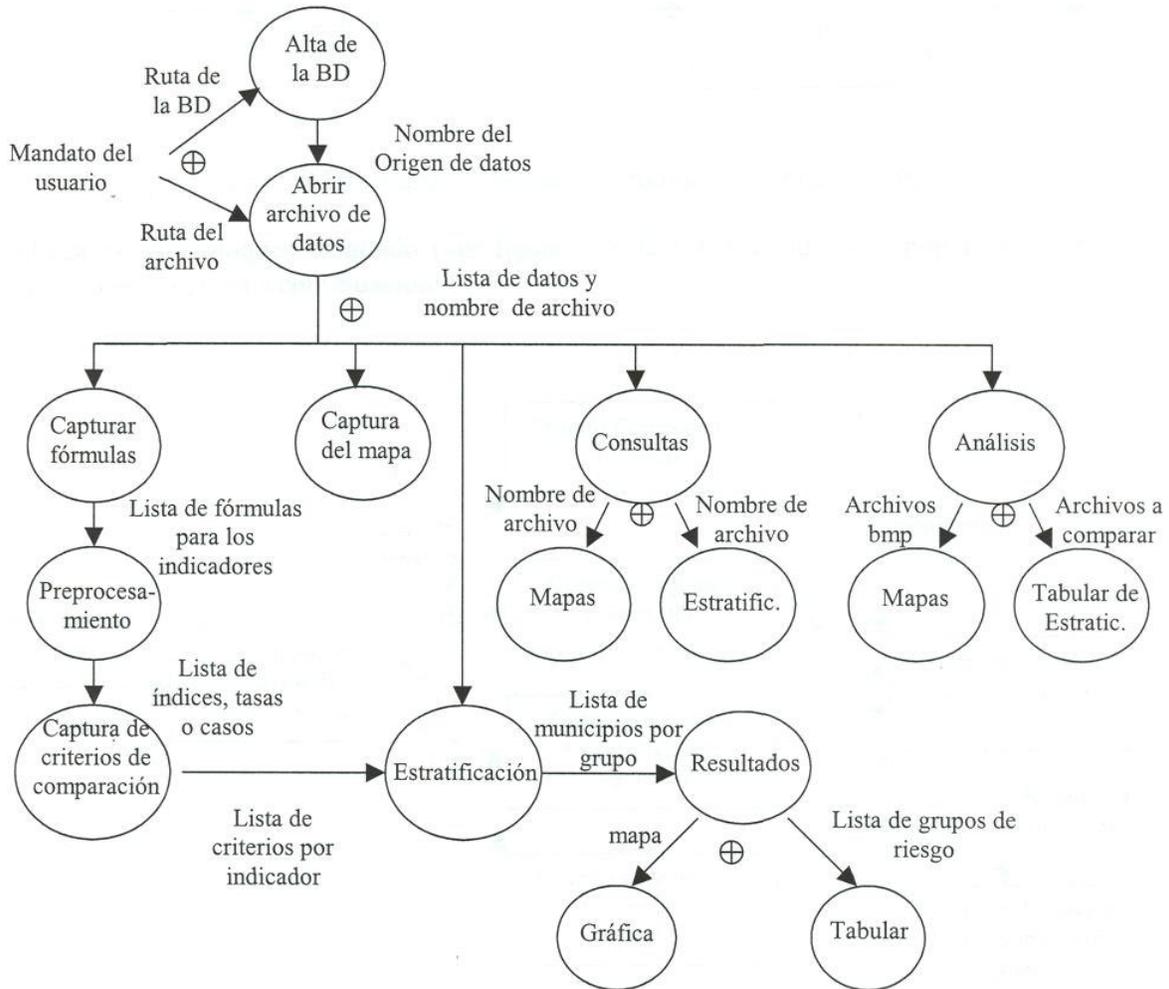


Figura 9. Diagrama de flujo de datos del SAD

Sin duda, uno de los módulos más importantes de esta herramienta es el de la estratificación de municipios en zonas de riesgo. El cual, proporciona un listado de grupos ordenados de forma descendente en función del promedio de riesgo de salud final para cada grupo. A continuación se muestra un diagrama a bloques en la figura 10 del proceso de estratificación.

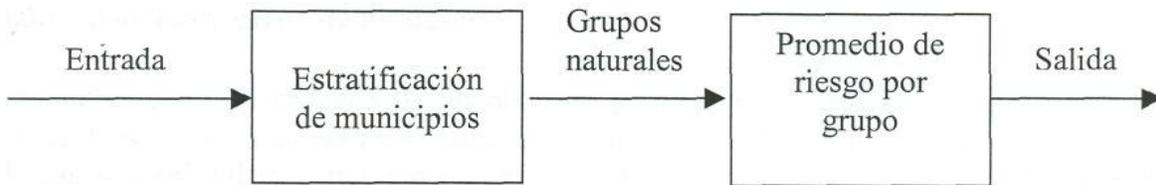


Figura 10. Diagrama a bloques del módulo de estratificación.

El diagrama a bloques detallado (ver figura 11) de las diferentes componentes de este proceso se visualiza a continuación.

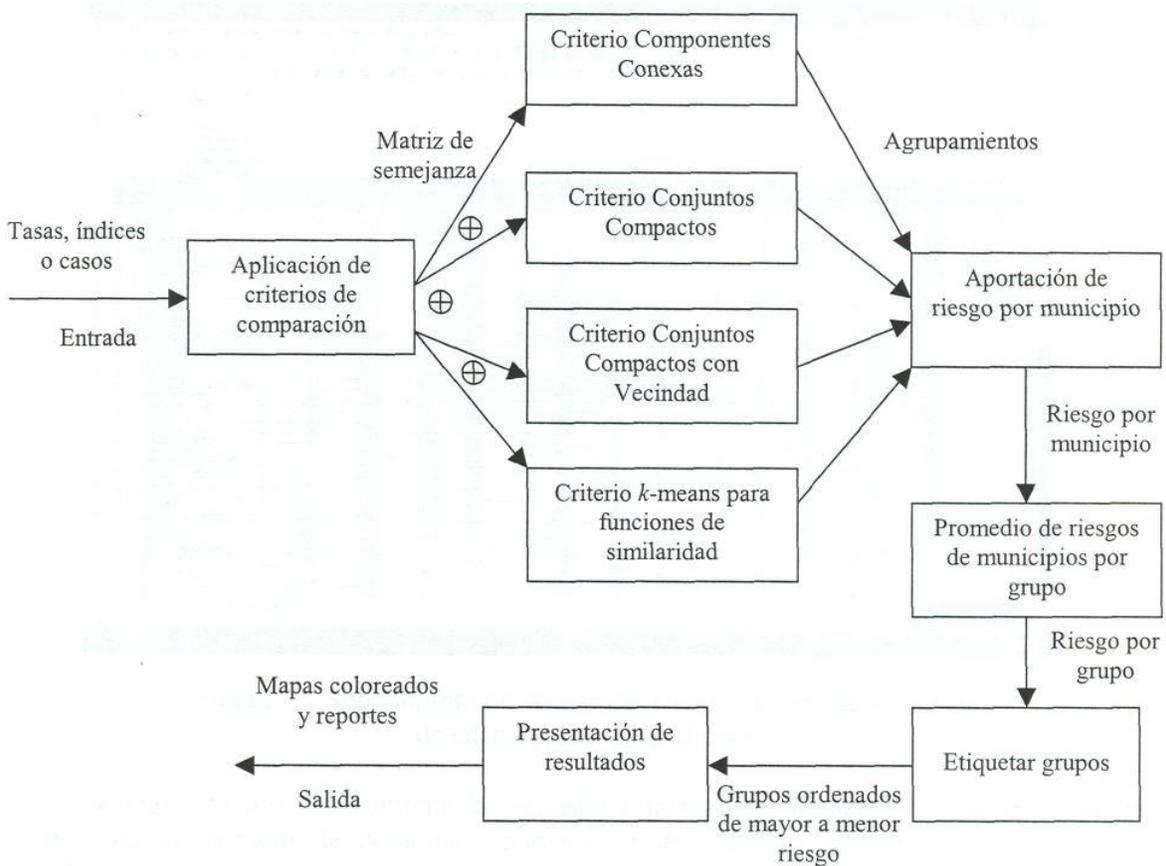


Figura 11. Diagrama detallado a bloques del módulo de estratificación

Es importante mencionar que debido al alto número de agolpamientos que genera el criterio

agrupacional conjuntos compactos. Fue necesario diseñar e implementar un nuevo criterio denominado conjuntos compactos con vecindad. Este algoritmo genera un número de agrupamientos entre las conexas y los compactos, pero sin perder de vista el objetivo de este trabajo que es el de obtener agrupamientos basados en semejanzas entre objetos.

### 4.3.1.- Análisis y carga de los datos

Los datos que utilizan los SSH los obtienen de diferentes fuentes, entre ellas: INEGI, de los centros de salud de las localidades entre otros. Estos datos son típicamente el número de casos (morbilidad) o número de defunciones (mortalidad). También se obtienen datos como el grado de marginalidad que tiene cada municipio en el Estado de Hidalgo, población de individuos para cada indicador y cada municipio, entre otros. Los indicadores proporcionados por los expertos en salud se encuentran definidos por tres tipos de variables: boléanos,  $k$  - valentes y numéricos. Y los valores que toman los indicadores numéricos se presentan en dominios diferentes.

Una de las necesidades de los expertos fue el de concentrar toda esa información en una hoja de cálculo en Excel y que el sistema fuera capaz de leer esos datos con un formato sencillo y fácil de manejar. Este formato se puede ver en la figura 12.

The screenshot shows a Microsoft Excel spreadsheet with the following data columns: No, Municipio, poblacion1, poblacion2, poblacion3, Morb. IRAS <1 año, Mort. IRAS <1 año, Morb. IRAS 1-4 años, Mort. EDAS <1 año, Mort. EDAS 1-4 años, Deteccion de EPV, Casos EPV, Mort. Deseu. <1 año, Mort. Accidentes <1 año, Cobertura de vac. <1 año, Cobertura Antisars. 14 meses, and Marginalidad. The rows list various municipalities such as ACATLAN, ACAXOCHITLAN, ACTOPAN, AGUA BLANCA DE TURBIDE, AJACUBA, ALFAJAYUCAN, ALMOLLOYA, APAN, EL APENAL, ATITALAQUA, ATLAPECO, ATOTONILCO EL GRANDE, ATOTONILCO DE TULA, CALAJAL, CARDONAL, CUAUTEPEC DE HINOJOSA, CHAPANTONGO, CHAPULHUACAN, CHILCUAUTLA, EL DOCHITLAN, EMILIANO ZAPATA, EPAZUYUCAN, FRANCISCO MADRID, HUASCA DE OCAMPO, HUAUTLA, HUAZALINGO, HUEHUETLA, HUEHUETLA, HUAUTLA DE REYES, HUCHAPAN, Ixmiquilpan, JACALA DE LEZAMA, JALTOCAN, JUAREZ HIDALGO, LOCLITA, and METEPEC.

Figura 12. Subconjunto de la base de datos en Excel para el grupo de edad del menor de 5 años.

En general este formato contiene la siguiente información: número de filas (número de municipios), número de columnas (número de indicadores), número de columnas de población, así como descripciones de los municipios. Además de que el formato no se restringe a un número de filas y columnas fijas. En otras palabras, puede contener más o

menos municipios, así como agregar o suprimir indicadores. En las columnas de población la herramienta detecta automáticamente cuantas columnas se utilizarán para realizar el preprocesamiento de los datos, por lo tanto también el formato es flexible en cuanto al número de columnas para las poblaciones. La forma en la que la herramienta obtiene estos datos se puede observar en la figura 13. En esta imagen se observa la ventana de abrir archivo comúnmente utilizada por Windows, la cual proporciona al usuario una forma fácil y sencilla de acceder a los datos de los SSH.



Figura 13. Acceso a los datos de los SSH en Excel.

Para poder contar con una herramienta flexible, que permita obtener la estratificación de municipios de cualquier grupo de edad, se planeó y diseñó una base de datos en Access. Esta base permite llevar un historial de las estratificaciones obtenidas por la herramienta para realizar los análisis de resultados correspondientes en diferentes periodos. El proceso de alta para la base de datos en el sistema solo se realiza una sola vez por grupo de edad, y la forma de darla de alta es en el administrador de orígenes de datos ODBC, el cual se visualiza en la figura 14.

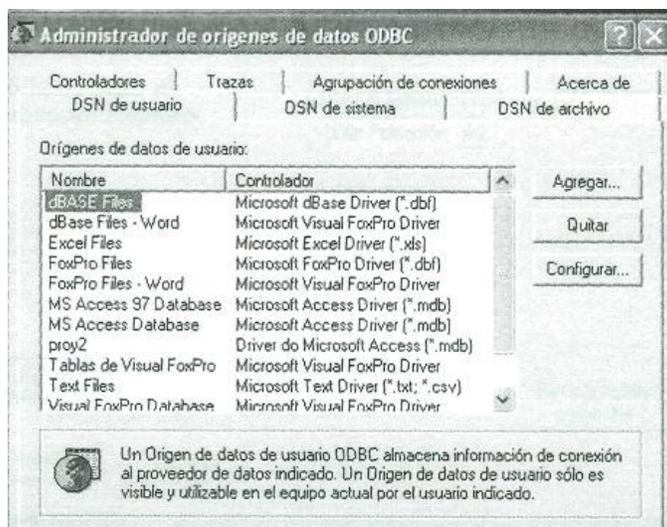


Figura 14. Módulo de alta para la base de datos en Access.

#### 4.3.2.- Preprocesamiento de los datos

Los datos requieren de un preprocesamiento para poder ser utilizados por los criterios de comparación. De hecho, en el procedimiento de estratificación manual que realizan los SSH también llevan acabo dicho preprocesamiento. Por ello los expertos requieren que la herramienta computacional cuente con la posibilidad de realizar esta tarea. Para el caso de indicadores que miden la morbilidad y la mortalidad se aplica una fórmula que da como resultado la tasa del indicador, y se compone de diferentes partes:

$$\frac{\text{No. Casos o No. Defunciones}}{\text{Poblacion}} \times k \text{ Habitantes} = \text{TASA}$$

Donde:

*k* Habitantes : Es una constante definida por un número de habitantes y que los expertos determinan el valor que puede tomar.

*No. Casos ó No. Defunciones:* Son datos ya almacenados en la hoja de Excel y que la herramienta los carga automáticamente.

*Población:* Es el número de individuos que se consideran en el indicador en cuestión.

*TASA:* Es la tasa obtenida para morbilidad o mortalidad según el indicador a evaluar.

Es de suma importancia mencionar que esta fórmula aplica únicamente a los indicadores que miden morbilidad y mortalidad. Para los demás indicadores no se les aplica fórmula alguna, únicamente se especifica que el dato que se cargó de Excel es el mismo que se utilizará para obtener la estratificación. La implementación computacional de este módulo se puede ver en la figura 15.

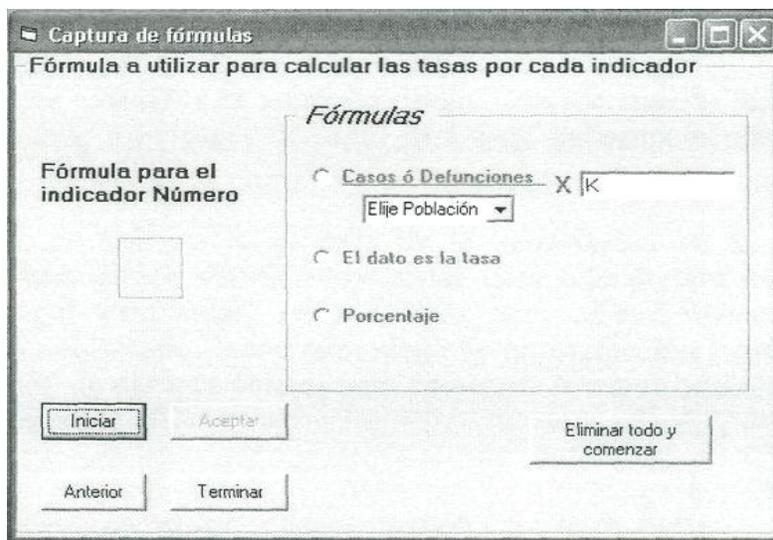


Figura 15. Captura de fórmulas para el preprocesamiento de datos para cada indicador. Las demás interfaces que conforman la herramienta son presentadas a lo largo de los capítulos 5 y 6. Las cuales fueron diseñadas en completa interacción con el usuario y el

experto, para lograr una herramienta a la medida de las necesidades y requerimientos de los usuarios.

#### 4.4.- Beneficios

Sin duda, esta herramienta traerá consigo importantes beneficios a los SSH, por mencionar algunos: reducir el tiempo de trabajo en la obtención de estratificaciones de municipios, contar con una herramienta computacional que automatice dichos procesos de estratificación, apoyo en la toma de decisiones, gestionar recursos a los municipios que más lo requieran, llevar un control de las estratificaciones obtenidas para un posterior evaluación, etc.

# Capítulo 5

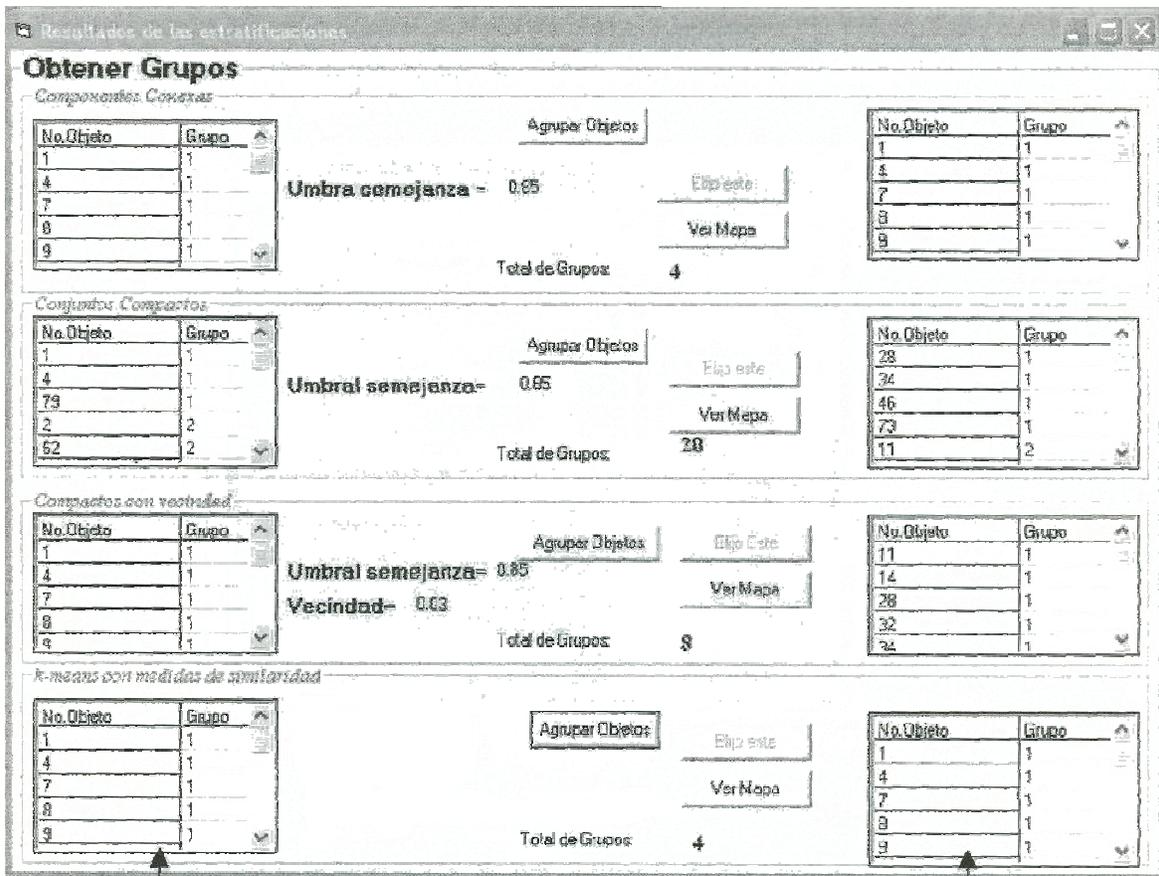
## DESARROLLO DEL SAD

Este capítulo se enfoca principalmente en el desarrollado y programación de los algoritmos y módulos del SAD (*Sistema de Agrupamiento de Datos*) para poder obtener las estratificaciones de los municipios. Así como las diferentes formas de presentación de resultados y de análisis de estratificaciones almacenadas. También se muestra el nuevo criterio agrupacional propuesto en esta tesis conjuntos compactos con vecindad.

### 5.1.- Programación e implementación de los criterios agrupacionales que utiliza la herramienta computacional

Para poder obtener estratificaciones de municipios en zonas de riesgo y ofrecer al usuario algoritmos del RLCP libres y restringidos se programaron e implementaron en esta herramienta los siguientes criterios agrupacionales: (componentes conexas, conjuntos compactos, conjuntos compactos con vecindad, y  $A>$ means para funciones de similaridad), y con ellos obtener agrupamientos (estratificaciones) de municipios y posteriormente etiquetar los grupos obtenidos en zonas de riesgo. Considerando la función de semejanza y las formas de evaluar los riesgos de salud para cada indicador como se anunciaron en el capítulo 2.

En la figura 16, se muestra la pantalla de la herramienta en la que se realizan los agrupamiento de municipios dividida en 4 zonas, cada zona pertenece a uno de los criterios agrupacionales aquí planteados, cada criterio agrupacional cuenta con botone para visualizar el mapa coloreado, botón para elegir la estratificación y almacenarla en la base de datos, y el botón de agrupar objetos para comenzar la estratificación de municipios para ese criterio agrupacional.



Agrupamientos naturales que arrojan los criterios

Agrupamientos ordenados en función al riesgo por grupo.

Figura 16. Pantalla de agrupamientos utilizando los criterios agrupacionales.

En la parte izquierda de la pantalla de resultados se visualizan los agrupamientos naturales de cada uno de los criterios agrupacionales. En la parte derecha se muestran los grupos ordenados en función al promedio de riesgo de salud por grupo, en otras palabras, están ordenados de forma descendente de mayor a menor riesgo de salud.

Además, la herramienta ofrece la opción de ver los municipios que pertenecen a un grupo así como la aportación de riesgo por municipio y el promedio de riesgo final por grupo. Esto se visualiza en la figura 17.

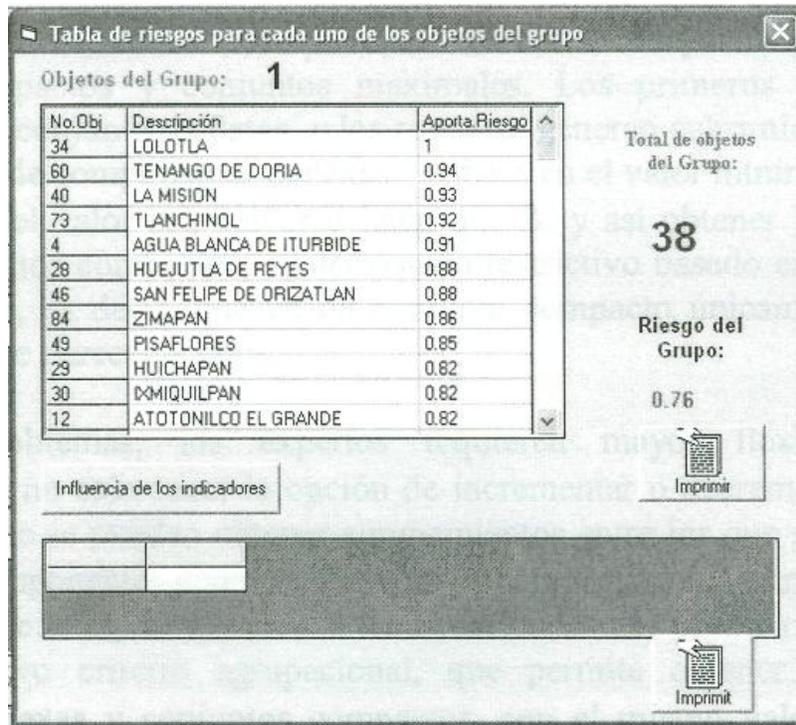


Figura 17. Información del grupo formado.

De igual forma, también es posible obtener el reporte de las características específicas de un grupo para poder ser analizado textualmente. Esto se visualiza en la figura 18.

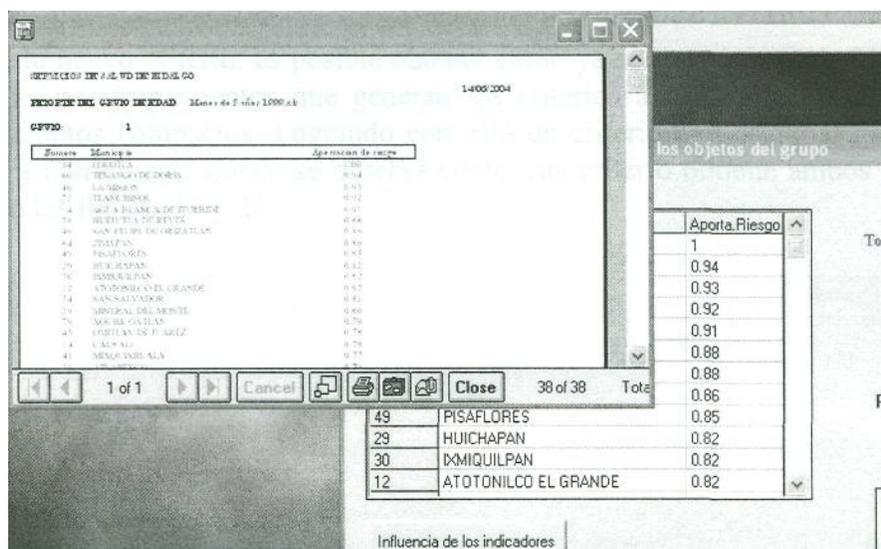


Figura 18. Reporte del análisis del grupo y sus aportaciones de riesgo.

## 5.2.- Programación del criterio agrupacional propuesto: Conjuntos Compactos con Vecindad

En el Reconocimiento Lógico Combinatorio de Patrones existen diversos criterios agrupacionales duros como: componentes conexas, conjuntos compactos, conjuntos fuertemente compactos y conjuntos maximales. Los primeros dos criterios obtienen particiones de un conjunto de datos, y los restantes generan cubrimientos en la colección de datos. El criterio de componentes conexas, se basa en el valor mínimo de semejanza el cual cumpla con que el valor sea mayor o igual que  $\beta_0$  y así obtener los grupos. Además, el criterio de conjuntos compactos, es demasiado restrictivo basado en la máxima semejanza entre dos objetos, es decir, que en un conjunto compacto únicamente se encuentran los objetos que más se parecen a ellos.

En algunos problemas, los expertos requieren mayor flexibilidad para generar agrupamientos, y no solo tener la opción de incrementar o decrementar el valor de  $\beta_0$ . En otras palabras, que se puedan obtener agrupamientos entre los que se obtienen al aplicar el criterio de las componentes conexas ó conjuntos compactos, con el mismo valor de  $\beta_0$ . En este trabajo de tesis, de acuerdo a las necesidades del problema y de los expertos, se propone un nuevo criterio agrupacional, que permita obtener estructuraciones entre componentes conexas y conjuntos compactos, con el mismo valor de  $\beta_0$  para los tres criterios agrupacionales.

El criterio propuesto, no solo obtiene subconjuntos de agrupamientos entre los criterios de componentes conexas y conjuntos compactos, usando el mismo valor para  $\beta_0$ . Además, este criterio con el valor apropiado de  $\beta_0$  puede obtener los agrupamientos que genera conjuntos compactos y componentes conexas. En este sentido, este nuevo criterio es una generalización de estos criterios de partición.

A través de este nuevo criterio, es posible obtener como ya se menciono anteriormente las estratificaciones o agrupamientos que generan los criterios agrupacionales componentes conexas y conjuntos compactos. Logrando con esto un criterio agrupacional general. Las imágenes de la herramienta donde se observa como este criterio obtiene ambos resultados se observan en las figuras 19 y 20.

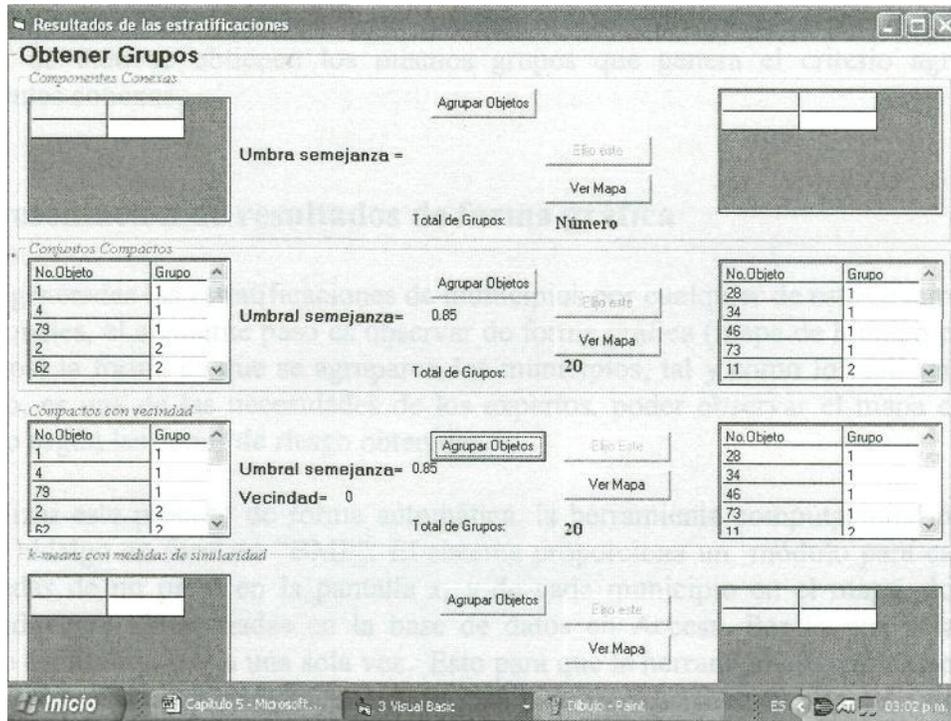


Figura 19. Obtención de conjuntos compactos utilizando el criterio de conjuntos compactos con vecindad.

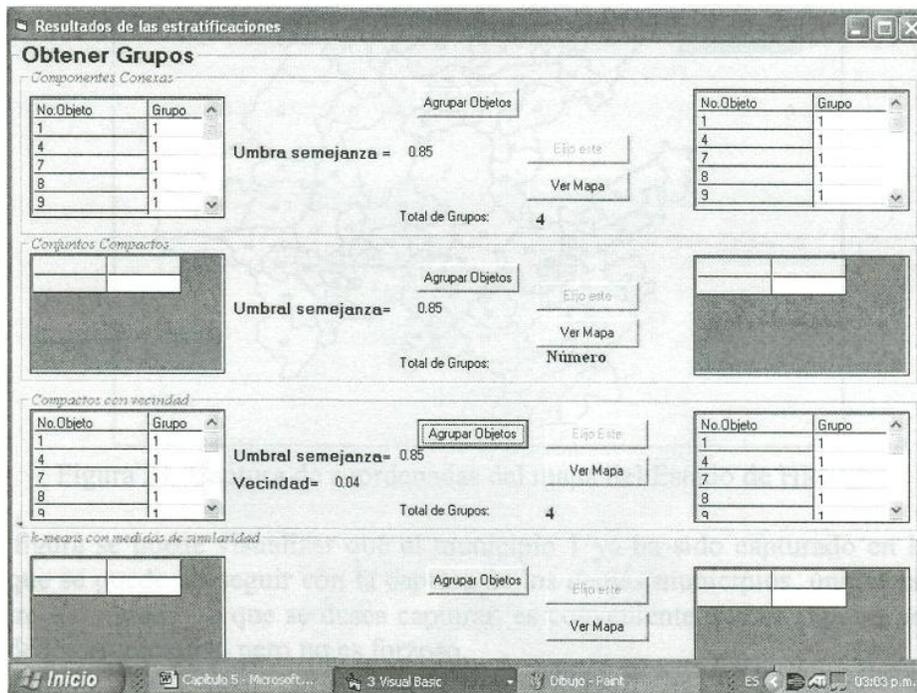


Figura 20. Obtención de componentes conexas utilizando el criterio de conjuntos compactos con vecindad. En esta figura se observa perfectamente que con el mismo umbral de semejanza y con una

vecindad de 0.06 se obtienen los mismos grupos que genera el criterio agrupacional componentes conexas.

### 5.3.- Presentación de resultados de forma gráfica

Una vez generadas las estratificaciones de municipios por cualquier de estos cuatro criterios agrupacionales, el siguiente paso es observar de forma gráfica (mapa de hidalgo con sus 84 municipios) la forma en que se agruparon los municipios, tal y como lo realizan los SSH. De hecho, es una de las necesidades de los expertos, poder observar el mapa de hidalgo coloreado según las zonas de riesgo obtenidas.

Para realizar este proceso de forma automática, la herramienta computacional requiere el mapa de hidalgo en formato "BMP". El sistema proporciona un módulo para capturar las coordenadas de un píxel en la pantalla  $x$ ,  $y$  de cada municipio en el mapa, donde estas coordenadas son almacenadas en la base de datos en Access. Por lo que solamente es necesario capturar el mapa una sola vez. Esto para que la herramienta computacional sea lo más flexible posible y acepte prácticamente cualquier tipo de mapa de bits, (ver figura 21.)

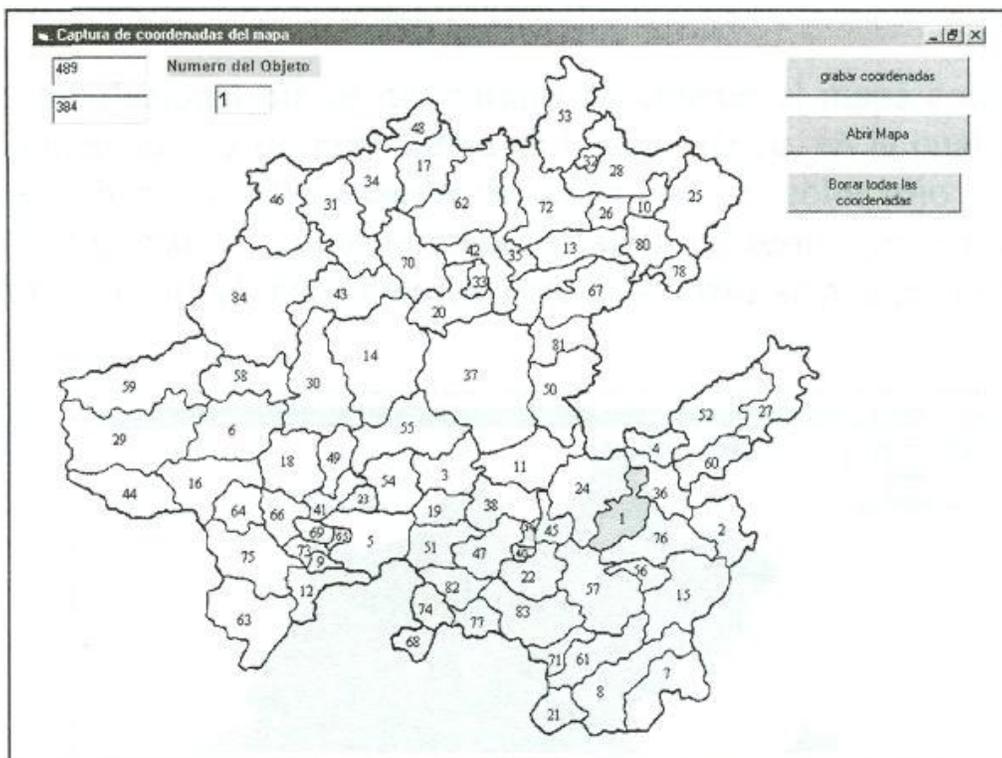


Figura 21. Captura de coordenadas del mapa del Estado de Hidalgo.

En esta figura se puede visualizar que el municipio 1 ya ha sido capturado en la base de datos y que se puede proseguir con la captura de los demás municipios, únicamente dando clic dentro del municipio que se desea capturar, es conveniente que se siga un orden en la captura de los municipios, pero no es forzoso.

Una vez que todas las coordenadas del mapa han sido capturadas, el mapa estará listo para poder ser coloreado según las estratificaciones obtenidas por cualquiera de los 4 criterios

agrupacionales. El mapa finalmente capturado se puede ver en la figura 22.

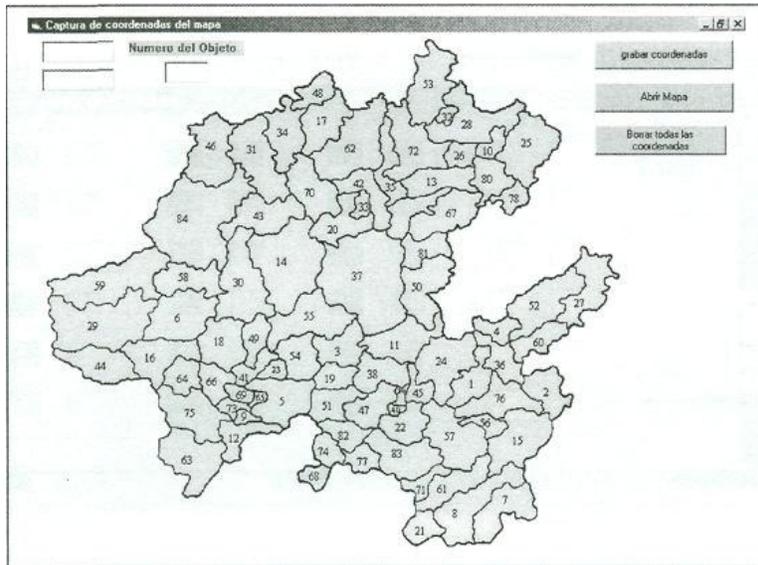


Figura 22. Apariencia del mapa cuando ha sido capturado por completo.

La herramienta computacional ofrece dos formas de obtener el mapa coloreado, según las estratificaciones obtenidas. La primera forma es la automática, en la cual dependiendo del número de grupos se hace una degradación del color rojo (el color rojo es el color que los SSH utilizan para etiquetar al grupo de mayor riesgo), posteriormente el azul, el verde y finalmente el color amarillo, el mapa coloreado de esta forma se puede observar en la figura

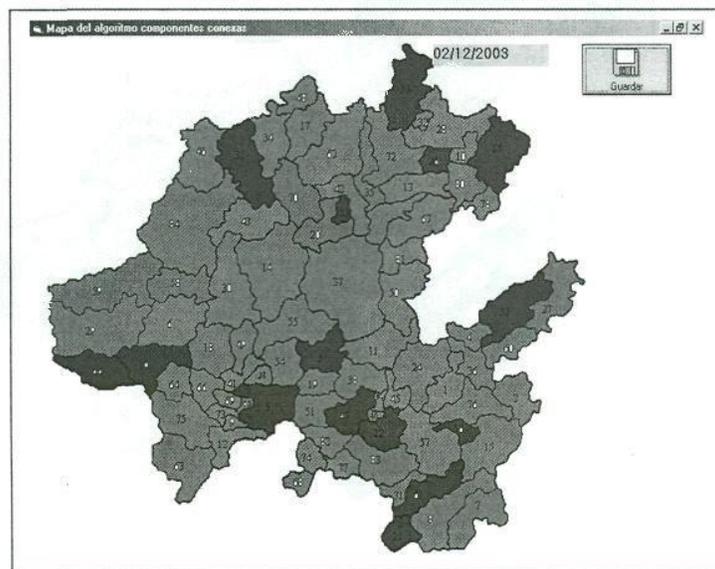


Figura 23. Representación de la estratificación en el mapa de forma automática.

La segunda forma de colorear el mapa es de forma manual, se pueden elegir los colores de una paleta de colores que se fue diseñada tomando en cuenta las necesidades del problema y de los expertos. De esta forma se degradan los 4 colores que utilizan los SSH, ofreciendo una variedad en la escala de los rojos, azules, verdes y amarillos. La paleta de colores que

permite manejar el sistema se puede observar en la figura 24.

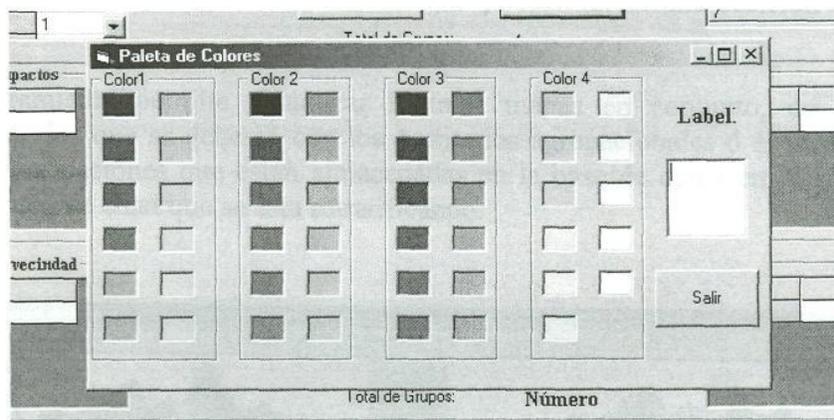


Figura 24. Paleta de colores según las necesidades del experto.

La misma estratificación de municipios del mapa de la figura 23 coloreada de forma manual se observa en la figura 25, el que el experto utilice colores de su preferencia le da a la herramienta mayor flexibilidad para abordar otro tipo de problemas similares al que se está estudiando.

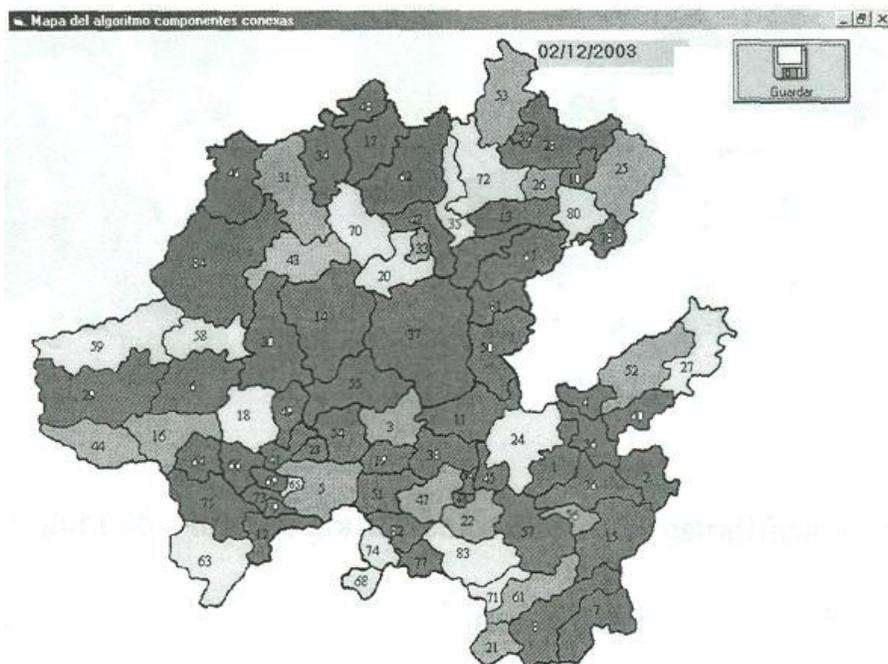


Figura 25 . Representación de la estratificación en el mapa de forma manual

#### 5.4.- Análisis gráfico de estratificaciones nuevas y anteriores

Otra de las necesidades de los SSH es comparar de forma gráfica el historial de estratificaciones nuevas con estratificaciones anteriores (ver figura 26). El hecho de

observar si algún municipio permanece o cambia de color es muy significativo para los expertos, ya que a partir de esta información pueden tomar una decisión o realizar alguna acción.

Esta herramienta permite visualizar hasta 4 mapas en conjunto, de estratificaciones anteriores, los que se generan con los 4 criterios agrupacionales ó el visualizar mapas de otras estratificaciones que están almacenadas en la base de datos en Access. Obviamente para el grupo de edad que se esta estratificando.

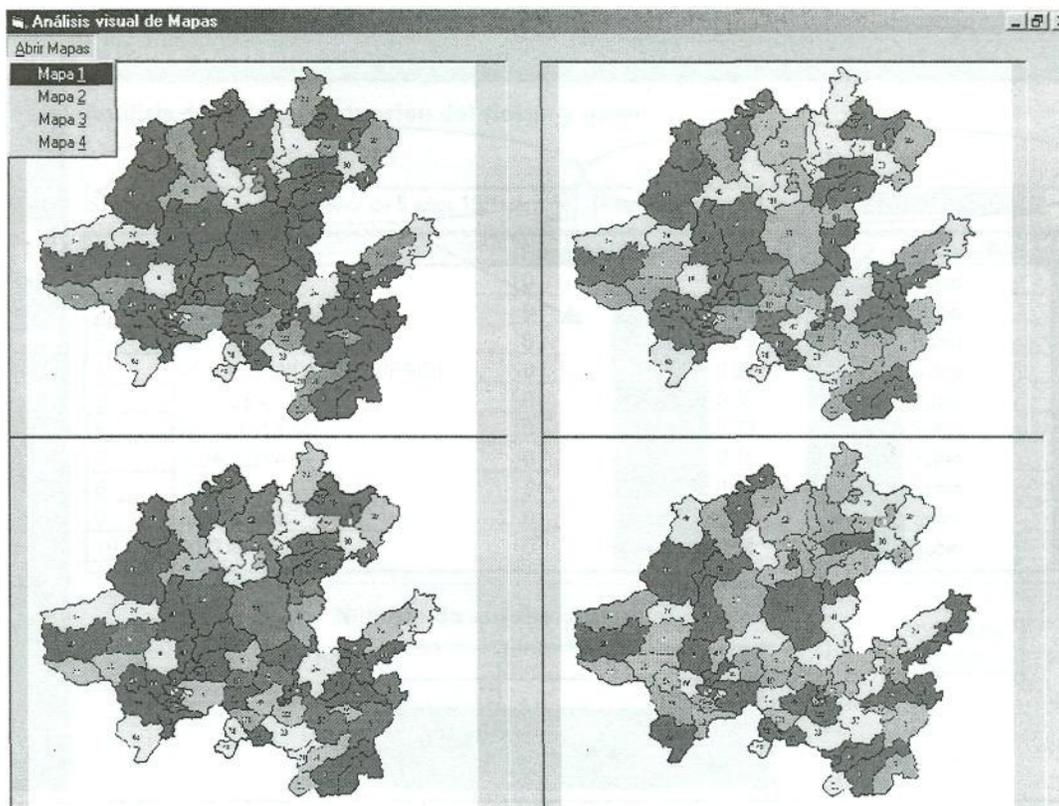


Figura 26. Análisis gráfico entre diferentes estratificaciones.

### 5.5.- Análisis tabular de estratificaciones nuevas y anteriores

Por otro lado, siempre es importante corroborar lo que se ve, con lo que se hizo. En esta parte, los expertos pueden comparar dos estratificaciones, ya sean con igual número de grupos o con número de grupos diferentes. El beneficio de esto es que de forma gráfica no se pueden visualizar a simple vista cuántos municipios pertenecen a cada grupo, y si existen mejoras en las dos estratificaciones a comparar en cuanto a la disminución del número de municipios pertenecientes al grupo de mayor riesgo. Peor aún, si han entrado a ese grupo otros municipios que pertenecían a un grupo de menor riesgo. Todo esto se puede observar en el módulo de comparación tabular de la herramienta computacional, (ver figuras 27 y 28). Es importante mencionar que este módulo no está diseñado para conocer cual es el

mejor algoritmo o cual de ellos obtuvo los mejores resultados. Además proporciona al experto información acerca de que municipio ha aumentado o disminuido en cuanto a la aportación de riesgo de una estratificación a otra . Todo esto puede realizarse con las estratificaciones generadas por los algoritmos en un mismo grupo de edad o en diferentes estratificaciones ya almacenadas en la base de datos referentes a periodos anteriores.

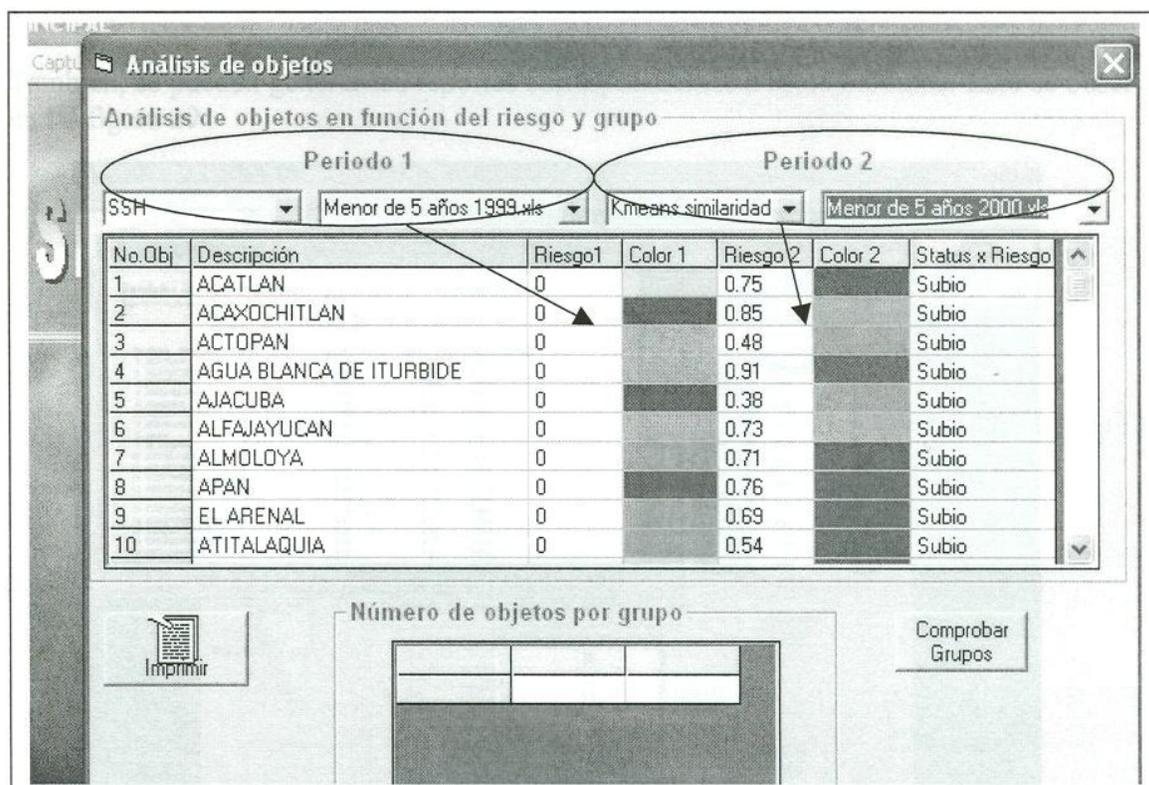


Figura 27. Análisis tabular entre un par de estratificaciones cualesquiera.

En esta figura, se está realizando una comparación entre la estratificación obtenida usando componentes conexas del archivo menor de 5 años 1999 contra la estratificación generada por k-means con el archivo menor de 5 años 2000.

En esta parte del análisis se muestra que ambas estratificaciones tienen el mismo número de grupos pero no el mismo número de municipios pertenecientes a cada grupo. Únicamente coinciden en el grupo 1.

**Número de objetos por grupo**

Grupo	conex	kmean
1	38	38
2	14	19
3	15	21
4	17	6

Figura 28. Número de municipios por grupo y estratificación.

También, se pueden generar los reportes correspondientes a estos módulos. Esto se observa en las figura 29.

**Servicios de salud de Hidalgo**

Reporte de resultados de las aperturas de riesgo

Pondera 1: Menor de 5 años 1999.xls  
Pondera 2: Menor de 5 años 2000.xls

No. Objeto	Descripción del objeto	Apertura rango 1	Apertura rango 2	Calificación de riesgo
1	ACATLAN	0.00	0.75	Subido
2	ACAYUCHITLAN	0.00	0.55	Subido
3	ACTOPAN	0.00	0.42	Subido
4	AGUA BLANCA DE TURB	0.00	0.91	Subido
5	AJACUBA	0.00	0.18	Subido
6	ALFARAYUCAN	0.00	0.73	Subido
7	ALBUQUERQUE	0.00	0.71	Subido
8	APAH	0.00	0.76	Subido
9	EL ARRIAL	0.00	0.69	Subido
10	ATLAPALCOA	0.00	0.54	Subido
11	ATLAPEXCO	0.00	0.70	Subido
12	ATOTONILCO EL GRANDE	0.00	0.32	Subido
13	ATOTONILCO DE TULA	0.00	0.75	Subido
14	CADZALTI	0.00	0.78	Subido
15	CARDENAL	0.00	0.66	Subido
16	CUAUTEMOC DE ITHYMOLO	0.00	0.41	Subido
17	CHAPANTONGO	0.00	0.61	Subido
18	CHAPULTEPEC	0.00	0.42	Subido
19	CHILCAGUATEL	0.00	0.61	Subido

**Número de objetos por grupo**

Grupo	No. Objetos 1	No. Objetos 2
1	27	39
2	19	12
3	19	13
4	19	20

Figura 29. Reporte del análisis tabular de estratificaciones.

### 5.6.- Consulta de estratificaciones almacenadas

Sin embargo, es muy importante y necesario el poder realizar consultas de estratificaciones generadas por el sistema en años o periodos anteriores. Por lo tanto este sistema también cuenta con un módulo que le permite al experto el poder obtener el reporte y el mapa correspondiente a alguna estratificación ya almacenada en la base de datos. Esto se puede visualizar en las figuras 30, 31 y 32.

No. Mun.	Descripción	Grupo	Aport. riesgo
60	TENANGO DE DORIA	1	0
69	TIZAYUCA	1	0
39	MINERAL DEL MONTE	1	0
77	TULANCINGO DE BRAVO	1	0
47	PACULA	1	0
41	MIXQUIAHUALA	1	0
61	TEPEAPULCO	1	0
48	PACHUCA DE SOTO	1	0
37	METZTITLAN	1	0
22	EPAZOYUCAN	1	0
5	AJACUBA	1	0
43	NICOLAS FLORES	1	0
84	ZIMAPAN	1	0
76	TULA DE ALLENDE	1	0
81	ZACUALTIPAN DE ANGELES	1	0
33	JUAREZ HIDALGO	1	0
8	APAN	1	0
34	LOLOTLA	1	0
30	IXMIQUILPAN	1	0
27	HUEHUETLA	1	0
29	HUICHAPAN	1	0

Figura 30. Consulta de la estratificación que realizó los SSH en 1999 para el grupo de edad del menor de 5 años.

Servicios de Salud de Hidalgo  
 Reporte de estratificaciones ya almacenadas  
 21/02/2004  
 Estratificación del grupo: Menor de 5 años 1999.xls  
 Con fecha: 12/10/2003  
 Algoritmo: SSH

No. Municipio	Descripción	Grupo	Aportación riesgo
60	TEHANGO DE DORIA	1	0.00
69	TIZAYUCA	1	0.00
39	MINERAL DEL MONTE	1	0.00
77	TULANCINGO DE BRAVO	1	0.00
47	PACULA	1	0.00
41	MIXQUIAHUALA	1	0.00
61	TEPEAPULCO	1	0.00
48	PACHUCA DE SOTO	1	0.00
37	METZITILAN	1	0.00
22	EPAZOYUCAN	1	0.00
5	AJACUBA	1	0.00
43	NICOLAS FLORES	1	0.00
84	ZIMAPAN	1	0.00
76	TULA DE ALLENDE	1	0.00
81	ZACUALTIPAN DE ANGELES	1	0.00
33	JUAREZ HIDALGO	1	0.00
8	APAN	1	0.00
34	LOLOTLA	1	0.00
30	IXMQUILPAN	1	0.00
27	HUEHUETLA	1	0.00
29	HUICHAPAN	1	0.00
2	ACAXOCHITLAN	1	0.00
13	ATOTONILCO DE TULA	1	0.00
63	TEPEH DE OCAMPO	1	0.00
74	TLAXCOAPAN	1	0.00
18	CHAPULHUACAN	1	0.00
71	TLAHUILTEPA	1	0.00
4	AGUA BLANCA DE ITURBIDE	2	0.00
67	TEZONTEPEC DE ALDAMA	2	0.00

1 of 2    Cancel    Close    84 of 84    Total:84    100%

Figura 31. Parte del reporte que generan las consulta.

Aquí se muestra parte del reporte que genera la consulta de una estratificación ya almacenada en la base de datos. De igual forma se puede obtener el mapa que se generó en función a esta estratificación. Como se ve en la figura 32.

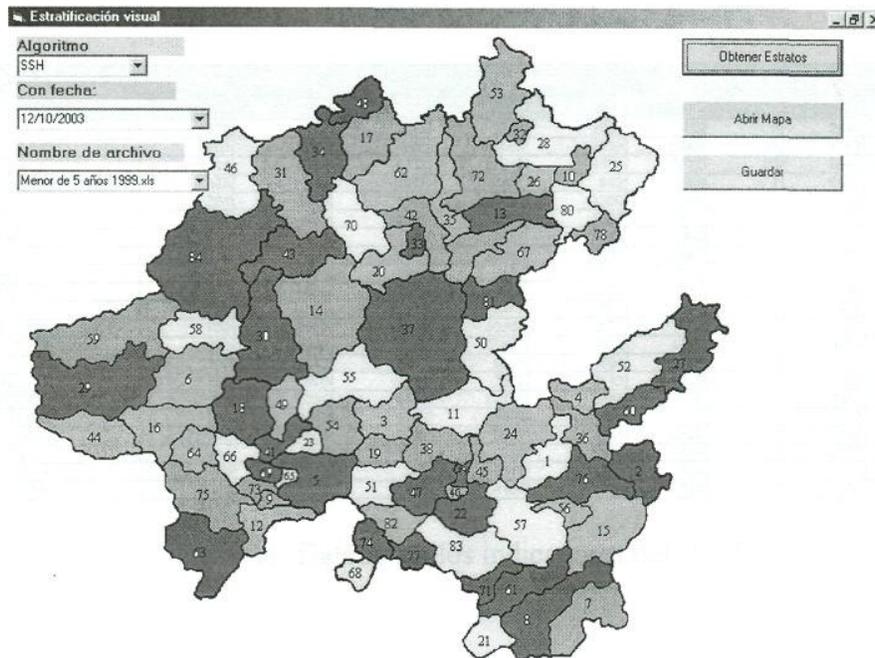


Figura 32. Mapa de la estratificación obtenida por los SSH en 1999 para el grupo de edad menor de 5 años.

# Capítulo 6

## EXPERIMENTACIÓN Y ANÁLISIS DE RESULTADOS

En este capítulo se muestran los resultados experimentales obtenidos, utilizando la herramienta computacional propuesta, usando datos reales del grupo de edad del menor de 5 años, proporcionados por los SSH, utilizando el formato propuesto en Excel. La base de datos de estudio consta de 84 municipios (objetos), 16 indicadores (rasgos o atributos) y 5 columnas de poblaciones para el preprocesamiento de los indicadores. Esta base de datos se puede ver en las tablas 11 y 12.

No	Municipio	población1	poblacion2	poblacion3	poblacion4	poblacion5	Morb IRAS < 1año	Morb IRAS < 1año	Morb.IRAS 1 - 4 años	Morb IRAS < 1 - 4 años	Morb EDAS < 1año
001	ACATLAN	578	2442	2453	9340	2526	1462	6	2840	5	277
002	ACAXOCHITLAN	1753	3926	5995	15653	3939	2849	41	4163	288	553
003	ACTOPAN	1214	4417	5313	17217	4500	2817	23	7541	4	608
004	AGUA BLANCA DE ITURBIDE	325	972	1477	3847	982	895	7	1590	5	83
005	AJACUBA	364	1315	1847	5120	1339	1540	10	3127	1	220
006	ALFAJAYUCAN	438	1495	1907	5892	1527	931	10	2052	3	91
007	ALMOLOYA	485	1172	1996	4457	1217	800	8	1562	6	186
008	APAN	1216	3554	5195	13862	3638	4545	20	9579	5	946
009	EL ARENAL	443	1627	1812	6120	1707	1066	9	2314	2	160
010	ATITALAQUIA	467	1950	2381	7415	2028	1584	10	3196	4	296
011	ATLAPEXCO	760	1597	2976	6376	1601	1531	0	3976	1	410
012	ATOTONILCO EL GRANDE	763	2635	3374	10365	2671	1659	22	3090	3	311
013	ATOTONILCO DE TULA	665	3092	2756	10443	3628	4777	7	9709	1	769
014	CALNALI	1180	1673	2931	6598	1694	1856	4	3431	4	363
015	CARDONAL	525	1998	2075	7572	2113	1919	13	4768	1	134
016	CUAUTEPEC DE HINOJOSA	1424	5399	6335	20150	5689	2886	28	7380	9	565
017	CHAPANTONGO	303	1142	1170	4502	1155	577	6	1602	2	79
018	CHAPULHUACAN	749	2535	2705	9462	2714	2222	8	5156	8	426
019	CHILCUAUTLA	414	1628	1916	6280	1684	1456	6	2977	3	199
020	ELOXOCHITLAN	86	170	362	713	165	328	0	1076	0	54
021	EMILIANO ZAPATA	342	1133	1568	4457	1148	838	5	2056	3	103

Tabla 11. Datos para los indicadores del 1-5

Mort. EDAS	Morb. EDAS	Mort. EDAS	Deteccion de EPV	Casos de	Mort. Desnu.	Mort. Desnu.	Mort. Acciden.	Cobertura de vac.	Cobertura Antisar.	Marginilidad
------------	------------	------------	------------------	----------	--------------	--------------	----------------	-------------------	--------------------	--------------

<1 año	1-4 año	1-4 año		EPV	<1 año	<1-4 años	<1-4 años	<1 año	14 meses	
6	686	2	0	1	0	0	2	100.00	100.00	3
36	1173	20	0	0	5	2	5	95.4	95.4	4
25	1530	9	1	2	3	0	0	97.22	100.00	1
6	343	1	0	2	1	0	0	100.00	100.00	2
10	733	0	1	1	0	1	0	78.95	78.95	1
9	363	3	0	0	1	0	3	100.00	100.00	3
7	430	3	0	1	2	0	0	100.00	100.00	2
21	2130	3	0	2	1	0	1	97.18	97.18	1
10	413	2	0	1	0	1	0	100.00	100.00	2
9	920	0	0	1	1	0	0	100.00	100.00	0
0	1457	1	0	1	0	0	0	100	100.00	4
21	742	4	0	2	1	0	0	69.81	69.81	2
7	1337	0	0	2	0	0	0	87.76	87.76	1
3	1253	2	0	1	1	0	0	100.00	100.00	4
14	430	2	0	0	0	1	1	100.00	100.00	3
26	1292	5	1	1	2	2	0	94.59	94.59	2
5	317	3	0	0	1	0	1	100.00	100.00	2
8	1406	6	1	0	0	0	1	100.00	100.00	4
6	731	0	0	0	0	1	0	100.0	100.00	3
0	204	0	1	0	0	0	0	0.0	0.0	3
6	295	0	1	1	0	0	1	100.0	100.00	0

Tabla 12. Datos para los indicadores del 6-16

## 6.1.- Análisis del grupo de edad menor de 5 años

Esta base de datos se refiere al grupo de edad menor de 5 años correspondiente a 1999. Cabe mencionar que estos datos fueron proporcionados por los Servicios de salud de Hidalgo. Dicha base está compuesta de las siguientes partes.

- 84 municipios que conforman al Estado de Hidalgo.
- 5 columnas de población de individuos que utilizan cada uno de los indicadores para el preprocesamiento de datos.
- 16 indicadores de daños a la salud, 13 indicadores numéricos, 2 Avalentes y 1 booleano.

Los 16 indicadores de salud que conforman a este grupo de edad se muestran a continuación:

- 1) Morbilidad por IRAS en menores de un año
- 2) Mortalidad por IRAS en menores de un año
- 3) Morbilidad por ERAS en niños de 1 a 4 años
- 4) Mortalidad por IRAS en niños de 1 a 4 años
- 5) Morbilidad por EDAS en menores de un año
- 6) Mortalidad por EDAS en menores de un año
- 7) Morbilidad por EDAS en niños de 1 a 4 años
- 8) Mortalidad por EDAS en niños de 1 a 4 años
- 9) Detección de EPV
- 10) Casos de EPV
- 11) Mortalidad por desnutrición en menores de un año
- 12) Mortalidad por desnutrición en niños de 1 a 4 años
- 13) Mortalidad por accidentes en niños de 1 a 4 años
- 14) Coberturas de vacunación en niños del año
- 15) Coberturas de vacunación anti sarampión a los 14 meses

## 16) Marginalidad

Los términos contenidos en estos indicadores se pueden consultar en el glosario de términos. La fórmula para el preprocesamiento, el criterio de comparación (ver criterios en el capítulo 2) y la forma de evaluar el riesgo (ver casos de riesgo en el capítulo 2) por indicador para este grupo de edad puede visualizarse en la tabla 13.

No indicador	Preprocesamiento	Tipo indicador	Criterio de comparación	Evaluación de la aportación de riesgo
1	Calculo de TASA	Numérico	<b>a</b>	Caso 1
2	Calculo de TASA	Numérico	<b>a</b>	Caso 1
3	Calculo de TASA	Numérico	<b>a</b>	Caso 1
4	Calculo de TASA	Numérico	<b>a</b>	Caso 1
5	Calculo de TASA	Numérico	<b>a</b>	Caso 1
6	Calculo de TASA	Numérico	<b>a</b>	Caso 1
7	Calculo de TASA	Numérico	<b>a</b>	Caso 1
8	Calculo de TASA	Numérico	<b>a</b>	Caso 1
9	Sin cálculo el dato pasa sin modificaciones.	Boleano 1= Detectan 0= No detectan	<b>h</b>	Caso 2
10	Sin cálculo el dato pasa sin modificaciones.	k-valente 0=Sin casos 1= un caso 2= Mas de un caso	<b>c</b>	Caso 1
11	Calculo de TASA	Numérico	<b>a</b>	Caso 1
12	Calculo de TASA	Numérico	<b>a</b>	Caso 1
13	Calculo de TASA	Numérico	<b>a</b>	Caso 1
14	Sin cálculo el dato pasa sin modificaciones.	Numérico	<b>a</b>	Caso 2
15	Sin cálculo el dato pasa sin modificaciones.	Numérico	<b>a</b>	Caso 2
16	Sin cálculo el dato pasa sin modificaciones.	k-valente 0= Muy baja 1= Baja 2= Mediana 3= Alta 4= Muy alta	<b>d</b>	Caso 1

Tabla 13. Análisis por indicador

Una vez que se conoce la forma en la que están descritos todos y cada uno de los 16 indicadores que conforman al grupo de edad "Menor de 5 años", se llevan acabo los cálculos de las tasas para los indicadores numéricos. Estos cálculos se visualizan en la tabla

14 y el reporte del indicador mortalidad por IRAS en menores de 1 año se presenta en la tabla 15.

A continuación se muestran algunos de los datos otorgados por los SSH con los que se realiza la siguiente experimentación. Es importante mencionar que estos datos necesitan ser procesados por el sistema para obtener las tasas, índices y porcentajes que requieren los criterios de comparación para realizar los agrupamientos de municipios. No se presenta toda la tabla completa debido a que los SSH mantienen estos datos como información confidencial.

La siguiente tabla, muestran los datos ya procesados, en otras palabras, son los índices, tasas, porcentajes y casos que la herramienta computacional utilizará para aplicar los criterios de comparación asignados a cada indicador y así obtener la matriz de semejanza. Finalmente mostrar al usuario los resultados (estratificación de municipios) en función de estos datos.

No	Municipio	Morb. IRAS <1 año	Morb. IRAS <1 año	Morb. IRAS 1-4 años	Morb. IRAS 1-4 años	Morb. EDAS <1 año	Morb. EDAS <1 año	Morb. EDAS 1-4 años	Morb. EDAS 1-4 años	Detección De EPV	Casos de EPV	Mort. Desun. <1 año	Mort. Desun. 1-4 años	Mort. Accident. 1-4	Cobertura De vac. <1 año	Cobertura Antisur. 14 meses	marginilidad
001	ACATLAN	596.00	10.38	1157.77	2.05	112.92	10.38	73.45	8.19	0	1	0.00	0.00	7.92	100.00	10.00	3
002	ACAXOCHITLAN	475.23	23.39	694.41	7.13	92.24	20.54	74.94	50.94	0	0	2.85	5.09	12.69	95.40	95.40	4
003	ACTOPAN	718.43	18.95	1419.35	0.91	114.44	19.77	88.87	20.38	1	2	2.47	0.00	0.00	97.22	100.00	1
004	AGUA BLANCA DE ITURBIDE	605.96	21.54	1076.51	5.14	56.19	18.46	89.16	10.29	0	2	3.08	0.00	0.00	100.00	100.00	3
005	AJACUBA	833.78	27.47	1693.02	0.76	119.11	27.47	143.16	0.00	1	1	0.00	7.60	0.00	78.95	78.95	1
006	ALFAJAYUCAN	488.20	22.83	1076.04	2.01	47.72	20.55	61.61	20.07	0	0	2.28	0.00	19.65	100.00	100.00	3
007	ALMOLOYA	400.80	16.49	782.57	5.12	93.19	14.43	96.48	25.60	0	1	4.12	0.00	0.00	100.00	100.00	2
008	APAN	874.88	16.45	1843.89	1.41	182.10	17.27	153.66	8.44	0	2	0.82	0.00	2.75	97.18	97.18	1
009	EL ARENAL	588.30	20.32	1277.04	1.23	88.30	22.57	67.48	12.29	0	1	0.00	6.15	0.00	100.00	100.00	2
010	ATITALAQUA	665.27	21.41	1342.29	2.05	124.32	19.27	124.07	0.00	0	1	2.14	0.00	0.00	100.00	100.00	0
011	ATLAPEXCO	514.45	0.00	1336.02	0.63	137.77	0.00	228.51	6.26	0	1	0.00	0.00	0.00	100.00	100.00	4
012	ATOTONILCO EL GRANDE	491.7	28.83	915.83	1.14	92.18	27.52	71.59	15.18	0	2	1.31	0.00	0.00	69.81	69.81	2
013	ATOTONILCO DE TULA	1733.31	10.53	3522.86	0.32	279.03	10.53	128.03	0.00	0	2	0.00	0.00	0.00	87.76	87.76	1
014	CALNALI	633.23	3.39	1170.59	2.39	123.85	2.54	189.91	11.95	0	1	0.85	0.00	0.00	100.00	100.00	4
015	CARDONAL	924.82	24.76	2297.83	0.50	64.58	26.67	56.79	10.01	0	0	0.00	5.01	4.73	100.00	100.00	3
016	CUAUTEPEC DE HINOJOSA	613.42	19.66	1164.96	1.67	89.19	18.26	64.12	9.26	1	1	1.40	3.70	0.00	94.59	94.59	2
017	CHAPANTONGO	493.16	19.80	1369.23	1.75	67.52	16.50	70.14	26.27	0	0	3.30	0.00	8.66	100.00	100.00	2
018	CHAPULHUACAN	821.44	10.68	1906.10	3.16	157.49	10.68	148.59	23.67	1	0	0.00	0.00	3.68	100.00	100.00	4
019	CHILCUAUTLA	759.92	14.49	1553.76	1.84	103.86	14.49	116.40	0.00	0	0	0.00	6.14	0.00	100.00	100.00	3

Tabla 14. Tasas, índices, porcentajes y casos de cada indicador para el grupo de edad “Menor de 5 años”

**SERVICIOS DE SALUD DE HIDALGO****REPORTE DEL GRUPO DE EDAD:****Menor de 5 años 1999.xls****INDICADOR:****Mort. IRAS < 1 año****04/12/2003**

<b>NO. MUNICIPIO</b>	<b>MUNICIPIO</b>	<b>TASA CALCULADA</b>
24	HUASCA DEOCAMPO	49.51
12	ATOTONILCO EL GRANDE	28.83
5	ÁJACUBA	27.47
59	TECOZAUTLA	26.79
63	TEPEJI DE OCAMPO	26.19
47	PACULA	25.97
38	TASQUILLO	25.70
74	TLAXCOAPAN	25.21
43	<b>NICOLÁS FLORES</b>	25.09
15	CARDONAL	24.76
29	HUICHAPAN	24.49
2	ACAXOCHITLAN	23.39
6	ALFAJAYUCAN	22.83
38	MINERAL DEL CHICO	22.73
39	MINERAL DEL MONTE	21.94
4	AGUA BLANCA DE ITURB	21.54
10	ATITALAQUIA	21.41
23	FRANCISCO I. MADERO	20.33
9	EL ARENAL	20.32
66	VILLA DE TEZONTEPEC	19.93
44	NOPALA DE VILLAGRAN	19.85
17	CHAPANTONGO	19.80
16	CUAUTEPEC DE HINOJOS	19.66
67	TEZONTEPEC DE ALDAMA	19.44
3	ACTOPAN	18.95

Tabla 15. Subconjunto del reporte de tasas para el indicador Mortalidad por IRAS en < 1 año.

## 6.2.- Estratificaciones obtenidas aplicando el SAD

En esta parte se muestran las estratificaciones obtenidas al aplicar la herramienta SAD, obteniendo como resultados los grupos generados y los municipios que pertenecen a cada uno de ellos, así como las aportaciones de riesgo por grupo y por municipio, en los diferentes criterios agrupacionales.

Es importante mencionar que el valor para el umbral de semejanza  $\beta_0$  utilizado en esta experimentación para esta base de datos, se obtuvo de dos maneras. Primera, de forma experimental como el resultado de una serie de pruebas con los expertos, La segunda, usando un resultado teórico que permite obtener este valor a partir de los valores de la matriz de semejanza de los objetos [SaG+ 2003]. A partir de este procedimiento, se obtiene el valor de semejanza que genera los grupos naturales óptimos maximizando la semejanza entre los objetos del mismo grupo y separándolos con los objetos de otros grupos.

El valor de semejanza entre municipios utilizado es 0.85 en un intervalo de 0 a 1, donde 1

es la mayor semejanza entre un par de municipios y 0 es la diferencia total entre dos municipios. Como se puede ver, el valor de semejanza que se utiliza permite generar resultados tan restrictivos como el experto lo necesite, independientemente del criterio agrupacional que se utilice.

### **6.2.1.- Estratificación generada utilizando componentes conexas**

Las estratificaciones obtenidas a través de la implementación de este criterio agrupacional se muestran en las tablas 16 a la 19. Donde se pueden observar los 4 grupos obtenidos con su respectivo análisis de indicadores en función de su riesgo. La presentación de resultados de forma gráfica (mapa) se muestra en la figura 33, este mapa está coloreado de forma manual para acercarse más al mapa que obtuvieron los SSH. Los grupos obtenidos al aplicar dicho criterio agrupacional son los siguientes:

Grupo: 1

Total de objetos: 38

Promedio de Riesgo del grupo: 0.76

Municipios del grupo : {34,60,40,73,4,28,46,84,49,29,30,12,54,39,79,45,14,41,11,8,32,13,1,69,75,76,7,66,48,9,77,55,64,36,67,23,10,51}

Grupo: 2

Total de objetos: 14

Promedio de Riesgo del grupo: 0.62

Municipios del grupo : {2,62,6,15,78,37,19,17,38,42,81,57,50,82}

Grupo: 3

Total de objetos: 15

Promedio de Riesgo del grupo: 0.51

Municipios del grupo : {33,26,47,25,53,22,44,52,3,61,31,16,5,56,21}

Grupo: 4

Total de objetos: 17

Promedio de Riesgo del grupo: 0.32

Municipios del grupo : {43,71,18,27,24,59,68,80,74,20,58,63,35,65,83,70,72}

A continuación, se presentan los reportes generados para cada grupo, así como el reporte de cómo influyen los indicadores para el cálculo del promedio de riesgo final por grupo.

En el siguiente grupo de tablas, se muestran los resultados de la estratificación obtenida utilizando componentes conexas. Se puede observar que en los reportes de aportaciones de riesgo, el indicador que más influye al riesgo de cada grupo no siempre es el mismo. Por ejemplo, para el grupo 1 en la tabla 16 el que más influyo fue el indicador (Detección de EPV) al igual que el grupo 2, pero nótese que en el grupo 4 (tabla 19) es el que menos influyo al riesgo del grupo, el que más influyo en este grupo es la marginalidad.

SERVICIOS DE SALUD DE HIDALGO			31/01/2004
REPORTE DEL GRUPO DE EDAD:		Menor de 5 años 1999.xls	
GRUPO:	1	Promedio de riesgo:	0.76
ALGORITMO:	conexas	Total de municipios del grupo	38
Numero	Municipio	Aportación de riesgo	
34	LOLOTLA	1.00	
60	TENANGO DE DORIA	0.94	
40	LA MISION	0.93	
73	TLANCHINOL	0.92	
4	AGUA BLANCA DE ITURBIDE	0.91	
28	HUEJUTLA DE REYES	0.88	
46	SAN FELIPE DE ORIZATLAN	0.88	
84	ZIMAPÁN	0.86	
49	PISAFLORES	0.85	
29	HUICHAPAN	0.82	
30	IXMIQUILPAN	0.82	
12	ATOTONILCO EL GRANDE	0.82	
54	SAN SALVADOR	0.81	
39	MINERAL DEL MONTE	0.80	
79	XOCHICOATLAN	0.79	
45	OMITLAN DE JUAREZ	0.78	
14	CALNALI	0.78	
41	MIXQUIAHUALA	0.77	
11	ATLAPEXCO	0.76	
8	APAN	0.76	
32	JALTOCAN	0.76	
13	ATOTONILCO DE TULA	0.75	
1	ACATLAN	0.75	
69	TIZAYUCA	0.74	
75	TOLCAYUCA	0.74	

SERVICIOS DE SALUD DE HIDALGO		
Reporte de promedios de aportaciones de riesgo por indicador para el grupo:		
1		
Con el algoritmo: conexas	Grupo de Edad:	Menor de 5 años 1999.xls
	indicador	Promedio de riesgo
	Deteccion de EPV	1.00
	Casos de EPV	0.78
	Marginalidad	0.52
	Morb. EDAS <1 año	0.20
	Morb. IRAS <1 año	0.19
	Morb. IRAS <1-4 años	0.19
	Morb. Desnu. <1 año	0.19
	Morb. EDAS <1-4 año	0.18
	Morb. EDAS <1 año	0.17
	Morb. IRAS <1 año	0.17
	Morb. IRAS <1-4 años	0.16
	Morb. Desnu <1-4 año	0.12
	Morb. EDAS <1-4 año	0.09
	Morb. Acceden. <1-4 año	0.08
	Cobertura de vac < 1 año	0.00
	Cobertura Antisar 14 meses	0.00

Tabla 16. Reportes para el grupo 1 aplicando el criterio Componentes Conexas.

SERVICIOS DE SALUD DE HIDALGO			31/01/2004
REPORTE DEL GRUPO DE EDAD: Menor de 5 años 1999.xls			
GRUPO:	2	Promedio de riesgo:	0.62
ALGORITMO:	conexas	Total de municipios del grupo:	14
Número	Municipio	Aportación de riesgo	
2	ACAXOCHITLAN	0.85	
62	TEPEHUACAN DE GUERRERO	0.76	
6	ALFAJAYUCAN	0.73	
15	CARDONAL	0.66	
78	XOCHITAPAN	0.66	
37	METZTITLAN	0.65	
19	CHILCUAUTLA	0.61	
17	CHAPANTONGO	0.61	
38	MINERAL DEL CHICO	0.59	
42	MOLANGO DE ESCAMILLA	0.59	
81	ZACUALTIPAN DE ANGELES	0.59	
57	SINGUILUCAN	0.56	
50	PROGRESO	0.44	
82	ZAPOTLAN DE JUAREZ	0.44	

Servicios de Salud de Hidalgo		31/01/2004
Reporte de promedios de aportaciones de riesgo por indicador		
para el grupo:	2	Grupo de Edad: Menor de 5 años 1999.xls
con el algoritmo:	conexas	
Indicador	Promedio de riesgo	
Deteccion de EPV	1.00	
Marginalidad	0.70	
Mort. Accident. 1-4 años	0.35	
Mort. Derm. 1-4 años	0.22	
Mort. Derm. < 1 año	0.20	
Mort. IRAS < 1 año	0.19	
Mort. EDAS < 1 año	0.19	
Mort. IRAS 1-4 años	0.18	
Morb. IRAS 1-4 años	0.17	
Morb. EDAS 1-4 años	0.17	
Morb. EDAS < 1 año	0.17	
Morb. IRAS < 1 año	0.16	
Mort. EDAS 1-4 años	0.15	
Cobertura Antiazar: 14 meses	0.00	
Cobertura de vac. < 1 año	0.00	
Casos de EPV	0.00	

Tabla 17. Reportes para el grupo 2 aplicando el criterio Componentes Conexas.

SERVICIOS DE SALUD DE HIDALGO			31/01/2004
REPORTE DEL GRUPO DE EDAD: Menor de 5 años 1999.xls			
GRUPO:	3	Promedio de riesgo:	0.51
ALGORITMO:	conexas	Total de municipios del grupo:	15
Número	Municipio	Aportación de riesgo	
33	JUAREZ HIDALGO	0.76	
26	HUAZALINGO	0.68	
47	PACULA	0.66	
25	HUAUTLA	0.61	
53	SAN BARTOLO TUTOTEPEC	0.61	
22	EPAZOYUCAN	0.59	
44	NOPALA DE VILLAGRAN	0.56	
52	SAN AGUSTIN TLAXIACA	0.49	
3	ACTOPAN	0.48	
61	TEPEAPULCO	0.45	
31	JACALA DE LEZAMA	0.42	
16	CUAUTEPEC DE HINOJOSA	0.41	
5	AJACUBA	0.38	
56	SANTIAGO TULANTEPEC	0.32	
21	EMILIANO ZAPATA	0.27	

Servicios de Salud de Hidalgo		31/01/2004
Reporte de promedios de aportaciones de riesgo por indicador para el grupo: 3		
con el algoritmo: conexas		Grupo de Edad: Menor de 5 años 1999.xls
Indicador	Promedio de riesgo	
Casos de EPV	0.73	
Marginalidad	0.55	
Mort. IRAS 1-4 años	0.26	
Mort. Desm. 1-4 años	0.25	
Mort. Acciden. 1-4 años	0.24	
Mort. Desm. < 1 año	0.24	
Mort. IRAS < 1 año	0.19	
Mort. EDAS < 1 año	0.18	
Morb. IRAS < 1 año	0.18	
Morb. IRAS 1-4 años	0.18	
Morb. EDAS 1-4 año	0.17	
Morb. EDAS < 1 año	0.17	
Mort. EDAS 1-4 año	0.14	
Cobertura de vac. < 1 año	0.01	
Cobertura Antisár. 14 meses	0.00	

Tabla 18. Reportes para el grupo 3 aplicando el criterio Componentes

SERVICIOS DE SALUD DE HIDALGO		31/01/2004
REPORTE DEL GRUPO DE EDAD: Menor de 5 años 1999.xls		
GRUPO: 4	Promedio de riesgo:	0.32
ALGORITMO: conexas	Total de municipios del grupo:	17
Número	Municipio	Aportación de riesgo
43	NICOLAS FLORES	0.70
71	TLAHUILTEPA	0.46
18	CHAPULHUACAN	0.42
27	HUEHUETLA	0.38
24	HUASCA DE OCAMPO	0.37
59	TECOZAUTLA	0.36
68	TIANGUISTENGO	0.33
80	YAHUALICA	0.32
74	TLAXCOAPAN	0.31
20	ELOXOCHITLAN	0.31
58	TASQUILLO	0.28
63	TEPEH DE OCAMPO	0.27
35	METEPEC	0.26
65	TETEPANGO	0.23
83	ZEMPOALA	0.22
70	TLAHUELILPAN	0.16
72	TLANALAPA	0.12

Servicios de Salud de Hidalgo		31/01/2004
Reporte de promedios de aportaciones de riesgo por indicador para el grupo: 4		
con el algoritmo: conexas		Grupo de Edad: Menor de 5 años 199
Indicador	Promedio de riesgo	
Marginalidad	0.63	
Mort. Desm. < 1 año	0.23	
Mort. Acciden. 1-4 años	0.22	
Morb. EDAS < 1 año	0.22	
Mort. EDAS 1-4 año	0.21	
Morb. IRAS < 1 año	0.20	
Mort. IRAS < 1 año	0.20	
Mort. EDAS < 1 año	0.20	
Morb. IRAS 1-4 años	0.19	
Morb. EDAS 1-4 año	0.19	
Mort. IRAS 1-4 años	0.13	
Mort. Desm. 1-4 años	0.08	
Cobertura Antisár. 14 meses	0.01	
Cobertura de vac. < 1 año	0.01	
Detección de EPV	0.00	

Tabla 19. Reportes para el grupo 4 aplicando el criterio Componentes Conexas.

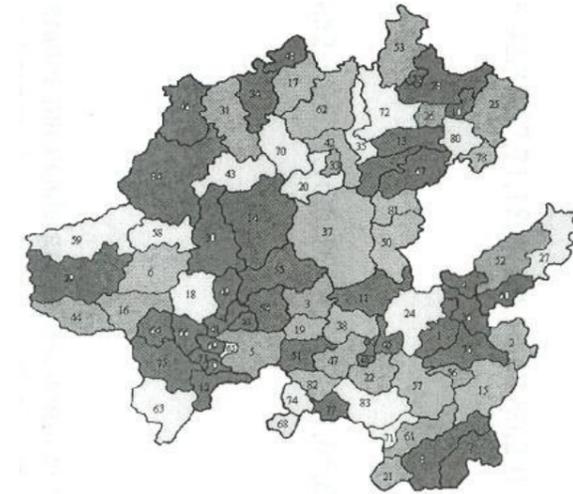


Figura 33. Resultados gráficos para el grupo de edad del menor de 5 años según componentes conexas.

### 6.2.2.- Estratificación utilizando A>means para el uso de funciones de similitud.

Las estratificaciones obtenidas a través de la implementación de este criterio agrupacional y con un número de grupos a formar igual a 4, se muestran en las tablas 20 a la 23 donde se pueden observar los 4 grupos obtenidos con su respectivo análisis de indicadores en función de su riesgo. La presentación de resultados de forma gráfica (mapa) se muestra en la figura 34, este mapa está coloreado de forma manual para acercarse más al mapa que obtuvieron los SSH. Los grupos obtenidos al aplicar dicho criterio agrupacional son los siguientes:

Grupo: 1

Total de objetos: 38

Promedio de Riesgo del grupo: 0.76

Municipios del grupo : {34,60,40,73,4,28,46,84,49,29,30,12,54,39,79,45,14,41,11,8,32,13,1,69,75,76,7,66,48,9,77,55,64,36,67,23,10,51}

Grupo: 2

Total de objetos: 19

Promedio de Riesgo del grupo: 0.54

Municipios del grupo : {2,62,6,43,15,78,37,19,42,38,57,71,18,27,24,59,68,80,20}

Grupo: 3

Total de objetos: 21

Promedio de Riesgo del grupo: 0.44

Municipios del grupo : {33,26,47,25,53,22,44,52,3,61,31,16,5,56,74,58,63,21,65,83,70}

Grupo: 4

Total de objetos: 6

Promedio de Riesgo del grupo: 0.41

Municipios del grupo: {17,81,50,82,35,72}

A continuación, se presentan los reportes generados para cada grupo, así como el reporte de cómo influyen los indicadores para el cálculo del promedio de riesgo final por grupo.

Para este grupo de tablas, se visualiza lo siguiente: el grupo 4 en la tabla 23, esta conformado por 6 municipios mientras que para componentes conexas esta conformado por 17 objetos. Sin duda es una diferencia considerable que solamente el experto puede decidir cual grupo representa mejor a los municipios en esa zona de riesgo. De igual forma, en los reportes de promedios de riesgo por indicador se muestra que para el grupo 1 tabla 20 también el indicador que más influyo al riesgo total del grupo es el mismo indicador que para el criterio componentes conexas. De hecho, el reporte de riesgos por indicador de la tabla 20 es el mismo que el de la tabla 16. Sin embargo, mientras que en el grupo 4 de componentes conexas el indicador que más influyo fue la marginalidad, en el grupo 4 de *k*-menas fue el indicador detección de EPV. La marginalidad fue el segundo indicador que más influyo.

SERVICIOS DE SALUD DE HIDALGO			31/01/2004
REPORTE DEL GRUPO DE EDAD:	Menor de 5 años 1999.xls		
GRUPO:	1	Promedio de riesgo:	0.76
ALGORITMO:	conexas	Total de municipios del grupo	38
Numero	Municipio	Aportación de riesgo	
34	LOLOTLA	1.00	
60	TENANGO DE DORIA	0.94	
40	LA MISION	0.93	
73	TLANCHINOL	0.92	
4	AGUA BLANCA DE ITURBIDE	0.91	
28	HUEJUTLA DE REYES	0.88	
46	SAN FELIPE DE ORIZATLAN	0.88	
84	ZIMAPÁN	0.86	
49	PISAFLORES	0.85	
29	HUICHAPAN	0.82	
30	IXMIQUILPAN	0.82	
12	ATOTONILCO EL GRANDE	0.82	
54	SAN SALVADOR	0.81	
39	MINERAL DEL MONTE	0.80	
79	XOCHICOATLAN	0.79	
45	OMITLAN DE JUAREZ	0.78	
14	CALNALI	0.78	
41	MIXQUIAHUALA	0.77	
11	ATLAPEXCO	0.76	
8	APAN	0.76	
32	JALTOCAN	0.76	
13	ATOTONILCO DE TULA	0.75	
1	ACATLAN	0.75	
69	TIZAYUCA	0.74	
75	TOLCAYUCA	0.74	
76	TULA DE ALLENDE	0.72	
7	ALMOLOYA	0.71	
66	VILLA DE TEZONTEPEC	0.71	
48	PACHUCA DE SOTO	0.70	

SERVICIOS DE SALUD DE HIDALGO		
Reporte de promedios de aportaciones de riesgo por indicador para el grupo:		
Con el algoritmo: conexas	Grupo de Edad:	Menor de 5 años 1999.xls
	<b>indicador</b>	<b>Promedio de riesgo</b>
	Deteccion de EPV	1.00
	Casos de EPV	0.78
	Marginalidad	0.52
	Morb. EDAS <1 año	0.20
	Morb. IRAS <1 año	0.19
	Morb. IRAS <1-4 años	0.19
	Morb. Desnu. <1 año	0.19
	Morb. EDAS <1-4 año	0.18
	Morb. EDAS <1 año	0.17
	Morb. IRAS <1 año	0.17
	Morb. IRAS <1-4 años	0.16
	Morb. Desnu <1-4 año	0.12
	Morb. EDAS <1-4 año	0.09
	Morb. Acceден. <1-4 año	0.08
	Cobertura de vac < 1 año	0.00
	Cobertura Antisar 14 meses	0.00

Tabla 20. Reportes para el grupo 1 usando *k*-means para funciones de similaridad

SERVICIOS DE SALUD DE HIDALGO			31/01/2004
REPORTE DEL GRUPO DE EDAD: Menor de 5 años 1999.xls			
GRUPO:	2	Promedio de riesgo:	0.54
ALGORITMO:	kmeans con similitud	Total de municipios del grupo:	19
Número	Municipio	Aportación de riesgo	
2	ACAXOCHITLAN	0.85	
62	TEPEHUACAN DE GUERRERO	0.76	
6	ALFAJAYUCAN	0.73	
43	NICOLAS FLORES	0.70	
15	CARDONAL	0.66	
78	XOCHIATIPAN	0.66	
37	METZITILAN	0.65	
19	CHILCUAUTLA	0.61	
42	MOLANGO DE ESCAMILLA	0.59	
38	MINERAL DEL CHICO	0.59	
57	SINGUILUCAN	0.56	
71	TLAHUILTEPA	0.46	
18	CHAPULHUACAN	0.42	
27	HUEHUETLA	0.38	
24	HUASCA DE OCAMPO	0.37	
59	TECOZAUTLA	0.36	
68	TIANGUISTENGO	0.33	
80	YAHUALICA	0.32	
20	ELOXOCHITLAN	0.31	

Servicios de Salud de Hidalgo		31/01/2004
Reporte de promedios de aportaciones de riesgo por indicador		
para el grupo:	2	Grupo de Edad: Menor de 5 años 1999.xls
con el algoritmo:	kmeans con similitud	
Indicador	Promedio de riesgo	
Marginalidad	0.87	
Detección de EPV	0.53	
Mort. Accident. 1-4 años	0.34	
Mort. EDAS 1-4 años	0.23	
Mort. Deter. 1-4 años	0.21	
Mort. Deter. < 1 año	0.20	
Mort. IRAS < 1 año	0.19	
Mort. EDAS < 1 año	0.18	
Mort. IRAS 1-4 años	0.18	
Morb. EDAS 1-4 años	0.17	
Morb. IRAS 1-4 años	0.17	
Morb. EDAS < 1 año	0.15	
Morb. IRAS < 1 año	0.15	
Cobertura Antibio: 14 meses	0.01	
Cobertura de vac. < 1 año	0.01	
Cuentos de EPV	0.00	

Tabla 21. Reportes para el grupo 2 usando k-means para funciones de similitud.

SERVICIOS DE SALUD DE HIDALGO			31/01/2004
REPORTE DEL GRUPO DE EDAD: Menor de 5 años 1999.xls			
GRUPO:	3	Promedio de riesgo:	0.44
ALGORITMO:	kmeans con similitud	Total de municipios del grupo:	21
Número	Municipio	Aportación de riesgo	
33	JUAREZ HIDALGO	0.76	
26	HUAZALINGO	0.68	
47	PACULA	0.66	
25	HUAUTLA	0.61	
53	SAN BARTOLO TUTOTEPEC	0.61	
22	EPAZOYUCAN	0.59	
44	HOPALA DE VILLAGRAN	0.56	
52	SAN AGUSTIN TLAXIACA	0.49	
3	ACTOPAN	0.48	
61	TEPEAPULCO	0.45	
31	JACALA DE LEZAMA	0.42	
16	CUAUTEPEC DE HINOJOSA	0.41	
5	AJACUBA	0.38	
56	SANTIAGO TULANTEPEC	0.32	
74	TLAXCOAPAN	0.31	
58	TASQUILLO	0.28	
63	TEPEL DE OCAMPO	0.27	
21	EMILIANO ZAPATA	0.27	
65	TETEPANGO	0.23	
83	ZEMPOALA	0.22	
70	TLAHUELILPAN	0.16	

Servicios de Salud de Hidalgo		31/01/2004
Reporte de promedios de aportaciones de riesgo por indicador		
para el grupo:	3	
con el algoritmo:	kmeans con similitud	Grupo de Edad: Menor de 5 años 1999.xls
Indicador	Promedio de riesgo	
Casos de EPV	0.32	
Marginalidad	0.49	
Mort. Desm. < 1 año	0.27	
Mort. Acciden. 1-4 años	0.22	
Mort. IRAS 1-4 años	0.22	
Mort. IRAS < 1 año	0.21	
Mort. EDAS < 1 año	0.20	
Morb. IRAS < 1 año	0.20	
Morb. EDAS < 1 año	0.19	
Morb. IRAS 1-4 años	0.19	
Mort. Desm. 1-4 años	0.18	
Morb. EDAS 1-4 años	0.16	
Mort. EDAS 1-4 años	0.14	
Cobertura de vac. < 1 año	0.00	
Cobertura Antioxiq. 14 meses	0.00	
Detección de EPV	0.00	

Tabla 22. Reportes para el grupo 3 usando &-means para funciones de similitud.

SERVICIOS DE SALUD DE HIDALGO		31/01/2004
REPORTE DEL GRUPO DE EDAD: Menor de 5 años 1999.xls		
GRUPO:	4	Promedio de riesgo: 0.41
ALGORITMO:	kmeans con similitud	Total de municipios del grupo: 6
Número	Municipio	Aportación de riesgo
17	CHAPANTONGO	0.61
81	ZACUALTIPAN DE ANGELES	0.59
50	PROGRESO	0.44
82	ZAPOTLAN DE JUAREZ	0.44
35	METEPEC	0.26
72	TLANALAPA	0.12

Servicios de Salud de Hidalgo		31/01/2004
Reporte de promedios de aportaciones de riesgo por indicador		
para el grupo:	4	
con el algoritmo:	kmeans con similitud	Grupo de Edad: Menor de 5 años 1999.xls
Indicador	Promedio de riesgo	
Detección de EPV	0.67	
Marginalidad	0.33	
Morb. EDAS < 1 año	0.27	
Morb. IRAS < 1 año	0.23	
Morb. IRAS 1-4 años	0.22	
Morb. EDAS 1-4 años	0.22	
Mort. Acciden. 1-4 años	0.19	
Mort. IRAS < 1 año	0.17	
Mort. EDAS < 1 año	0.17	
Mort. Desm. < 1 año	0.13	
Mort. EDAS 1-4 años	0.12	
Mort. IRAS 1-4 años	0.09	
Mort. Desm. 1-4 años	0.07	
Cobertura Antioxiq. 14 meses	0.00	
Cobertura de vac. < 1 año	0.00	
Casos de EPV	0.00	

Tabla 23. Reportes para el grupo 4 usando &-means para funciones de similaridad

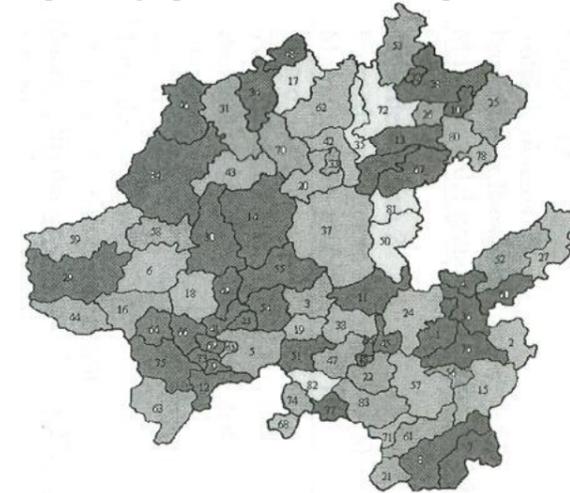


Figura 34. Resultados gráficos para el grupo de edad del menor de 5 años utilizando A:-means para funciones de similaridad.

### 6.2.3.- Estratificación generada usando conjuntos compactos

Las estratificaciones obtenidas a través de la implementación de este criterio agrupacional se muestran en las tablas 24 y 25. Donde únicamente se mostrarán los 2 primeros grupos debido al número alto de grupos formados. Así como su respectivo análisis de indicadores en función de su riesgo. La presentación de resultados de forma gráfica (mapa) se muestra en la figura 35, este mapa está coloreado de forma manual. Los grupos obtenidos al aplicar dicho criterio agrupacional son los siguientes:

Grupo : 1	Grupo : 11
Total de objetos: 4	Total de objetos:2
Promedio de Riesgo del grupo: 0.92	Promedio de Riesgo del grupo: 0.60
Municipios del grupo : {34,73,46,28}	Municipios del grupo : {17,81}
Grupo: 2	Grupo : 12
Total de objetos: 6	Total de objetos: 2
Promedio de Riesgo del grupo: 0.84	Promedio de Riesgo del grupo: 0.60
Municipios del grupo : {60,40,49,14,32,11}	Municipios del grupo : {67,23}
Grupo : 3	Grupo : 13
Total de objetos: 6	Total de objetos: 7
Promedio de Riesgo del grupo: 0.82	Promedio de Riesgo del grupo: 0.53
Municipios del grupo : {84,29,30,12,54,45}	Municipios del grupo : {33,22,44,52,3,31,16}
Grupo : 4	Grupo : 14
Total de objetos: 3	Total de objetos: 2
Promedio de Riesgo del grupo: 0.82	Promedio de Riesgo del grupo: 0.50
Municipios del grupo : {4,79,1}	Municipios del grupo : {10,51}
Grupo : 5	Grupo : 15
Total de objetos: 6	Total de objetos: 2
Promedio de Riesgo del grupo: 0.76	Promedio de Riesgo del grupo: 0.44
Municipios del grupo : {39,41,8,13,75,66}	Municipios del grupo : {50,82}
Grupo : 6	Grupo : 16
Total de objetos: 3	Total de objetos: 6
Promedio de Riesgo del grupo: 0.76	Promedio de Riesgo del grupo: 0.44
Municipios del grupo : {2,62,78}	Municipios del grupo : {43,71,18,27,68,80}
Grupo : 7	Grupo : 17
Total de objetos: 4	Total de objetos: 4
Promedio de Riesgo del grupo: 0.71	Promedio de Riesgo del grupo: 0.36
Municipios del grupo : {69,76,48,77}	Municipios del grupo : {61,5,56,21}
Grupo : 8	Grupo : 18
Total de objetos: 5	Total de objetos: 3
Promedio de Riesgo del grupo: 0.67	Promedio de Riesgo del grupo: 0.35
Municipios del grupo : {7,9,55,64,36}	Municipios del grupo : {24,59,20}
Grupo : 9	Grupo : 19
Total de objetos: 4	Total de objetos: 3
Promedio de Riesgo del grupo: 0.64	Promedio de Riesgo del grupo: 0.25
Municipios del grupo : {26,47,25,53}	Municipios del grupo : {58,35,83}
Grupo: 10	Grupo : 20
Total de objetos: 7	Total de objetos: 5
Promedio de Riesgo del grupo: 0.63	Promedio de Riesgo del grupo: 0.22
Municipios del grupo : {6,15,37,19,38,42,57}	Municipios del grupo : {74,63,65,70,72}

Como solamente se muestran 2 tablas representativas de los agrupamientos que generó conjuntos compactos, no se puede realizar un análisis detallado de los riesgos calculados para cada indicador.

En la figura 35, se observa el alto número de grupos formados por este criterio, por lo que fue necesario obtener el mapa coloreado de forma manual para permitir visualizar mejor los grupos. En este grupo de tablas se puede observar que el número de municipios pertenecientes a cada grupo es relativamente pequeño o dicho de otra forma es más estricta la estratificación. Como de puede ver en las tablas 24 y 25.

SERVICIOS DE SALUD DE HIDALGO			31/01/2004
<b>REPORTE DEL GRUPO Menor de 5 años 1999.xls</b>			
<b>DE EDAD:</b>			
<b>GRUPO:</b>	1	<b>Promedio de riesgo:</b>	0.92
<b>ALGORITMO:</b>	Conjuntos compactos	<b>Total de municipios del grupo</b>	4
	<b>Numero</b>	<b>Municipio</b>	<b>Aportación de riesgo</b>
	34	LOLOTLA	1.00
	73	TLANCHINOL	0.92
	46	SAN FELIPE DE ORIZATLAN	0.88
	28	HUEJUTLA DE REYES	0.88
SERVICIOS DE SALUD DE HIDALGO			31/01/2004
<b>Reporte de promedios de aportaciones de riesgo por indicador para el grupo: 1</b>			
<b>Con el algoritmo: conjuntos compactos</b>	<b>Grupo de Edad:</b>	<b>Menor de 5 años 1999.xls</b>	
	<b>indicador</b>	<b>Promedio de riesgo</b>	
	Deteccion de EPV	1.00	
	Casos de EPV	0.78	
	Marginalidad	0.52	
	Morb. EDAS <1 año	0.20	
	Morb. IRAS <1 año	0.19	
	Morb. IRAS <1-4 años	0.19	
	Morb. Desnu. <1 año	0.19	
	Morb. EDAS <1-4 año	0.18	
	Morb. EDAS <1 año	0.17	
	Morb. IRAS <1 año	0.17	
	Morb. IRAS <1-4 años	0.16	
	Morb. Desnu <1-4 año	0.12	
	Morb. EDAS <1-4 año	0.09	
	Morb. Acceden. <1-4 año	0.08	
	Cobertura de vac < 1 año	0.00	
	Cobertura Antisar 14 meses	0.00	

Tabla 24. Reportes para el grupo 1 usando el criterio Conjuntos Compactos.

SERVICIOS DE SALUD DE HIDALGO			31/01/2004
REPORTE DEL GRUPO DE EDAD: Menor de 5 años 1999.xls			
GRUPO:	2	Promedio de riesgo:	0.84
ALGORITMO:	conjuntos compactos	Total de municipios del grupo:	6

Número	Municipio	Aportación de riesgo
60	TENANGO DE DORIA	0.94
40	LA MISION	0.93
49	PISAFLORES	0.85
14	CALNALI	0.78
32	JALTOCAN	0.76
11	ATLAPEXCO	0.76

Servicios de Salud de Hidalgo		31/01/2004
Reporte de promedios de aportaciones de riesgo por indicador para el grupo: 2		
con el algoritmo: conjuntos compactos		Grupo de Edad: Menor de 5 años 1999.xls

Indicador	Promedio de riesgo
Deteccion de EPV	1.00
Marginalidad	1.00
Casos de EPV	0.50
Mort. Desem. 1-4 años	0.45
Mort. IRAS 1-4 años	0.28
Morb. EDAS 1-4 año	0.24
Morb. EDAS < 1 año	0.17
Mort. Desem. < 1 año	0.16
Morb. IRAS 1-4 años	0.15
Morb. IRAS < 1 año	0.15
Mort. Acciden. 1-4 años	0.15
Mort. EDAS 1-4 año	0.09
Mort. IRAS < 1 año	0.07
Mort. EDAS < 1 año	0.06
Cobertura Antisars. 14 meses	0.00
Cobertura de vac. < 1 año	0.00

Tabla 25. Reportes para el grupo 2 usando el criterio Conjuntos Compactos.

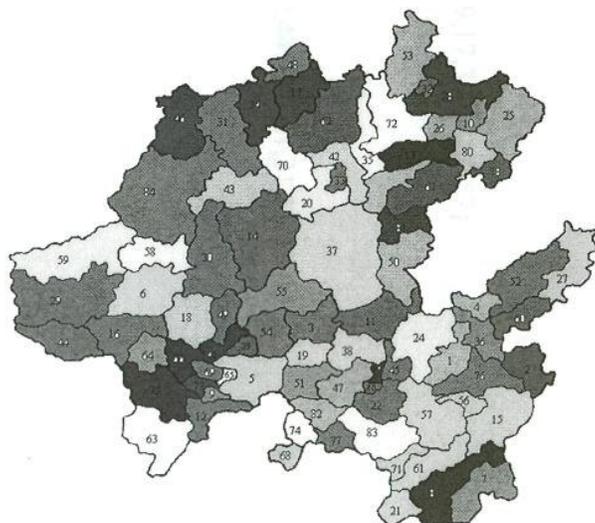


Figura 35. Resultados gráficos para el grupo de edad del menor de 5 años aplicando el criterio de conjuntos compactos

#### 6.2.4.- Estratificación obtenida utilizando conjuntos compactos con vecindad

Las estratificaciones obtenidas a través de la implementación de este criterio agrupacional usando  $ct = 0.036$ , se muestran de las tablas 26 a la 30 donde se pueden observar los 5 grupos obtenidos con su respectivo análisis de indicadores en función de su riesgo. La presentación de resultados de forma gráfica (mapa) se muestra en la figura 36, este mapa está coloreado de forma manual para acercarse más al mapa que obtuvieron los SSH. Los grupos obtenidos al aplicar dicho criterio agrupacional son los siguientes:

Grupo: 1

Total de objetos: 38

Promedio de Riesgo del grupo: 0.76

Municipios del grupo : {34,60,40,73,4,28,46,84,49,29,30,12,54,39,79,45,14,41,11,8,32,13,1,69,75,76,7,66,48,9,77,55,64,36,67,23,10,51}

Grupo: 2

Total de objetos: 12

Promedio de Riesgo del grupo: 0.66

Municipios del grupo : {2,62,6,15,78,37,19,17,38,42,81,57}

Grupo: 3

Total de objetos: 15

Promedio de Riesgo del grupo: 0.51

Municipios del grupo : {33,26,47,25,53,22,44,52,3,61,31,16,5,56,21}

Grupo: 4

Total de objetos: 2

Promedio de Riesgo del grupo: 0.44

Municipios del grupo: {50,82}

Grupo: 5

Total de objetos: 17

Promedio de Riesgo del grupo: 0.32

Municipios del grupo : {43,71,18,27,24,59,68,80,74,20,58,63,35,65,83,70,72}

A continuación, se presentan los reportes generados para cada grupo, así como el reporte de cómo influyen los indicadores para el cálculo del promedio de riesgo final por grupo.

En estas tablas, se observa otra estratificación, ahora son 5 grupos los obtenidos, pero como se visualiza en la tabla 29 el grupo 4 solamente tiene 2 municipios. En la tabla 26, se observa otra vez que el grupo 1 es el mismo que se ha generado con los demás criterios agrupacionales de hecho también son idénticos en cuanto a los indicadores que más influyeron para el riesgo del grupo. Por otro lado, en el grupo 5 (tabla 30), ahora se diferencia que el indicador que más influyó fue nuevamente la marginalidad al igual que sucedió con componentes conexas.

SERVICIOS DE SALUD DE HIDALGO			31/01/2004
REPORTE DEL GRUPO DE EDAD:			Menor de 5 años 1999.xls
GRUPO:	1		Promedio de riesgo: 0.76
ALGORITMO:	Conjuntos compactos con vecindad		Total de municipios del grupo 38
Numero	Municipio	Aportación de riesgo	
34	LOLOTLA	1.00	
60	TENANGO DE DORIA	0.94	
40	LA MISION	0.93	
73	TLANCHINOL	0.92	
4	AGUA BLANCA DE ITURBIDE	0.91	
28	HUEJUTLA DE REYES	0.88	
46	SAN FELIPE DE ORIZATLAN	0.88	
84	ZIMAPÁN	0.86	
49	PISAFLORES	0.85	
29	HUICHAPAN	0.82	
30	IXMIQUILPAN	0.82	
12	ATOTONILCO EL GRANDE	0.82	
54	SAN SALVADOR	0.81	
39	MINERAL DEL MONTE	0.80	
79	XOCHICOATLAN	0.79	
45	OMITLAN DE JUAREZ	0.78	
14	CALNALI	0.78	
41	MIXQUIAHUALA	0.77	
11	ATLAPEXCO	0.76	
8	APAN	0.76	
32	JALTOCAN	0.76	
13	ATOTONILCO DE TULA	0.75	
1	ACATLAN	0.75	
69	TIZAYUCA	0.74	
75	TOLCAYUCA	0.74	
76	TULA DE ALLENDE	0.72	

SERVICIOS DE SALUD DE HIDALGO			31/01/2004
Reporte de promedios de aportaciones de riesgo por indicador para el grupo:			Menor de 5 años 1999.xls
Con el algoritmo:	1	Grupo de Edad:	
compactos con vecindad			
indicador	Promedio de riesgo		
Deteccion de EPV	1.00		
Casos de EPV	0.78		
Marginalidad	0.52		
Morb. EDAS <1 año	0.20		
Morb. IRAS <1 año	0.19		
Morb. IRAS <1-4 años	0.19		
Morb. Desnu. <1 año	0.19		
Morb. EDAS <1-4 año	0.18		
Morb. EDAS <1 año	0.17		
Morb. IRAS <1 año	0.17		
Morb. IRAS <1-4 años	0.16		
Morb. Desnu <1-4 año	0.12		
Morb. EDAS <1-4 año	0.09		
Morb. Acceden. <1-4 año	0.08		
Cobertura de vac < 1 año	0.00		
Cobertura Antisar 14 meses	0.00		

Tabla 26. Reportes para el grupo 1 aplicando el nuevo criterio Conjuntos Compactos con Vecindad.

Servicios de Salud de Hidalgo		31/01/2004
Reporte de promedios de aportaciones de riesgo por indicador para el grupo: 5		
con el algoritmo: conjuntos compactos con vecindad		Grupo de Edad: Menor de 5 años 1999.xls
Indicador	Promedio de riesgo	
Marginalidad	0.63	
Mort. Desm. < 1 año	0.23	
Mort. Acciden. 1-4 años	0.22	
Morb. EDAS < 1 año	0.22	
Mort. EDAS 1-4 años	0.21	
Morb. IRAS < 1 año	0.20	
Mort. IRAS < 1 año	0.20	
Mort. EDAS < 1 año	0.20	
Morb. IRAS 1-4 años	0.19	
Morb. EDAS 1-4 años	0.19	
Mort. IRAS 1-4 años	0.13	
Mort. Desm. 1-4 años	0.08	
Cobertura Antisár. 14 meses	0.01	
Cobertura de vac. < 1 año	0.01	
Detección de EPV	0.00	
Casos de EPV	0.00	

SERVICIOS DE SALUD DE HIDALGO		
REPORTE DEL GRUPO DE EDAD: Menor de 5 años 1999.xls		
GRUPO:	2	Promedio de riesgo:
ALGORITMO:	conjuntos compactos con vecindad	Total de municipios del grupo:
Número	Municipio	Aportación de riesgo
2	ACAXOCHITLAN	0.85
62	TEPEHUACAN DE GUERRERO	0.76
6	ALFAJAYUCAN	0.73
15	CARDONAL	0.66
78	XOCHIATIPAN	0.66
37	METZTITLAN	0.65
19	CHILCUAUTLA	0.61
17	CHAPANTONGO	0.61
38	MINERAL DEL CHICO	0.59
42	MOLANGO DE ESCAMILLA	0.59
81	ZACUALTIPAN DE ANGELES	0.59
57	SINGUILUCAN	0.56

Tabla 27. Reportes para el grupo 2 aplicando el nuevo criterio Conjuntos Compactos con Vecindad.

Servicios de Salud de Hidalgo

31/01/2004

Reporte de promedios de aportaciones de riesgo por indicador

para el grupo: 2

con el algoritmo: conjuntos compactos con vecindad

Grupo de Edad: Menor de 5 años 1999.xls

Indicador	Promedio de riesgo
Deteccion de EPV	1.00
Marginalidad	0.77
Mort. Acciden. 1-4 años	0.41
Mort. Desm. 1-4 años	0.26
Mort. Desm. < 1 año	0.24
Mort. IRAS 1-4 años	0.19
Mort. IRAS < 1 año	0.19
Mort. EDAS < 1 año	0.18
Morb. IRAS 1-4 años	0.18
Mort. EDAS 1-4 años	0.17
Morb. EDAS 1-4 años	0.17
Morb. IRAS < 1 año	0.17
Morb. EDAS < 1 año	0.15
Cobertura Antisoz. 14 meses	0.00
Cobertura de vac. < 1 año	0.00

SERVICIOS DE SALUD DE HIDALGO

31/01/2004

REPORTE DEL GRUPO DE EDAD: Menor de 5 años 1999.xls

GRUPO: 3

Promedio de riesgo: 0.51

ALGORITMO: conjuntos compactos con vecindad

Total de municipios del grupo: 15

Número	Municipio	Aportación de riesgo
33	JUAREZ HIDALGO	0.76
26	HUAZALINGO	0.68
47	PACULA	0.66
25	HUAUTLA	0.61
53	SAN BARTOLO TUTOTEPEC	0.61
22	EPAZOYUCAN	0.59
44	NOPALA DE VILLAGRAN	0.56
52	SAN AGUSTIN TLAXIACA	0.49
3	ACTOPAN	0.48
61	TEPEAPULCO	0.45
31	JACALA DE LEZAMA	0.42
16	CUAUTEPEC DE HINOJOSA	0.41
5	AJACUBA	0.38
56	SANTIAGO TULANTEPEC	0.32
21	EMILIANO ZAPATA	0.27

Tabla 28. Reportes para el grupo 3 aplicando el nuevo criterio Conjuntos Compactos con Vecindad.

Servicios de Salud de Hidalgo

31/01/2004

Reporte de promedios de aportaciones de riesgo por indicador

para el grupo: 3

con el algoritmo: conjuntos compactos con vecindad

Grupo de Edad: Menor de 5 años 1999.xls

Indicador	Promedio de riesgo
Casos de EPV	0.73
Marginalidad	0.55
Mort. IRAS 1-4 años	0.26
Mort. Desm. 1-4 años	0.25
Mort. Acciden. 1-4 años	0.24
Mort. Desm. < 1 año	0.24
Mort. IRAS < 1 año	0.19
Mort. EDAS < 1 año	0.18
Morb. IRAS < 1 año	0.18
Morb. IRAS 1-4 años	0.18
Morb. EDAS 1-4 años	0.17
Morb. EDAS < 1 año	0.17
Mort. EDAS 1-4 años	0.14
Cobertura de vac. < 1 año	0.01
Cobertura Antisoz. 14 meses	0.00

SERVICIOS DE SALUD DE HIDALGO		31/01/21
REPORTE DEL GRUPO DE EDAD: Menor de 5 años 1999.xls		
GRUPO:	4	Promedio de riesgo: 0.44
ALGORITMO:	conjuntos compactos con vecindad	Total de municipios del grupo: 2
Número	Municipio	Aportación de riesgo
50	PROGRESO	0.44
82	ZAPOTLAN DE JUAREZ	0.44

Tabla 29. Reportes para el grupo 4 aplicando el nuevo criterio Conjuntos Compactos con Vecindad

Servicios de Salud de Hidalgo		31/01/2004
Reporte de promedios de aportaciones de riesgo por indicador para el grupo: 4		
con el algoritmo: conjuntos compactos con vecindad		Grupo de Edad: Menor de 5 años 1999.xls
Indicador	Promedio de riesgo	
Detección de EPV	1.00	
Magnitud	0.25	
Morb. EDAS < 1 año	0.24	
Mort. EDAS < 1 año	0.22	
Mort. IRAS < 1 año	0.20	
Morb. EDAS 1-4 años	0.17	
Morb. IRAS < 1 año	0.15	
Morb. IRAS 1-4 años	0.13	
Mort. IRAS 1-4 años	0.12	
Mort. EDAS 1-4 años	0.06	
Cobertura Antisár. 14 meses	0.01	
Cobertura de vac. < 1 año	0.01	
Mort. Accidan. 1-4 años	0.00	
Casos de EPV	0.00	
Mort. Derm. < 1 año	0.00	

SERVICIOS DE SALUD DE HIDALGO		31/01/2004
REPORTE DEL GRUPO DE EDAD: Menor de 5 años 1999.xls		
GRUPO:	5	Promedio de riesgo: 0.32
ALGORITMO:	conjuntos compactos con vecindad	Total de municipios del grupo: 17
Número	Municipio	Aportación de riesgo
43	NICOLAS FLORES	0.70
71	TLAHUILTEPA	0.46
18	CHAPULHUACAN	0.42
27	HUEHUETLA	0.38
24	HUASCA DE OCAMPO	0.37
59	TECOZAUTLA	0.36
68	TIANGUISTENGO	0.33
80	YAHUALICA	0.32
74	TLAXCOAPAN	0.31
20	ELOXOCHITLAN	0.31
58	TASQUILLO	0.28
63	TEPEH DE OCAMPO	0.27
35	METEPEC	0.26
65	TETEPANGO	0.23
83	ZEMPOALA	0.22
70	TLAHUELIPAN	0.16
72	TLANALAPA	0.12

Tabla 30. Reportes para el grupo 5 aplicando el nuevo criterio Conjuntos Compactos con Vecindad.

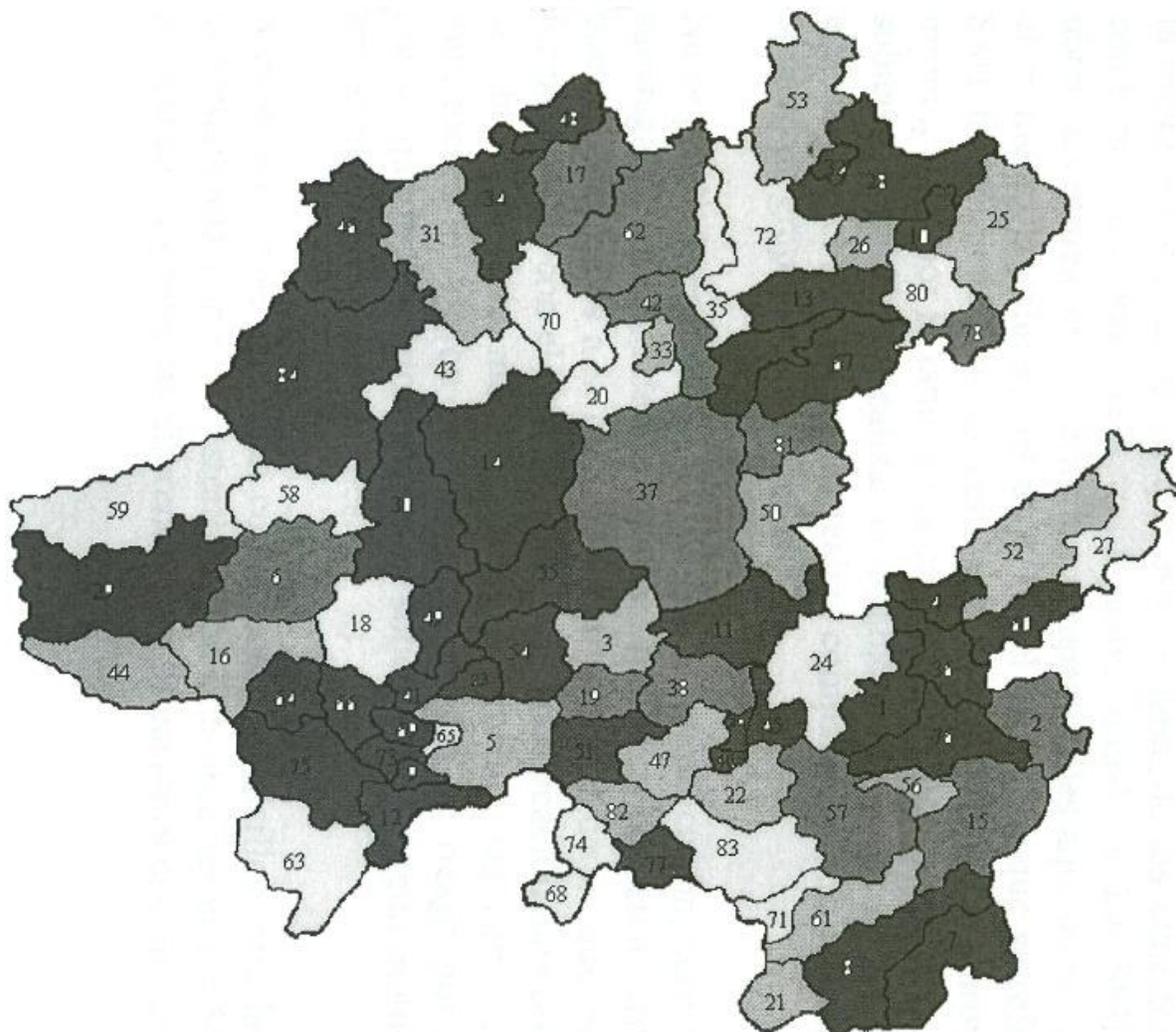


Figura 36. Resultados gráficos para el grupo de edad del menor de 5 años aplicando el nuevo criterio conjuntos compactos con vecindad.

### 6.3.- Análisis entre estratificaciones

Esta herramienta cuenta con un módulo que le permite al usuario realizar una revisión sencilla entre dos estratificaciones ya obtenidas por el sistema SAD. En ambos periodos se debe elegir el algoritmo y el archivo del grupo de edad que generó la estratificación que se desea verificar. En este análisis el usuario puede verificar la mejoría de un municipio de un periodo a otro, por ejemplo (año anterior con el actual). En otras palabras, si la aportación de riesgo de un municipio ha mejorado o empeorado al paso de un tiempo, si ha permanecido en el mismo grupo o ha cambiado para mejorar o empeorar. Además, de un análisis global entre el total de municipios de cada grupo para ambas estratificaciones. Esto no significa que el módulo sirve para determinar cuál es el mejor algoritmo o cuál es la mejor estratificación ya que esto es tarea de los expertos. A continuación se realizan comparaciones entre la estratificación actual de los SSH contra las estratificaciones obtenidas utilizando los criterios agrupacionales aquí manejados. Cabe mencionar que por el momento la columna (Riesgo 1) del periodo 1 que se refiere a la estratificación de los SSH no contiene valores debido a que ellos no calcularon aportaciones de riesgo por municipio. Y por lo tanto en la columna (Status) se observa que las aportaciones de riesgo subieron. También se realizan comparaciones entre los diferentes resultados que arrojan los algoritmos del sistema, no se realiza comparación con conjuntos compactos debido al gran número de grupos que este obtiene.

No se puede determinar cuál es la mejor estratificación, pero si se puede elegir la que mejor satisfaga los criterios de los expertos, solo ellos pueden determinar cuál es la estratificación que se apega más a la realidad. Es importante decir que esta herramienta se utilizará de forma paralela con el método actual de estratificación para poder evaluar los resultados que se obtengan. Este módulo de la herramienta así como el módulo que muestra que indicadores influyeron más en el promedio de riesgo por grupo son contribuciones al trabajo que los expertos no habían solicitado en ningún momento, son atribuciones para el apoyo en la toma de decisiones.

A continuación se presentan comparaciones sencillas entre las estratificaciones que genera la herramienta contra la estratificación que obtuvieron los SSH. Así también, se realizan comparaciones entre las diferentes estratificaciones que aporta la herramienta.

### 6.3.1- Comparación de la estratificación actual y la obtenida con el criterio agrupacional componentes conexas

El análisis comparativo entre la estratificación actual de los SSH contra la obtenida utilizando componentes conexas se muestra en la figura 37. Esta tabla contiene la siguiente información: total de grupos por algoritmo, cantidad de municipios pertenecientes a cada grupo, aportaciones de riesgo por municipios así como los cambios que dichas aportaciones de riesgo sufren de un periodo a otro y el color que se le asignó al municipio en dicha estratificación según el grupo. El reporte correspondiente se puede ver en la tabla 31.

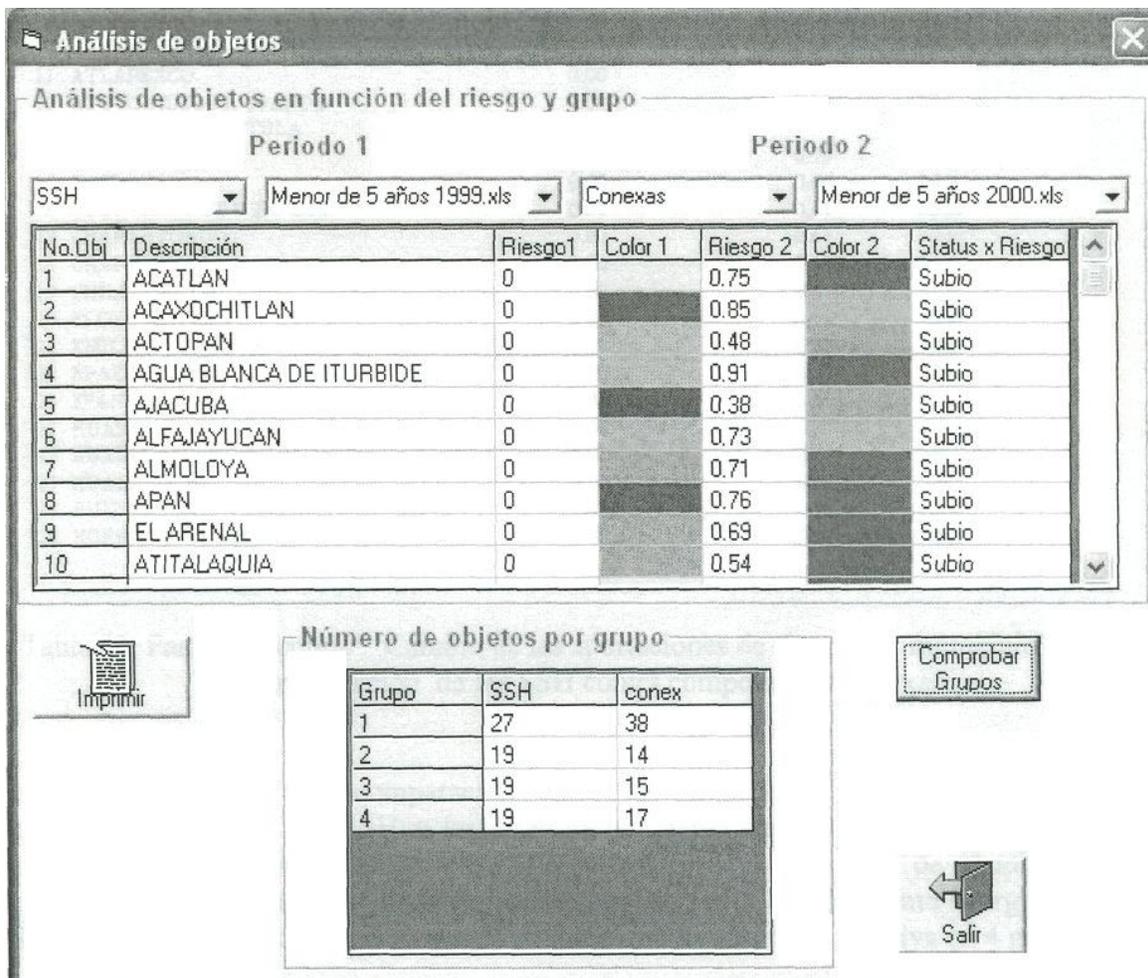


Figura 37. Análisis comparativo entre la estratificación actual (SSH) y la obtenida utilizando componentes conexas.

Periodo 1: **Menor de 5 años 1999.xls**  
 Periodo 2: **Menor de 5 años 1999.xls**

No. Objeto	Descripción del objeto	Aporta. Riesgo 1	Aporta. Riesgo 2	Cambio de riesgo
1	ACATLAN	0.00	0.75	Subió
2	ACAXOCHITLAN	0.00	0.85	Subió
3	ACTOPAN	0.00	0.48	Subió
4	AGUA BLANCA DE ITURBI	0.00	0.91	Subió
5	AJACUBA	0.00	0.38	Subió
6	ALFAJAYUCAN	0.00	0.73	Subió
7	ALMOLOYA	0.00	0.71	Subió
8	APAN	0.00	0.76	Subió
9	EL ARENAL	0.00	0.69	Subió
10	ATITALAQUIA	0.00	0.54	Subió
11	ATLAPEXCO	0.00	0.76	Subió
12	ATOTONILCO EL GRANDE	0.00	0.82	Subió
13	ATOTONILCO DE TULA	0.00	0.75	Subió
14	CALNALI	0.00	0.78	Subió
15	CARDONAL	0.00	0.66	Subió
16	CUAUTEPEC	0.00	0.41	Subió
17	CHAPANTONGO	0.00	0.61	Subió
18	CHAPULHUACAN	0.00	0.42	Subió
19	CHILCUAUTLA	0.00	0.61	Subió
20	ELOXOCHITLAN	0.00	0.31	Subió
21	EMILIANO ZAPATA	0.00	0.27	Subió
22	APAZOYUCAN	0.00	0.59	Subió
23	FRANCISCO I MADERO	0.00	0.59	Subió
24	HUASCA DE OCAMPO	0.00	0.37	Subió
25	HUAUTLA	0.00	0.61	Subió
26	HUAZALINGO	0.00	0.68	Subió
27	HUEHUETLA	0.00	0.38	Subió
28	HUEJUTLA DE REYES	0.00	0.88	Subió

Tabla 31. Parte del reporte "Cambio de las aportaciones de riesgo por municipio" " entre la estratificación de los SSH contra componentes conexas.

Una vez realizada una comparación entre los municipios que pertenecen a cada estratificación (la de los SSH y la generada por componentes conexas) se observó que ambos resultados corresponden en un 28% en función del total de municipios que pertenecen a cada grupo, además de considerar que en el agrupamiento otorgado por los SSH son 84 municipios que se reparten en forma prácticamente equitativa en 4 grupos.

### 6.3.2- Comparación de la estratificación actual y la obtenida utilizando k-means para uso de funciones de similitud

El análisis comparativo entre la estratificación actual de los SSH contra la obtenida utilizando k-means para uso de funciones de similaridad, se muestra en la figura 38. Esta tabla contiene la siguiente información: total de grupos por algoritmo, cantidad de municipios pertenecientes a cada grupo, aportaciones de riesgo por municipios así como los cambios que dichas aportaciones de riesgo sufren de un periodo a otro y el color que se le asignó al municipio en dicha estratificación según el grupo. El reporte correspondiente se puede ver en la tabla 32.

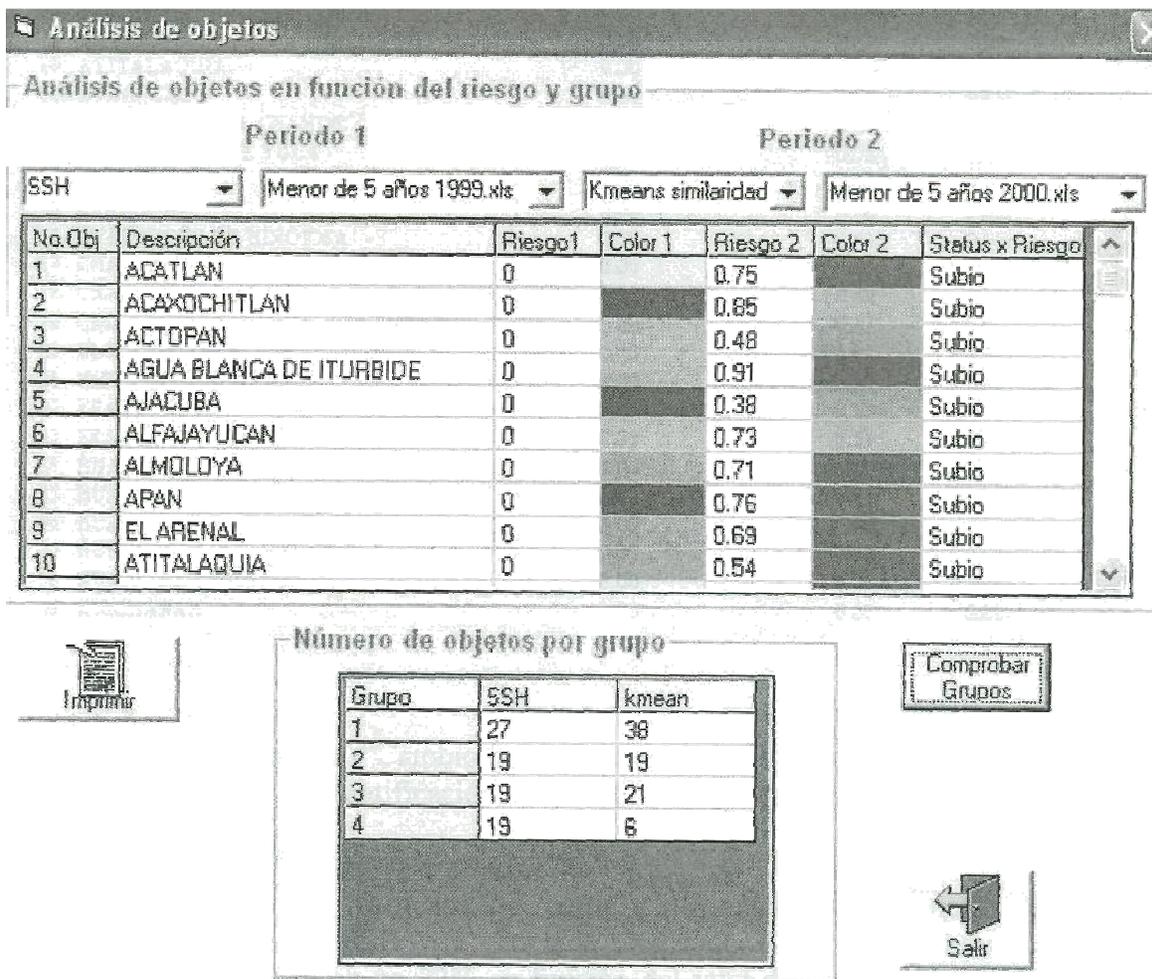


Figura 38. Análisis comparativo entre la estratificación actual (SSH) y la obtenida utilizando &-means para uso de funciones de similaridad.

No. Objeto	Descripción del objeto	Aporta. Riesgo 1	Aporta. Riesgo 2	Cambio de riesgo
1	ACATLAN	0.00	0.75	Subió
2	ACAXOCHITLAN	0.00	0.85	Subió
3	ACTOPAN	0.00	0.48	Subió
4	AGUA BLANCA DE ITURBI	0.00	0.91	Subió
5	AJACUBA	0.00	0.38	Subió
6	ALFAJAYUCAN	0.00	0.73	Subió
7	ALMOLOYA	0.00	0.71	Subió
8	APAN	0.00	0.76	Subió
9	EL ARENAL	0.00	0.69	Subió
10	ATITALAQUIA	0.00	0.54	Subió
11	ATLAPEXCO	0.00	0.76	Subió
12	ATOTONILCO EL GRANDE	0.00	0.82	Subió
13	ATOTONILCO DE TULA	0.00	0.75	Subió
14	CALNALI	0.00	0.78	Subió
15	CARDONAL	0.00	0.66	Subió
16	CUAUTEPEC	0.00	0.41	Subió
17	CHAPANTONGO	0.00	0.61	Subió
18	CHAPULHUACAN	0.00	0.42	Subió
19	CHILCUAUTLA	0.00	0.61	Subió
20	ELOXOCHITLAN	0.00	0.31	Subió
21	EMILIANO ZAPATA	0.00	0.27	Subió
22	APAZOYUCAN	0.00	0.59	Subió
23	FRANCISCO I MADERO	0.00	0.59	Subió
24	HUASCA DE OCAMPO	0.00	0.37	Subió
25	HUAUTLA	0.00	0.61	Subió
26	HUAZALINGO	0.00	0.68	Subió
27	HUEHUETLA	0.00	0.38	Subió
28	HUEJUTLA DE REYES	0.00	0.88	Subió
29	HUICHAPAN	0.00	0.82	Subió

Tabla 32. Parte del reporte "Cambio de las aportaciones de riesgo por municipio" " entre la estratificación de los SSH contra &-means.

Una vez realizada una comparación entre los municipios que pertenecen a cada estratificación (la de los SSH y la generada utilizando &-means) se observó que ambos resultados corresponden en un 23% en función del total de municipios que pertenecen a cada grupo, además de considerar que en el agrupamiento otorgado por los SSH son 84 municipios que se reparten en forma prácticamente equitativa en 4 grupos.

### 6.3.3- Comparación de la estratificación actual y la obtenida utilizando conjuntos compactos con vecindad

A continuación se muestra en la figura 39, el análisis comparativo entre la estratificación

actual de los SSH contra la obtenida utilizando conjuntos compactos con vecindad. En este análisis se observa que no se puede visualizar el número de municipios por grupo debido a que son dos estratificaciones con número de grupos diferentes.

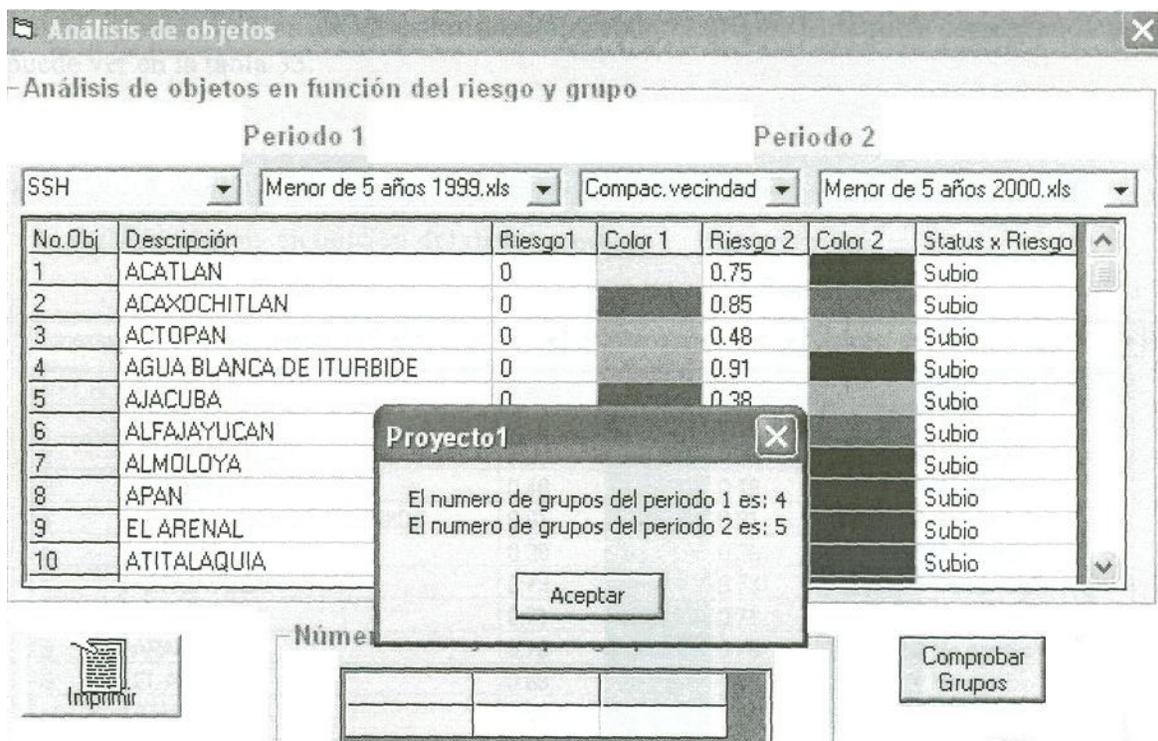


Figura 39. Análisis comparativo entre la estratificación actual (SSH) y la obtenida utilizando conjuntos compactos con vecindad.

Una vez realizada una comparación entre los municipios que pertenecen a cada estratificación (la de los SSH y la generada usando conjuntos compactos con vecindad) se observó que ambos resultados corresponden en un 27% en función del total de municipios que pertenecen a cada grupo, además de considerar que en el agrupamiento otorgado por los SSH son 84 municipios que se reparten en forma prácticamente equitativa en 4 grupos.



### Reporte de cambios de las aportaciones de riesgo

Periodo 1: Menor de 5 años 1999.xls

Periodo 2: Menor de 5 años 1999.xls

No. Objeto	Descripción del objeto	Aporta. Riesgo 1	Aporta. Riesgo 2	Cambio de riesgo
1	ACATLAN	0.75	0.75	No cambio
2	ACAXOCHITLAN	0.85	0.85	No cambio
3	ACTOPAN	0.48	0.48	No cambio
4	AGUA BLANCA DE ITURBI	0.91	0.91	No cambio
5	AJACUBA	0.38	0.38	No cambio
6	ALFAJAYUCAN	0.73	0.73	No cambio
7	ALMOLOYA	0.71	0.71	No cambio
8	APAN	0.76	0.76	No cambio
9	EL ARENAL	0.69	0.69	No cambio
10	ATITALAQUIA	0.54	0.54	No cambio
11	ATLAPEXCO	0.76	0.76	No cambio
12	ATOTONILCO EL GRANDE	0.82	0.82	No cambio
13	ATOTONILCO DE TULA	0.75	0.75	No cambio
14	CALNALI	0.78	0.78	No cambio
15	CARDONAL	0.66	0.66	No cambio
16	CUAUTEPEC	0.41	0.41	No cambio
17	CHAPANTONGO	0.61	0.61	No cambio
18	CHAPULHUACAN	0.42	0.42	No cambio
19	CHILCUAUTLA	0.61	0.61	No cambio
20	ELOXOCHITLAN	0.31	0.31	No cambio
21	EMILIANO ZAPATA	0.27	0.27	No cambio

Tabla 33. Parte del reporte "Cambio de las aportaciones de riesgo por municipio" entre la estratificación obtenida usando componentes conexas contra k-means.

Una vez realizada una comparación entre los municipios que pertenecen a cada estratificación (usando componentes conexas y A:-means) se observó que ambos resultados corresponden en un 77% en función del total de municipios que pertenecen a cada grupo, además de considerar que en el agolpamiento otorgado por los SSH son 84 municipios que se reparten en forma prácticamente equitativa en 4 grupos.

En la figura 41 se puede visualizar el análisis gráfico de estos resultados, donde la experiencia de los expertos es fundamental para la interpretación de dichos mapas

coloreados en función del criterio agrupacional utilizado.

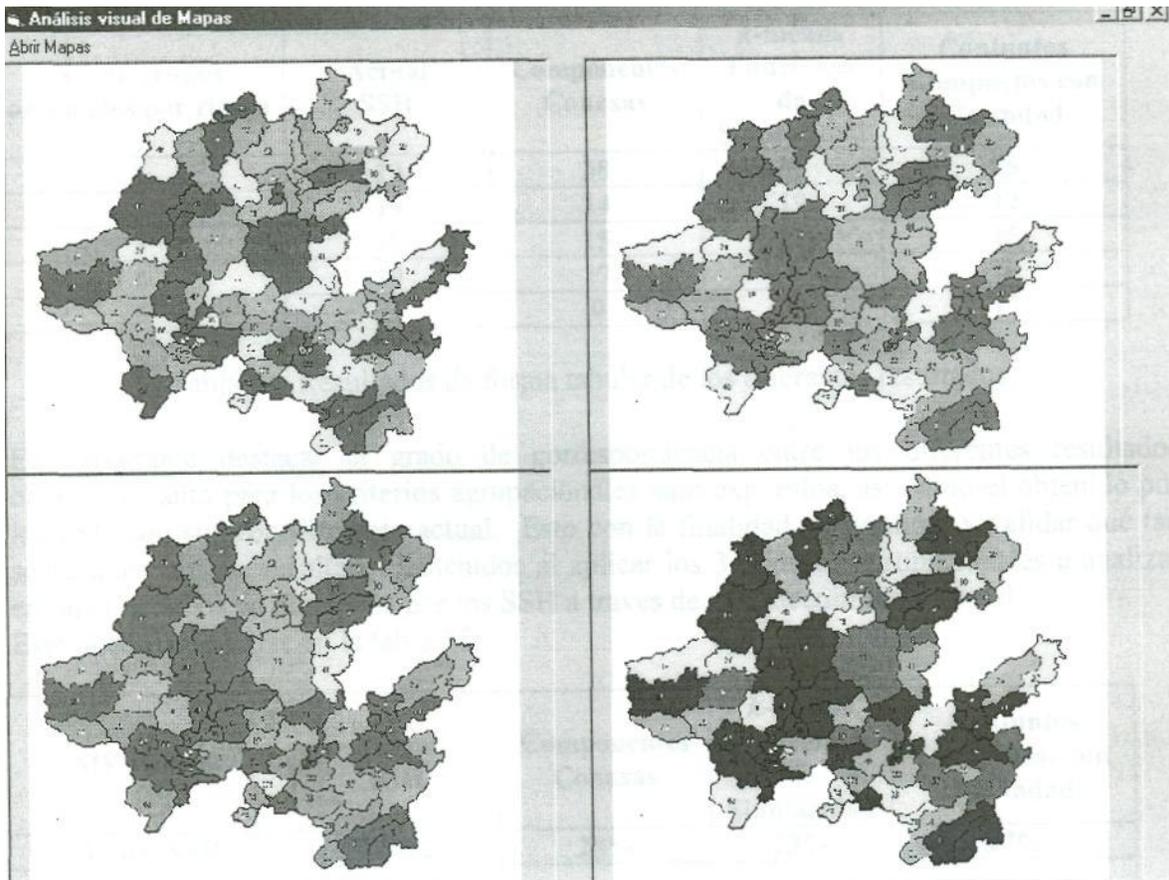


Figura 41. Análisis gráfico entre estratificaciones para el mismo grupo de edad.

#### 6.4.- Resultados entre estratificaciones.

En resumen, los resultados obtenidos (número de municipios por grupo) de los diferentes agrupamientos se presentan en la tabla 34.

No. De grupos ordenados por riesgo	Actual SSH	Componentes Conexas	k-means Funciones de Similaridad	Conjuntos Compactos con Vecindad
1	29	38	38	38
2	19	14	19	12
3	19	15	21	15
4	19	17	6	2
5	0	0	0	17

Tabla 34. Resultados de forma tabular de los diferentes resultados.

Es importante destacar el grado de correspondencia entre los diferentes resultados obtenidos, tanto para los criterios agrupacionales aquí expuestos, así como el obtenido por los SSH con su procedimiento actual. Esto con la finalidad de verificar y validar qué tan semejantes son los resultados obtenidos al aplicar los 3 criterios agrupacionales a analizar en función de los ya obtenidos por los SSH a través de su procedimiento actual. Esto puede observarse en la tabla 35.

MÉTODOS	Actual SSH	Componentes Conexas	k-means Funciones de Similaridad	Conjuntos compactos con Vecindad
Actual SSH	100%	28%	23%	27%
Componentes Conexas	28%	100%	77%	97%
k-means Funciones de Similaridad	23%	77%	100%	97%
Conjuntos compactos con Vecindad	27%	97%	97%	100%

Tabla 35. Correspondencia general entre métodos en función del número de municipios pertenecientes a cada grupo.

Sin embargo, debido a que el primer grupo es el más importante para los expertos, porque es el que presenta el mayor promedio de riesgo final, se realiza una comparación únicamente para el grupo 1 de todas las estratificaciones generadas.

El grupo. 1 de la estratificación obtenida por los SSH esta conformado por 27 municipios. En la tabla 36 se muestra el número de municipios con los que concuerdan cada uno de los criterios agrupacionales expuestos.

		k-means	
--	--	---------	--

TÉCNICAS	Componentes Conexas	Funciones de Similaridad	Conjuntos compactos con Vecindad
Números de municipios que concuerdan con el grupo 1 de los SSH	13 de 27	13 de 27	13 de 27

Tabla 36. Correspondencia entre número de municipios del grupo 1 de los SSH contra componentes conexas, k-means para funciones de similaridad y conjuntos compactos con vecindad.

Por otro lado, realizando este mismo procedimiento entre el grupo 1 de los 3 criterios agrupacionales a analizar se puede observar que generan el mismo grupo. Esto se visualiza en la tabla 37.

TÉCNICAS	Componentes Conexas	k-means Funciones de Similaridad	Conjuntos compactos con Vecindad
Números de municipios que concuerdan entre los 3 criterios agrupacionales para el grupo 1	38 de 38	38 de 38	38 de 38

Tabla 37. Correspondencia entre número de municipios del grupo 1 (de mayor riesgo) para componentes conexas, k-means para funciones de similaridad y conjuntos compactos con vecindad.

En conclusión, se puede visualizar que los resultados obtenidos por los diferentes criterios agrupacionales aquí expuestos se asemejan a bajo grado a los ya obtenidos por los SSH. Esto no significa que estos resultados sean erróneos. Debido a que los expertos no habían considerado algunas cosas importantes tales como el tipo de variable, pesos informacionales de los indicadores, forma de evaluar el riesgo de cada indicador y por consecuencia el riesgo del municipio entre otros.

Por otro lado, en forma experimental, se observa que los resultados que genera esta herramienta computacional aplicando los diferentes criterios agrupacionales se asemejan bastante, a pesar de que son criterios diferentes. No se realizó la comparación con los resultados que arroja el criterio de conjuntos compactos debido a la cantidad alta de grupos (20 grupos) que genera, y que no son de mucha utilidad para los expertos.

## Conclusiones y Trabajo futuro

En este capítulo se presentan las conclusiones del trabajo de tesis desarrollado, así como el trabajo futuro exponiendo las direcciones que se pueden seguir a partir de los resultados obtenidos. También se presentan las publicaciones generadas como resultado de este trabajo.

## Conclusiones

Con el desarrollo de esta tesis, se cumplió con el objetivo de proporcionar a los expertos de los SSH una herramienta computacional diseñada según sus requerimientos y necesidad capaz de obtener estratificaciones de municipios en zonas o grupos de riesgo de salud. Además de contar con módulos específicos para el apoyo en la toma de decisiones, como son: a) manejo del historial de las estratificaciones obtenidas en periodos anteriores y poder ser consultadas las veces que deseen, b) comparación de estratificaciones, permitiendo verificar la mejoría de los municipios, c) cálculo automático de las tasas, índices y porcentajes que anteriormente realizaban de forma manual y por separado. También es importante mencionar que la lectura de los datos de los SSH se realiza de forma automática y sencilla para los expertos, utilizando un formato en Excel que ellos mismos propusieron, así como un módulo para el coloreo automático de municipios del mapa del Estado de Hidalgo en función del grupo que corresponda. Este coloreo puede ser de forma manual o automática dependiendo de las necesidades de los expertos.

Pueden contar con una base de datos específica para cada uno de los diferentes grupos de edad que conforman al programa "Atención Integral", permitiendo realizar un análisis comparativo entre estratificaciones de forma gráfica o tabular, entre otras ventajas. Esta herramienta fue diseñada específicamente para proporcionar a los expertos en salud una solución al problema aquí planteado. Sin embargo, debido al diseño flexible del sistema desarrollado como el poder capturar cualquier tipo de gráfico (en formato BMP) como el mapa de Hidalgo. Además de aceptar un número variable de objetos e indicadores o variables (dependiendo de las características técnicas del hardware). El sistema es capaz de proporcionar soluciones a problemas de agrupamiento de datos similares.

Cabe mencionar que gracias al problema aquí planteado fue posible crear un nuevo criterio agrupacional derivado del criterio conjuntos compactos, denominado conjuntos compactos con vecindad, el cual ya ha sido aceptado y publicado. Este nuevo criterio agrupacional genera agrupamientos intermedios entre los agrupamientos que generan componentes conexas y conjuntos compactos, además de que es capaz de obtener los resultados que generan estos dos criterios agrupacionales con un mismo umbral de semejanza. Sin embargo, una de las partes que requirió más tiempo en el desarrollo de esta tesis, definitivamente fue el obtener los criterios de comparación adecuados para cada uno de los indicadores que conforman el grupo de edad del menor de 5 años. En este proceso fue indispensable la colaboración e interacción continúa con el experto. Así como determinar cómo se evaluaría el riesgo de salud para cada indicador y la aportación informacional de cada uno de los indicadores. Esto último con el objetivo de poder etiquetar y determinar la aportación de riesgo total por grupo y conocer cuál es el grupo y qué municipios tiene un mayor riesgo de salud. Los resultados obtenidos han sido expuestos al experto, quien ha dado su visto bueno a las estructuraciones generadas. Además, el experto está convencido

que el simple hecho de automatizar el programa "Atención Integral" es sin duda un beneficio para la Institución, considerando que se pondrá a prueba de forma paralela con el procedimiento actual para poder evaluar su comportamiento. Es importante recalcar que los beneficios de este sistema serán a largo y mediano plazo, ya que se obtendrán resultados cada año.

También es indispensable decir que la herramienta fue realizada en función de las necesidades de los expertos y que esto trae consigo que dicha herramienta este enfocada a este tipo de problemas. Esta herramienta también tiene sus limitaciones las cuales se consideran para trabajo futuro.

## 6.2- Trabajo futuro

El trabajo realizado, pone de manifiesto diversos aspectos que podrían desarrollarse en trabajos futuros. En primer lugar es importante, poder introducir otros criterios de comparación, así como nuevos criterios agrupacionales del RLCP. Otra parte es la de diseñar y desarrollar un módulo de captura del gráfico más elaborado, que no acepte únicamente archivos en formato BMP. También, un módulo que determine de forma automática el valor adecuado de  $\alpha$  para el criterio agrupacional conjuntos compactos con vecindad. Otro trabajo importante es implementar un algoritmo de agrupamiento conceptual basado en el RLCP, para obtener propiedades de los grupos previamente generados con los algoritmos ya implementados.

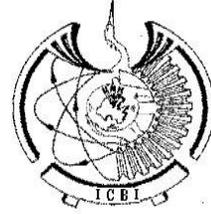
## Publicaciones generadas

1. López C. Marco and Sánchez G. "A new clustering criterion in pattern recognition", Proc. WSEAS en Artificial Intelligence, Knowledge Engineering and Data Bases (ADCED 2004 en CD), Salzburg, Austria, 2004.
2. López C. Marco and Sánchez G. "A new clustering criterion in pattern recognition", WSEAS TRANSACTIONS on COMPUTERS, Issue 3, Volume 3, July 2004 pp. 558-562.



UNIVERSIDAD AUTÓNOMA DEL  
ESTADO DE HIDALGO

INSTITUTO DE CIENCIAS BÁSICAS E  
INGENIERÍA



CENTRO DE INVESTIGACIÓN EN TECNOLOGÍAS DE INFORMACIÓN Y  
SISTEMAS

Pachuca, Hgo. a 19 de octubre de 2001

Dr. Alberto .Jonguitud Falcón.  
Director General de los Servicios de Salud de Hidalgo.  
PRESENTE

At'n, Dra, Ana Ma. Tavarèz Jiménez.  
Directora de los Servicios de Salud de Hidalgo.

El que suscribe, Profesor Investigador del Centro de Investigación en Tecnologías de Información y Sistemas (CITIS), de la Universidad Autónoma del Estado de Hidalgo (UAEH), solicita a ustedes otorguen su valioso apoyo de no mediar inconveniente alguno a los alumnos de la Maestría en Ciencias Computacionales del CITIS.

Genoveva Ruiz Casañas y  
Marco Antonio Lopez Caviedes

para que les sea autorizado establecer comunicación con personal de la secretaria que dignamente representan, y obtener información sobre la metodología y resultados de la estratificación de municipios de riesgo para la atención integral., y del programa de arranque parejo en la vida, ya que se pretende automatizar dichos procesos por medio de tecnicas de Reconocimiento de Patrones que se imparten en la Maestría del CITIS

Sin otro particular por el momento y agradeciendo de antemano su invaluable apoya les envió un cordial saludo.

A T E N T A M E N T E

Dr. Guillermo Sánchez Díaz  
Profesor - Investigador del CITIS

SERVICIOS DE SALUD DE HIDALGO	
DIRECCION GENERAL	
<b>RECIBI</b>	
FECHA:	18/10/01
NOMBRE:	Caviedes
NUMERO:	1334

Recibido en  
18-10-2001  
13:34 p.m.

c.cp. Expediente.

ASUNTO: CONSTANCIA

Pachuca Hgo. a 1 de Junio de 2004

A QUIEN CORRESPONDA:

El que suscribe DR. JUAN RUIZ CASAÑAS MAESTRO EN SALUD PÚBLICA POR EL INSTITUTO NACIONAL DE SALUD PÚBLICA DE LA SSA, ADSCRITO A LOS SERVICIOS DE SALUD DE HIDALGO (SSH). Y ASESOR DE LA TESIS DE MAESTRÍA "HERRAMIENTA COMPUTACIONAL DE VALIDACIÓN Y VERIFICACIÓN PARA LA ESTRATIFICACIÓN DE MUNICIPIOS EN ZONAS DE RIESGO PARA LA ATENCIÓN INTEGRAL DE LA SALUD EN EL ESTADO DE HIDALGO" la cual es sustentada por el C. LIC. EN COMPUTACIÓN MARCO ANTONIO LÓPEZ CAVIEDES Para obtener el Grado de "MAESTRÍA EN CIENCIAS COMPUTACIONALES, POR LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO", hace constar lo siguiente:

1. El sistema ha sido utilizado por el suscrito para establecer la "Estratificación de Municipios de Riesgo en el Estado de Hidalgo" para la Atención Integral a la Salud en el grupo de edad: Menor de 5 años, encontrándose un ambiente amigable en el manejo del mismo, aún para personas con pocos conocimientos en materia de Informática.
2. La versatilidad del sistema, permite simplificar los procesos de: Lectura de base de datos de Excel; Cálculo automático de tasas, porcentajes e índices que requieren los indicadores para los diferentes grupos de edad. Así mismo permite obtener la aportación informacional por cada indicador haciendo uso de la estadística; Generar estratificaciones de los municipios utilizando 4 criterios agrupacionales de la familia del "Reconocimiento de Patrones", Clasificación "No supervisada", con la opción de obtener un número de grupos libres o restringido según las necesidades del usuario; Visualizar los resultados de forma tabular y de forma gráfica (mapa coloreado automática y manualmente).
3. Utilizando el Sistema, a través de los años, se podrá evaluar la aplicación de Programas de Salud y su impacto en la población a la que están destinados, tomando en cuenta la "Estratificación de Municipios de Riesgo".
4. El Sistema soporta cualquier número de municipios o indicadores por lo que se puede utilizar para la estratificación de "Municipios o Localidades de Riesgo" en otros Estados de la República Mexicana o de otros países, contando con la información primaria para construir dichos indicadores por grupo de edad (los que también pueden variar).
5. Los resultados obtenidos, son congruentes con la experiencia que se tiene en los Servicios de Salud de Hidalgo en cuanto a la ubicación en los estratos de riesgo de los 84 municipios que integran el Estado de Hidalgo, según los grupos de edad que se manejan.
6. Todo lo anterior se obtiene ya que para el diseño del Sistema se realizó un trabajo constante, coordinado y continuo de los Asesores (interno y Externo) y del sustentante, de acuerdo a las necesidades del usuario (SSH), que se fueron planteando y afinando durante el desarrollo de dicho Sistema, contándose también con el apoyo de catedráticos de la UAEH y de otras instituciones Educativas y Científicas.

**BIBLIOGRAFÍA:**

- [A1E+ 1997] Alba Cabrera E. y Lazo Cortés M. (1997) "Nuevas extensiones del concepto de testor para diferentes tipos de funciones de semejanza". Tesis doctoral, pp. 8-10. La Habana, Cuba.
- [AuJ+ 1970] Auguston J. y Minker J. (1970) "An analysis of some graph theoretical cluster techniques". J. ACM 17, pp. 571-588.
- [CaJ 1992] Carbonell J. (1992) "Machine Learning, Paradigms and Methods". MIT Press, U.S.A.
- [CLUSTER] CLUSTER 3.0 para windows, Macintosh y Linux, [www.bonsai.ims.u-tokyo.ac.jp](http://www.bonsai.ims.u-tokyo.ac.jp)
- [CON 1993] CONAPO (1993), "Indicadores socioeconómicos e índices de marginación municipal", México.
- [DeM+ 2003] De Cáceres M., Oliva F. Y Font Xavier (2003) "GINKO, un programa de análisis multivariante orientado a la clasificación basada en distancias", Congreso Nacional de Estadística e investigación Operativa. Depto. de Biología vegetal, Depto. de Estadística, Universidad de Barcelona, España.
- [DuB+ 1980] Dubes R. y Jain A. (1980) "Clustering Methodologies in Exploratory Data Analysis Advances in Computers". Academic Press, Vol. 19 New York, U.S.A.
- [DuR+ 2000] Duda R. y Hart P y Stork G. (2000) "Pattern Classification " 2<sup>nd</sup> ed. Wiley, New York, NY. U.S.A.
- [EsL 1977] Escudero L. (1977) "Reconocimiento de Patrones". Paraninfo. Madrid, España.
- [FMS+ 1994] Fundación Mexicana para la Salud (1994) "Propuestas para el avance del Sistema de Salud en México", Informe final, México.
- [FrC+ 1999] Fraley C. y Raftery A. (1999) "MCLUST: software for model-based cluster and discriminant análisis", Technical report 342, Dept. Statistics, University of Washington.
- [GaJ+ 1999] Garcia Serrano J. y Martínez Trinidad F. (1999) "Algoritmo C-means: extensión al caso duro, para el uso de funciones de similaridad" Memorias IV Simposio Iberoamericano de Reconocimiento de Patrones, Modelos matemáticos y herramientas computacionales para el análisis de datos, ICMAF pp. 317-326, La Habana, Cuba.
- [ISCLE] Integral Solutions, Clementine Data Mining, System: Tech Report Basingstoke RG21 4RG, U.K.
- [JaA+ 1996] Jain A. " Data clustering: A Review" (1996) Departament of Computer Science, Michigan State University, Technical Report

- [KaK+1991] Kaufman K. A., Michalski R.S. , Kershchberg L. (1991) " Mining for Knowledge in Databases :Goals and general description of trie INLEN system" Knowledge Discovery in Databases, Menlo Park, AAAI/MIT pp. 449-462.
- [KaL+ 1990] Kaufman L. y Rousseeuw P. (1990) " Finding Groups in Data: An Introduction to Cluster Analysis". John Wiley & Sons.
- [K1W 1996] Klosgen, W. Explora (1996): A multipattern and multistrategy discovery assistant (Advances in Knowledge Discovery and Data Minig) Menlo Park AAAI/MIT pp. 249-271.
- [KuS+ 1989] Kumate J. G. Soberon (1989) " Salud para todos ¿Utopía o realidad? ", el Colegio Nacional de México.
- [KuS+ 1993] Kumate J. G. Soberon (1993) " Medicina Preventiva, Medicina Curativa y Salud Pública en el Estado Moderno y DérezDdad Social", FCE, México.
- [LoM+2004] López C. Marco and Sánchez G. "A new clustering criterion in pattern recognition", WSEAS TRANSACTIONS on COMPUTERS, Issue 3, Volume 3, July2004 pp. 558-562.
- [LoS+1997] López Dérez S., Lazo Cortés M. Y Estrada García H. (1997) "Electrodiagnóstico médico utilizando las herramientas de Reconocimiento de Patrones". Memorias II Simposio Iberoamericano de Reconocimiento de Patrones, CIMAF-IPN pp. 237-244, La habana, Cuba.
- [MaF+ 2000] Martínez F., Ruiz J. y Lazo M. (2000) " Structuralization of Universes". Fuzzy Sets & Systems, Vol 112/3 pp. 485-500.
- [MaJ+2000] Martines Trinidad J., Velasco Sánchez M. y Contreras Arévalo E. (2000) "Feature Selections for classification of patients with Uveitis" 5<sup>th</sup> Iberoamerican syposium on pattern recognition, Instituto Superior Técnico. Universidad Técnica de Lisboa, Lisbon Portugal, pp. 459-466.
- [MÍG+ 1995] Milligan W. Glenn y Cheng Richard (1995)"&-means clustering algorithms with internal influence measurement", Versión 2.0, Ohio State University.
- [NgR+ 1994] Ng. Raymond y Jiawei H.(1994) "Efficient and Effective Clustering Methods for Spatial Data Mining". Proc. 20<sup>th</sup> Int. Conf. On Very Large data Bases, Santiago, Chile, pp. 144-155.
- [OM+ 1996] Ortíz Posadas MR, Martínez Trinidad F. y Ruíz Shulcloper J. (1996) "A new approach to differential diagnosis of diseases ". Int J Biomed Comput 40 pp. 179-185
- [OrM+ 1998] Ortíz Posadas M, Vega Alvarado L., Jiménez Jacinto V. y Lazo Cortés M. (1998) "El concepto de analogía en medicina. Una función de semejanza para pacientes con fisuras de paladar". Memorias III Simposio

Iberoamericano de Reconocimiento de Patrones, CINVESTAV-IPN, pp. 247-256. México.

- [PNS+ 2001] Programa Nacional de Salud (2001-2005) México.
- [RuJ+ 1992] Ruiz S. José et al. (1992) " PROGNOSIS y sus aplicaciones a las Geociencias". Memorias del III Congreso Iberoamericano de Inteligencia Artificial IBERAMIA'92. pp. 561-586. La Habana, Cuba
- [RuJ+ 1995a] Ruiz S. José y Lazo C. Manuel (1995) " Introducción al Reconocimiento de Patrones: Enfoque Lógico Combinatorio". Serie verde No. 51, CINVESTAV-IPN, México
- [RuJ+ 1995b] Ruiz S. José y Martínez T. (1995) " Clasificación sin aprendizaje y con aprendizaje parcial". (Enfoque Lógico Combinatorio). CINVESTAV-IPN, México
- [RuJ+ 1999a] Ruiz S. José, Guzmán A. Adolfo y Martínez Trinidad Francisco (1999) " Enfoque Lógico Combinatorio al Reconocimiento de Patrones" CIC-IPN, México
- [RuJ+ 2002a] Ruiz J. y Mongi A. (2002) "Logical Combinatorial Pattern Recognition: a review ". recent Research Developments in Pattern Recognition, S.G. Pándala (Ed), Transworld Research Networks, Kerala, India, pp. 133-176.
- [RuJ+ 2002b] Ruiz J., Sánchez Díaz G. y Mongi A. Abidi. (2002) " Clustering mixed incomplete data", Heuristics and Optimization for Knowledge Discovery, Ruhul Sarker, Hussein Abbass, and Charles S. Newton, by Idea Group Publishing, U.S.A., ISBN: 1-930708-26-2, pp.89-107.
- [SaG+ 2003] Sánchez G., Martínez F. y Waissman J. (2003), "A new method to determine a similarity threshold in clustering problems ". WSEAS Transactions on Computers. Volumen 2, pp. 191-196.
- [ScE+ 1996] Schikuta Erich y Erhart Martín (1996) "The BANG-Clustering System: Grid-Based Data Analysis", Institute of applied Computer Science and Information systems-, University of Vienna A-1010 Vienna, Austria
- [SGRAPH] Silicon Graphics, Inc. <http://www.sgi.com>
- [SÍN+ 1993a] Sistema Nacional de Salud (1993) "Boletín de Información Estadística Daños a la salud", Vol. II núm. 13 Secretaria de Salud
- [SiN+ 1993b] Sistema Nacional de Salud (1993)"Boletín de Información Estadística Recursos y Servicios ", Vol. II núm. 13, Secretaria de Salud
- [SSA+ 1999] SSA(1999-2005) "Programa Estatal de Salud". Hidalgo. México

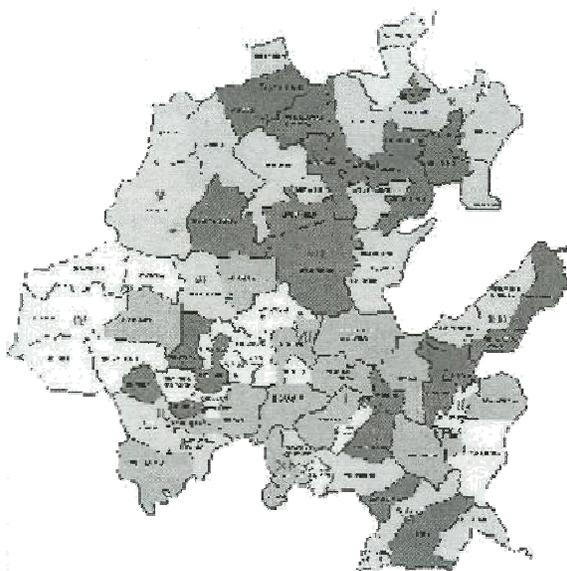
- [SSA+ 1994] SSA(1994) "Modelo de Atención a la Salud para Población Abierta ".  
Revisión y Actualización. México.
- [SSH 2000] Servicios de Salud de Hidalgo (2000) "Evaluación Sexenal", México.
- [ZhY+2001] Zhao Ying y Karypis George (2001) "Clustering in Life Sciences",  
Department of Computer Science, University of Minnesota, Minneapolis,  
MN 55455

# ANEXOS

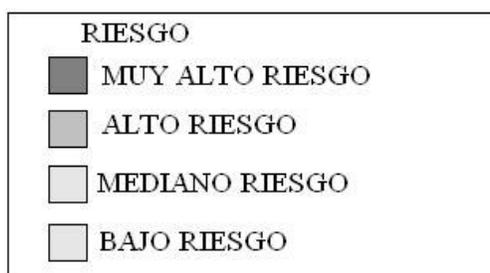
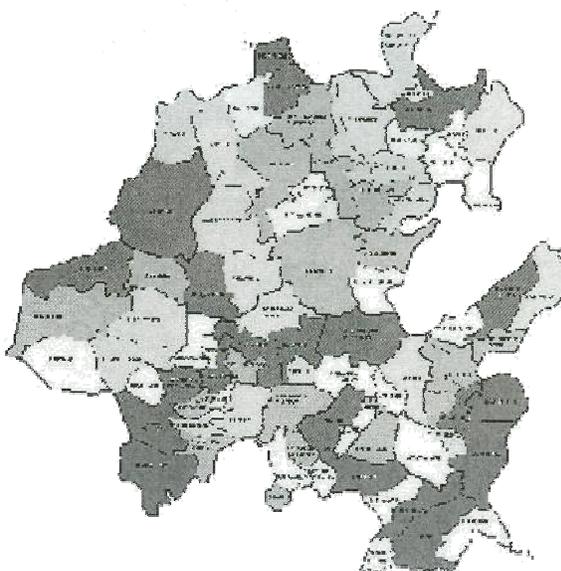
## ANEXO A

### ESTRATIFICACIÓN DE MUNICIPIOS EN ZONAS DE RIESGO PARA LA ATENCIÓN INTEGRAL POR GRUPO DE EDAD

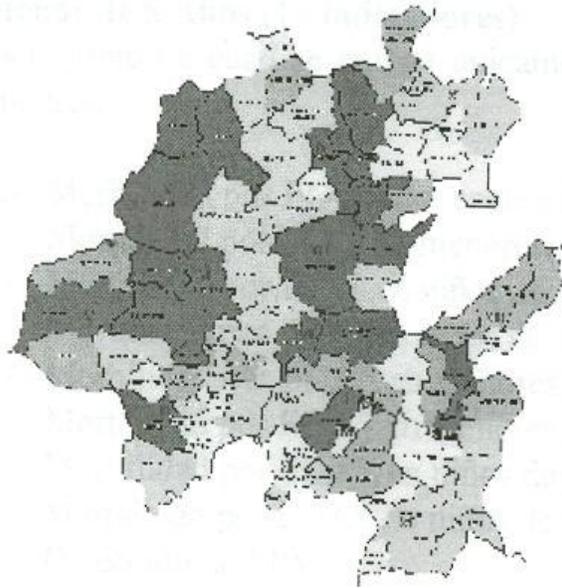
#### DEL ADOLESCENTE



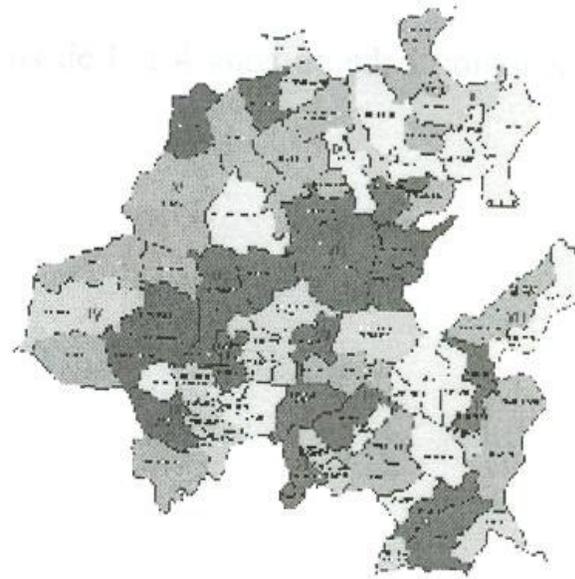
#### DE LA MUJER



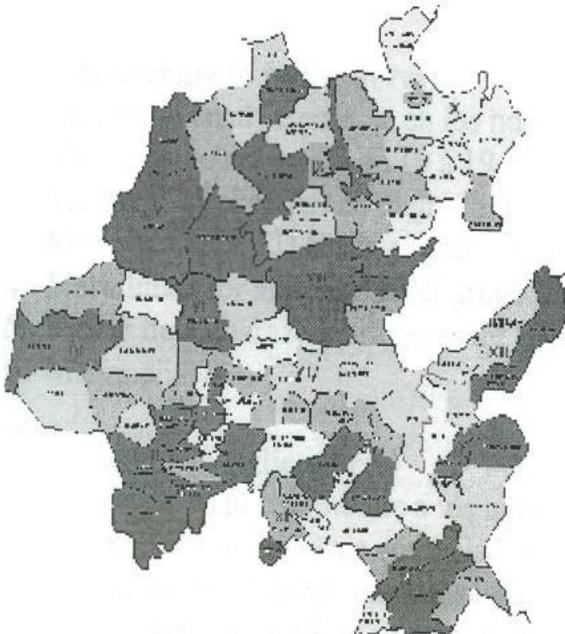
DEL ADULTO



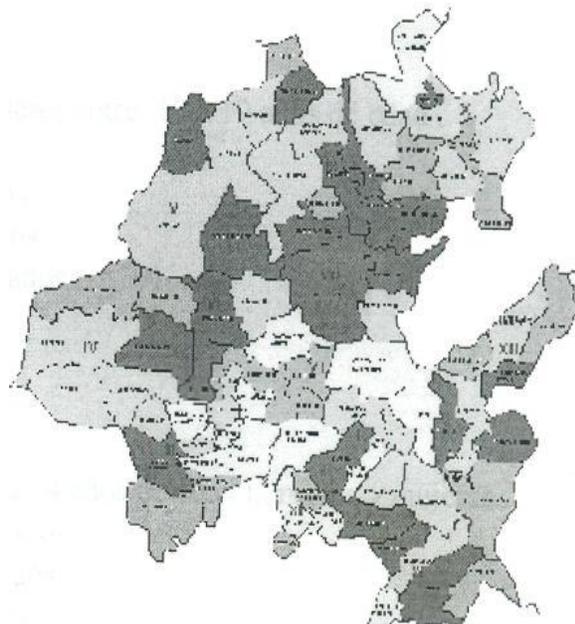
DEL ADULTO MAYOR



DEL MENOR DE 5 AÑOS



DEL ESCOLAR



## ANEXO B

## **Indicadores de salud por grupo de edad**

### **Menor de 5 Años (18 indicadores)**

Este grupo de edad se enfoca únicamente en los niños de 0 a 4 años de edad hombres y mujeres.

- 1.- Morbilidad por ERAS en menores de un año
- 2.- Mortalidad por IRAS en menores de un año
- 3.- Morbilidad por IRAS en niños de 1 a 4 años
- 4.- Mortalidad por IRAS en niños de 1 a 4 años
- 5.- Morbilidad por EDAS en menores de un año
- 6.- Mortalidad por EDAS en menores de un año
- 7.- Morbilidad por EDAS en niños de 1 a 4 años
- 8.- Mortalidad por EDAS en niños de 1 a 4 años
- 9.- Detección de EPV
- 10.- Casos de EPV
- 11.- Morbilidad por desnutrición en menores de un año
- 12.- Mortalidad por desnutrición en menores de un año
- 13.- Morbilidad por desnutrición en niños de 1 a 4 años
- 14.- Mortalidad por desnutrición en niños de 1 a 4 años
- 15.- Mortalidad por accidentes en niños de 1 a 4 años
- 16.- Coberturas de vacunación en niños del año
- 17.- Coberturas de vacunación antisarampión a los 14 meses
- 18.-Marginalidad

### **Del adolescente (6 indicadores)**

Este grupo de edad se enfoca en los hombres y mujeres entre 10 a 19 años de edad .

- 1.- Mortalidad en el grupo de 10 a 19 años
- 2.- Morbilidad por IRAS en el grupo de 10 a 19 años
- 3.- Morbilidad por EDAS en el grupo de 10 a 19 años
- 4.- Morbilidad por ambiasis en el grupo de 10 a 19 años
- 5.- Riesgo de embarazo en mujeres de 10 a 19 años
- 6.- Marginalidad

### **Edad escolar (6 indicadores)**

Este grupo de edad comprende a los niños entre 5 a 14 años de edad hombres y mujeres.

- 1.- Morbilidad por ambiasis en población de 5 a 14 años
- 2.- Morbilidad por EDAS en población de 5 a 14 años
- 3.- Morbilidad por IRAS en población de 5 a 14 años
- 4.- TX. Antirrábicos en población de 5 a 14 años
- 5.- Mortalidad por accidentes en población de 5 a 14 años
- 6.- Marginalidad

### **De la mujer (11 indicadores)**

En este grupo sólo se consideran los mujeres entre los 20 a 45 años

- 1.- Mortalidad por Ca. Cu.
- 2.- Número de defunciones por Ca.Cu.
- 3.- Defunciones por Ca. Mama
- 4.- Mortalidad por tumores malignos
- 5.- Mortalidad materna
- 6.- Mortalidad perinatal
- 7.- Mortalidad por Ca.Cu. (insitu e invasor)
- 8.- Morbilidad por ITS
- 9.- Casos y defunciones por defectos del tubo neural
- 10.- %papaniculau
- 11.- Grado de marginación

### **Del adulto (8 indicadores)**

En este grupo sólo se consideran los hombres entre los 20 a 45 años

- 1.- Tasa de mortalidad por accidentes, traumatismos
- 2.- Tasa de mortalidad por diabetes mellitas
- 3.- Tasa de mortalidad por hipertensión arterial
- 4.- Tasa de mortalidad por cirrosis hepática
- 5.- Tasa de mortalidad por tuberculosis pulmonar
- 6.- Tasa de morbilidad por diabetes mellitas
- 7.- Tasa de morbilidad por hipertensión arterial
- 8.- Tasa de morbilidad por tuberculosis pulmonar

### **Del adulto mayor (7 indicadores)**

En este grupo de edad los adultos mujeres y hombres mayores de 45 años son los que se consideran.

- 1.- Tasa de mortalidad por tumores malignos
- 2.- Tasa de mortalidad por diabetes mellitus
- 3.- Tasa de mortalidad por hipertensión arterial
- 4.- Tasa de mortalidad por enfermedad del corazón
- 5.- Tasa de mortalidad por enfermedad del cerebro vascular
- 6.- Tasa de morbilidad por hipertensión arterial
- 7.- Tasa de morbilidad por tuberculosis por tuberculosis pulmonar

## ANEXO C

### EJEMPLO DEL PROCEDIMIENTO DE ESTRATIFICACIÓN DE MUNICIPIOS EN ZONAS DE RIESGO QUE REALIZAN LOS SSH.

Paso 1.-

Se utilizará solamente el indicador de Mortalidad general en el grupo de edad del adolescente donde la tasa de dicho indicador se calcula de la siguiente manera:

$$Mort. = \frac{No. Casos \text{ o } No. Definiciones}{Poblacion} \times 100,000 \text{ habitantes}$$

Paso 2.-

Las tasas para este indicador se ordenan de forma descendente como se puede apreciar en las tablas A y B del Anexo D.

Paso 3.-

Una vez que las tasas ya están ordenadas se procede a estratificar los municipios por medio de una distribución por cuartiles :

$$Q0 = \frac{84}{100} \times 25 = 21 \quad Q1 = \frac{84}{100} \times 50 = 42 \quad Q2 = \frac{84}{100} \times 75 = 63 \quad Q3 = \frac{84}{100} \times 100 = 84$$

En las tablas A y B del Anexo D se puede observar la estratificación de municipios para el indicador de Mortalidad general según los resultados del procedimiento por cuartiles.

Paso 4.-

La ponderación del municipio de Actopan es de 4 como se puede ver en tabla C del Anexo F, por pertenecer a la zona de riesgo 4, y así sucesivamente con los demás municipios.

Paso 5.-

Se suman todas las ponderaciones por municipio de todos los indicadores (ver tabla 2) del Anexo F.

$$Ponderacion \ final = \sum_1^{84} pond_j \quad \text{donde } j = 1,2,3 \text{ y } 4$$

Donde:

$pond_i$  = ponderación de cada indicador para  $i = 1, \dots$ , total de indicadores, para el grupo de edad por municipio.

Paso 6.-

Ordenar las sumas de ponderaciones finales de todos los municipios de forma descendente para el indicador de Mortalidad General del grupo de edad del adolescente

Paso 7.-

Por último se utiliza nuevamente la distribución por cuartiles para obtener la ponderación final y así lograr estratificar los municipios por grupos de edad en zonas de riesgo como se puede observar en la tabla C del Anexo F, etiquetando las ponderaciones finales. La estratificación de municipios en zonas de riesgo que los SSH obtuvieron para el grupo de edad del menor de 5 años se observan en el Anexo E.

## ANEXO D

Tabla A. Mortalidad por municipio para el grupo de edad del adolescente 199  
"Distribución por Cuartiles"

MUNICIPIO	No. DE DEF.	POB.	TASA	
SAN AGUSTIN MEZQUITITLAN	4	2049	195.22	
TEPETITLAN	4	2118	188.86	
AJACUBA	5	3383	147.80	
METEPEC	3	2544	117.92	
HUAZALINGO	3	2616	114.68	
CHAPULHUACAN	6	5385	111.42	
EMILIANO ZAPATA	3	2945	101.87	
CHILCUAUTLA	4	3969	100.78	
ALFAJAYUCAN	4	4066	98.38	
TETEPANGO	2	2078	98.3896.25	
MINERAL DEL MONTE	3	3198	93.81	
ZIMAPAN	9	10089	89.21	
ATOTONILCO EL GRANDE	6	6772	88.60	
TLANALAPA	2	2341	85.43	
ACTOPAN	9	10898	82.58	
TIZAYUCA	8	9729	82.23	
ATITALAQUIA	4	4884	81.90	
EPAZOYUCAN	2	2458	81.37	
TEPEAPULCO	9	11362	79.21	
ALMOLOYA	2	2564	78.00	Ponderación 3
SAN AGUSTIN TLAXIACA	4	5155	77.59	DE 77.59 a +
CARDONAL	4	5185	77.15	
HUASCA DE OCAMPO	3	3990	75.19	
TULA DE ALLENDE	14	19890	70.39	
LA MISION	2	2849	70.20	
CHAPANTONGO	2	2943	67.96	
TLAHUILTEPA	2	3014	66.36	
HUAUTLA	4	6407	62.43	
PACULA	1	1616	61.88	
PACHUCA DE SOTO	30	48737	61.55	
TIANGUISTENGO	2	3261	61.33	
ZAPOTLAN DE JUAREZ	2	3275	61.07	
OMITLAN DE JUAREZ	1	1643	60.86	
TEPEJI DE OCAMPO	9	15132	59.48	
MIXQUIAHUALA	5	8786	56.97	
SAN SALVADOR	4	7099	56.35	
TEZONTEPEC DE ALDAMA	5	9059	55.19	
SANTIAGO TULANTEPEC	3	5456	54.99	
XOCHICOATLAN	1	1859	53.79	
CUAUTEPEC DE HINOJOSA	6	11450	52.40	
HUEHUETLA	3	5902	50.83	Ponderación 2
NICOLAS FLORES	1	1986	50.35	DE 50.35 A 77.15

Tabla B. Mortalidad por municipio para el grupo de edad

del adolescente 1998  
"Distribución por Cuartiles"

MUNICIPIO	No. DE DEF.	POB.	TASA	
TEPEHUACAN DE GUERRERO	3	5991	50.08	
TULANCINGO	13	26544	48.98	
TENANGO DE DORIA	2	4216	47.44	
JALTOCAN	1	2120	47.17	
ATLAPEXCO	2	4410	45.35	
APAN	4	8921	44.84	
TASQUILLO	2	4563	43.83	
AGUABLANCA DE ITURBIDE	1	2306	43.37	
MINERAL DE LA REFORMA	3	6998	42.87	
TOLCAYUCA	1	2345	42.64	
IXMIQUILPAN	8	19161	41.75	
TECOZAUTLA	3	7389	40.60	
LOLOTLA	1	2476	40.39	
YAHUALICA	2	5027	39.79	
MOLANGO	1	2549	39.23	
TLAXCOAPAN	2	5150	38.83	
ACATLAN	2	5179	38.62	
METZTITLAN	2	5229	38.25	
ZEMPOALA	2	5416	36.93	
TLANCHINOL	3	8245	36.39	
<b>ATOTONILCO DE TULA</b>	<b>2</b>	<b>5717</b>	<b>34.98</b>	<b>Ponderación 1 DE 36.39 A</b>
ACAXOCHITLAN	3	8734	34.35	
JACALA	1	3066	32.62	
TLAHUELILPAN	1	3213	31.12	
ORIZATLAN	3	9944	30.17	
NOPALA	1	3362	29.74	
SINGUILUCAN	1	3380	29.59	
SANTIAGO DE ANAYA	1	3430	29.15	
TEPATEPEC	2	7289	27.44	
EL ARENAL	1	3658	27.34	
XOCHIATIPAN	1	3939	25.39	
HUICHAPAN	2	9108	21.96	
SAN BARTOLA TUTOTEPEC	1	4724	21.17	
PROGRESO	1	4780	20.92	
HUEJUTLA DE REYES	5	25449	19.65	
CALNALI	0	4019	0	
EL OXOCHITLAN	0	638	0	
JUAREZ HIDALGO	0	706	0	
MINERAL DEL CHICO	0	1818	0	
PISA FLORES	0	4172	0	
TEZONTEPEC	0	2079	0	
<b>ZACUALTIPAN</b>	<b>0</b>	<b>5437</b>	<b>0</b>	<b>Ponderación 0 0 A 34.98</b>

ANEXO E

ESTRATIFICACIÓN FINAL DE LOS SSH PARA  
GRUPO DE EDAD DEL MENOR DE 5 AÑOS.

No.	MUNICIPIO	COLORES PARA EL MAPA
002	CAXOCHITLAN	ROJO
005	AJACUBA	ROJO
008	APAN	ROJO
013	ATOTONILCO DE TULA	ROJO
018	CHAPULHUACAN	ROJO
022	EPAZOYUCAN	ROJO
027	HUEHUETLA	ROJO
029	HUICHAPAN	ROJO
030	IXMIQUILPAN	ROJO
033	JUAREZ HIDALGO	ROJO
034	LOLOTLA	ROJO
037	METZTITLAN	ROJO
039	MINERAL DEL MONTE	ROJO
041	MIXQUIAHUALA	ROJO
043	NICOLAS FLORES	ROJO
047	PACULA	ROJO
048	PACHUCA DE SOTO	ROJO
060	TENANGO DE DORIA	ROJO
061	TEPEAPULCO	ROJO
063	TEPEJI DE OCAMPO	ROJO
069	TIZAYUCA	ROJO
071	TLAHUILTEPA	ROJO
074	TLAXCOAPAN	ROJO
076	TULA DE ALLENDE	ROJO
077	TULANCINGO DE BRAVO	ROJO
081	ZACUALTIPAN DE ANGELES	ROJO
084	ZIMAPAN	ROJO
003	ACTOPAN	AZUL
004	AGUA BLANCA DE ITURBIDE	AZUL
006	ALFAJAYUCAN	AZUL
012	ATOTONILCO EL GRANDE	AZUL
015	CARDONAL	AZUL
016	CUAUTEPEC DE HINOJOSA	AZUL
020	ELOXOCHITLAN	AZUL
024	HUASCA DE OCAMPO	AZUL
026	HUAZALINGO	AZUL
035	METEPEC	AZUL
040	LA MISION	AZUL
044	NOPALA DE VILLAGRAN	AZUL
049	PICAFLORES	AZUL
053	SAN BARTOLA TUTOTEPEC	AZUL
056	SANTIAGO TULANTEPEC	AZUL

062	TEPEHUACAN DE GUERRERO	AZUL
064	TEPETITLAN	AZUL
067	TEZONTEPEC DE ALDAMA	AZUL
082	ZAPOTLAN DE JUAREZ	AZUL
007	ALMOLOYA	VERDE
009	EL ARENAL	VERDE
010	ATITALAQUIA	VERDE
014	CALNALI	VERDE
017	CHAPANTONGO	VERDE
019	CHILCUAUTLA	VERDE
031	JACALA DE LEZAMA	VERDE
032	JALTOCAN	VERDE
036	SAN AGUSTIN MEZQUITITLAN	VERDE
038	MINERAL DEL CHICO	VERDE
042	MOLANGO DE ESCAMILLA	VERDE
045	OMITLAN DE JUAREZ	VERDE
054	SAN SALVADOR	VERDE
059	TOCOZAUTLA	VERDE
072	TLANALAPA	VERDE
073	TLANCHINOL	VERDE
075	TOLCAYUCA	VERDE
078	XOCHIATIPAN	VERDE
079	ACATLAN	AMARILLO
001	ATLAPEXCO	AMARILLO
011	EMILIANO ZAPATA	AMARILLO
023	FRANCISCO I. MADERO	AMARILLO
025	HUAUTLA	AMARILLO
028	HUEJUTLA DE REYES	AMARILLO
046	SAN FELIPE DE ORIZATLAN	AMARILLO
050	PROGRESO	AMARILLO
051	MINERAL DE LA REFORMA	AMARILLO
052	SAN AGUSTIN TLAXIACA	AMARILLO
055	SANTIAGO DE ANAYA	AMARILLO
057	SINGUILUCAN	AMARILLO
058	TASQUILLO	AMARILLO
065	TETEPANGO	AMARILLO
066	VILLA DE TEZONTEPEC	AMARILLO
068	TIALGUISTENGO	AMARILLO
070	TLAHUELILPAN	AMARILLO
080	YAHUALICA	AMARILLO
083	ZEMPOALA	AMARILLO

## ANEXO F

Tabla C. Procedimiento de ponderación de los indicadores para el grupo de edad del adolescente de los SSH.

No.	MUNICIPIO	MORT. GRAL	POND	MORBIL GRAL	POND	MORT POR ACCID	POND	MARGIN ALIDAD POND	SUMA POND	POND FINAL	RIESGO
1	ACATLAN	38.62	2	1405.5	1	0.00	1	3	7	1	BAJO
2	ACAXOCHITLAN	34.35	1	1236.3	1	17.16	3	3	8	1	BAJO
3	ACTOPAN	82.58	4	2291.7	1	22.73	3	1	9	2	MEDIO
4	AGUA BLANCA DE ITUR	43.37	2	2278.6	1	21.62	3	3	9	2	MEDIO
5	AJACUBA	147.80	4	2512.2	2	14.65	2	1	9	2	MEDIO
6	ALFAJAYUCAN	98.38	4	2864.7	2	61.24	4	3	13	4	MUY ALTO
7	ALMOLOYA	78.00	4	1884.3	1	38.35	4	3	12	3	ALTO
8	APAN	44.84	2	4194.5	3	22.24	3	1	9	2	MEDIO
9	EL ARENAL	27.34	1	1900.6	1	0.00	1	2	5	1	BAJO
10	ATITALAQUIA	81.90	4	1614.8	1	20.13	3	1	9	2	MEDIO
11	ATLAPEXCO	45.35	2	4940.9	4	0.00	1	4	11	3	ALTO
12	ATOTONILCO EL GRAN	88.60	4	2408.5	2	29.38	4	2	12	3	ALTO
13	ATOTONILCO DE TULA	34.98	1	4927.7	4	8.27	2	1	8	1	BAJO
14	CALNALI	0.00	1	3197.5	3	0.00	1	3	8	1	BAJO
15	CARDONAL	77.15	4	4040.2	3	37.94	4	3	14	4	MUY ALTO
16	CUAUTEPEC DE HINOJO	52.40	3	1986.2	1	0.00	1	2	7	1	BAJO
17	CHAPATONGO	67.96	3	2268.9	1	0.00	1	3	8	1	BAJO
18	CHAPULHUACAN	111.42	4	3407.8	3	27.26	4	3	14	4	MUY ALTO
19	CHILCUAUTLA	100.78	4	3178.7	3	24.91	3	3	13	4	MUY ALTO
20	ELOXOCHITLAN	0.00	1	9357.3	4	0.00	1	3	9	2	MEDIO
21	EMILIANO ZAPATA	101.87	4	1634.4	1	16.88	3	1	9	2	MEDIO
22	EPAZOYUCAN	81.37	4	3107.5	3	0.00	1	2	10	3	ALTO
23	TEPATEPEC	27.44	1	1675.4	1	13.18	2	1	5	1	BAJO
24	HUASCA DE OCAMPO	75.19	3	1753.2	1	11.92	2	3	9	2	MEDIO
25	HUAUTLA	62.43	3	3467.4	3	15.49	3	4	13	4	MUY ALTO
26	HUAZALINGO	114.68	4	2556.4	2	19.02	3	4	13	4	MUY ALTO
27	HUEHUETLA	50.83	3	1747.6	1	33.80	4	4	12	3	ALTO

## **GLOSARIO DE TÉRMINOS**

**Ambiasis:** Infección por Amibas

**Ca.Cu.:** Cáncer Cervicouterino

**EDAS:** Enfermedades Diarreicas Agudas

**EPV:** Enfermedades Prevenibles por Vacunación

**IRAS:** Infecciones Respiratorias Agudas

**ITS:** Infecciones de Transmisión Sexual

**TX:** Abreviatura de tratamiento

**Morbilidad:** La morbilidad es un indicador de daños a la salud de la población (No. De casos sobre población por  $k$  habitantes) que junto con otros indicadores como la mortalidad (No. de defunciones sobre población por  $k$  habitantes), sociodemográficos y de servicios de salud, describen la situación de salud y riesgo de enfermedad de la población.