

UNIVERSIDAD CENTRAL “MARTA ABREU” DE LAS VILLAS
FACULTAD DE MATEMÁTICA, FÍSICA Y COMPUTACIÓN
DEPARTAMENTO DE COMPUTACIÓN



“Uso de los métodos Scan en la detección de conglomerados de enfermos en Cifuentes”

Tesis presentada en opción al grado académico de Máster en Computación Aplicada

Autor: Dr. Jorge E. Díaz Chalala

Tutores: M.Sc. Laureano Rodríguez Corvea
Dra. Gladys Casas Cardoso

Santa Clara
2010

SÍNTESIS

El perfeccionamiento de los sistemas de Vigilancia en Salud, ha hecho imprescindible la vinculación de la epidemiología con otras ciencias, los aportes de la estadística y la computación han sido de gran utilidad. Se han desarrollado novedosos métodos de detección de epidemias que pueden ser utilizados en estudios de enfermedades transmisibles y no transmisibles.

Los métodos Scan son muy útiles en la detección de aglomeraciones temporales de enfermos, sin embargo se ha demostrado la superioridad de las variantes borrosas de este método con relación a las clásicas.

En este trabajo se propone una metodología que ayuda a determinar la presencia de conglomerados de enfermos y a caracterizar las enfermedades en estudio basada en la aplicación de los métodos Scan Clásicos y Borrosos, los cuales fueron implementados en un software realizado sobre el paquete Mathematica.

Utilizando el software elaborado se realizó el procesamiento a varias enfermedades y trastornos de la conducta presentes en el municipio de Cifuentes que eran de interés para los epidemiólogos.

ÍNDICE

INTRODUCCIÓN.....	1
CAPÍTULO I. TÉCNICAS DE DETECCIÓN DE CONGLOMERADOS DE ENFERMOS. SU UTILIZACIÓN EN EPIDEMIOLOGÍA.....	5
1.1 EL SISTEMA DE VIGILANCIA EN SALUD EN CUBA	5
1.2 TÉCNICAS DE DETECCIÓN DE CONGLOMERADOS	8
1.2.1 Métodos para investigar conglomerados espaciales	11
1.2.2 Métodos para investigar conglomerados temporales	12
1.2.3 Métodos de detección de conglomerados espacio - temporales.....	12
1.2.4 Otros métodos de detección de conglomerados	13
1.3 ACERCA DE LA METODOLOGÍA GENERAL	14
1.4 LA INTEGRACIÓN DE LOS FACTORES DE RIESGO A LOS MÉTODOS DE DETECCIÓN DE CONGLOMERADOS	16
1.5 INTRODUCCIÓN A LA LÓGICA BORROSA	18
1.5.1 Funciones de pertenencia	21
1.5.2 Borrosificador.....	23
1.5.3 Desborrosificador.....	24
1.6 PAQUETES ESTADÍSTICOS PARA LA DETECCIÓN DE CONGLOMERADOS	25
1.6.1 Paquetes generales	25
1.6.2 Sistema estadístico para la detección de conglomerados: EpiDet.....	26
1.7 CONSIDERACIONES FINALES	27
CAPÍTULO II. LOS MÉTODOS SCAN.....	28
2.1 LOS MÉTODOS SCAN CLÁSICOS	28
2.1.1 Método Scan Clásico en una línea para la detección de conglomerados temporales	28
2.1.2 El método Scan sobre un círculo	31
2.1.3 Consideraciones sobre los métodos Scan	31
2.1.4 Limitaciones del método Scan.....	32
2.1.5 El Scan Lineal sobre el Mathematica	32
2.2 LOS MÉTODOS SCAN BORROSOS	33
2.2.1 El método Scan Borroso sobre una Línea.....	34
2.2.2 El método Scan Borroso sobre un círculo	38
2.2.3 Algunas consideraciones sobre los métodos Scan Borrosos.....	39
2.2.4 El Scan Lineal Borroso sobre el Mathematica	39
2.3 METODOLOGÍA PARA LA APLICACIÓN DE LOS MÉTODOS SCAN EN LA DETECCIÓN DE CONGLOMERADOS DE ENFERMOS	45
2.4 CONSIDERACIONES FINALES	47
CAPÍTULO III. APLICACIONES.....	48
3.1 RECOPIACIÓN DE LOS DATOS	48
3.1.1 Procedimiento de recolección de datos sobre el Mathematica.....	49
3.2 RESULTADOS NO SIGNIFICATIVOS.....	50
3.2.1 Enfermedades Cerebrovasculares	50
3.2.2 Accidentes	53
3.2.3 Suicidios.....	54
3.3 RESULTADOS SIGNIFICATIVOS PARA TODOS LOS PARÁMETROS	56
3.3.1 Meningoencefalitis Viral.....	56
3.3.2 Hepatitis A	59
3.4 RESULTADOS SIGNIFICATIVOS PARA ALGUNOS PARÁMETROS	61
3.4.1 Enfermedades del Corazón	61
3.4.2 Tumores Malignos	64
3.4.3 Intentos Suicidas	67
3.5 CONSIDERACIONES FINALES	70

CONCLUSIONES.....	72
RECOMENDACIONES	73

INTRODUCCIÓN

Con la mejora de las condiciones de saneamiento, la evolución de los sistemas y servicios de salud, el descubrimiento de nuevos y más eficaces agentes terapéuticos y vacunas, así como de los métodos de control de enfermedades, la capacidad de expansión de las epidemias de enfermedades transmisibles de mayor prevalencia ha disminuido con el tiempo. En contraste, asistimos al desarrollo de enfermedades degenerativas, a los daños derivados de la contaminación y agentes tóxicos presentes en el ambiente, y a la emergencia de nuevas enfermedades transmisibles, (Coutin Marie 2008).

Una de las necesidades que plantea el comportamiento de estas entidades es el perfeccionamiento de los sistemas de vigilancia, de manera que puedan identificar cuando una agregación de casos de enfermos observada en un área geográfica determinada, en un período de tiempo limitado, o teniendo en cuenta ambos escenarios a la vez, es superior a lo esperado, y si ello representa un brote, (Consellería 2002).

En Cuba se ha redimensionado y desarrollado un modelo de vigilancia en salud que responda a las necesidades del sistema sanitario, de acuerdo con los cambios del patrón epidemiológico y las transformaciones operadas en las condiciones de salud de la población. La vigilancia en salud se ha ido desarrollando en el país y sus resultados tienen un impacto decisivo en la mejoría de gestión y de los indicadores de salud a todos los niveles.

Las técnicas estadísticas pueden ayudar a los epidemiólogos a reconocer si una agrupación de casos fuera de lo usual es estadísticamente significativa, y puede corresponder al inicio de una epidemia, (Casas Cardoso 2003).

La modelación matemática o estadística de las epidemias no es un tema novedoso en nuestros tiempos, sin embargo la mayoría de las investigaciones en este campo se limitaron al estudio de enfermedades transmisibles. La epidemiología en Cuba presta también atención a las enfermedades no transmisibles, pero existe la limitación de que esta teoría de modelación está menos desarrollada. Las técnicas de *clustering* juegan aquí un papel fundamental.

Las enfermedades no transmisibles (ENT) constituyen las primeras causas de muerte en Cuba desde hace varios años y se conoce además que entre ellas se encuentran las primeras causas de mortalidad prematura (accidentes, suicidio, tumores malignos, etc.), (Álvarez Sintés 2008).

La epidemiología es la ciencia que se ocupa del estudio de las epidemias. Para dar respuestas a todas las preguntas relacionadas con el origen y la evolución de las epidemias, se ve en la necesidad de auxiliarse de otras ciencias entre las que juegan un papel fundamental la matemática y la computación. Dentro de la matemática es la estadística la que desempeña esta labor con una influencia cada vez mayor. Una de las tareas más importantes de los epidemiólogos es detectar a tiempo pequeños focos de enfermedades para evitar que se conviertan en epidemias de serias complejidades y de difícil control.

Existen numerosos métodos estadísticos reportados en la literatura que resuelven este problema. Las técnicas Scan son unas de las más utilizadas.

El test Scan, es una prueba estadística para detectar agrupaciones en una serie temporal de casos de un cierto evento o enfermedad. La hipótesis nula establece que los casos ocurren al azar en el tiempo, frente a la alternativa de que los casos se agrupan en ciertos períodos. Es un método útil para detectar “picos” en la incidencia de una enfermedad y, al no tener un componente espacial, es adecuado para estudios de enfermedades en los que la población se distribuye de forma desigual en el área de estudio, (Rodríguez 2007).

En este contexto se desarrolla el presente trabajo. Su objetivo general es aplicar las técnicas Scan para la detección de conglomerados a los datos de enfermedades recogidos en el municipio de Cifuentes para determinar la existencia de aglomeraciones inusitadas de casos. Los objetivos específicos son:

1. Estudiar los registros de enfermedades presentes en el municipio de Cifuentes para seleccionar las más importantes desde el punto de vista epidemiológico.
2. Proponer una metodología general para la aplicación de los métodos Scan.

3. Realizar el procesamiento a las enfermedades seleccionadas acorde a la metodología propuesta y llegar a conclusiones estadísticas y epidemiológicas que expliquen el comportamiento de tales enfermedades.
4. Realizar el procesamiento con todos los casos y según el sexo del paciente.

Para dar cumplimiento a estos objetivos fue necesario plantearse y solucionar algunas tareas de investigación, entre las que se encuentran:

1. Estudiar elementos de estadística y de métodos de detección de conglomerados en investigaciones epidemiológicas (para cumplimentar los objetivos específicos 1 y 2)
2. Implementar los métodos Scan en un *software* de propósito general como el Mathematica para poder utilizarlo en la detección de conglomerados (para cumplimentar los objetivos 3 y 4)

El primer paso para la realización de este trabajo fue la confección del marco teórico. Para ello se realizó una amplia revisión de la literatura consultando libros, artículos y páginas de internet, entre otras fuentes. Sus elementos esenciales se encuentran expuestos de manera resumida en el primer capítulo de la presente tesis. Como conclusión de la elaboración del marco teórico se enuncia la siguiente hipótesis de investigación:

“La formalización de una metodología basada en la aplicación de los métodos Scan (Clásico y Borroso) ayuda a determinar la presencia de conglomerados y a caracterizar las enfermedades en estudio”

El trabajo está formado por tres capítulos: el primero constituye una revisión bibliográfica sobre el Sistema de Vigilancia Epidemiológica en Cuba y la utilidad de los métodos de detección de conglomerados en el estudio de posibles epidemias. Se hace una breve descripción de la metodología general para investigar conglomerados y también se exponen algunos elementos de lógica borrosa.

En el capítulo 2 se hace énfasis en el método Scan en sus variantes Clásica y Borrosa para la detección de conglomerados temporales sobre un círculo y sobre una línea. Se presenta de manera breve el *software* elaborado sobre el paquete Mathematica. El capítulo culmina con la formulación de una metodología para la aplicación de estas técnicas.

En el tercer capítulo se describen brevemente las afecciones y trastornos de la conducta estudiados y se muestran los resultados obtenidos al aplicar la metodología anteriormente propuesta. Además se hace un análisis sencillo, de acuerdo al criterio de especialistas en Higiene y Epidemiología, de las posibles causas que determinaron la existencia de conglomerados de enfermos en algunas enfermedades en los años estudiados.

Finalmente se presentan las conclusiones de la tesis, así como algunas recomendaciones que abren futuras líneas de investigación.

CAPÍTULO I. TÉCNICAS DE DETECCIÓN DE CONGLOMERADOS DE ENFERMOS. SU UTILIZACIÓN EN EPIDEMIOLOGÍA

En este capítulo se hace una revisión bibliográfica sobre el Sistema de Vigilancia Epidemiológica en Cuba y la utilidad de los métodos de detección de conglomerados en el estudio de posibles epidemias. Se hace una breve descripción de la metodología general para investigar conglomerados y también se exponen algunos elementos de lógica borrosa.

1.1 El sistema de vigilancia en salud en Cuba

Las condiciones económico-sociales que han caracterizado la situación del país, con las consecuentes limitaciones en los recursos materiales y financieros, los cambios ocurridos en los perfiles de riesgo, morbilidad y mortalidad en los últimos años, con mayor relevancia hacia las enfermedades no transmisibles, el deterioro ambiental y sus implicaciones en la salud de la población, especialmente en las zonas urbanas, y la emergencia de nuevos problemas, han puesto de manifiesto la necesidad de perfeccionar los sistemas de vigilancia epidemiológica a cada nivel de la organización de salud, los sistemas de detección activa y la capacidad de respuesta, de modo que permita identificar cambios a corto plazo y ejecutar las alternativas de intervención más rápidas, oportunas y eficientes, (Feal Cañizares 1999).

La vigilancia en salud es esencial en los procesos de prevención de enfermedades y promoción de salud, es una herramienta vital en la ubicación de recursos del sistema de salud, así como la evaluación de la eficiencia de programas y servicios, (Berdasquera Corcho 2000). Ha sido definida en nuestro contexto como el seguimiento, recolección sistemática, análisis e interpretación de datos sobre eventos de salud o condiciones relacionadas, para ser utilizadas en la planificación, implementación y evaluación de programas de Salud pública, incluyendo como elementos básicos la diseminación de dicha información a los que necesitan conocerla para lograr una acción de prevención y control más efectiva y dinámica en los diferentes niveles, (Guerrero Pupo 2008).

A partir del primer semestre de 1993 se definió un grupo de estrategias por el Área de Higiene y Epidemiología del Ministerio de Salud Pública (MINSAP), con el objetivo central de dar una respuesta mucho más eficiente a la situación sanitaria del país, una de cuyas prioridades fue dirigida al establecimiento de un sistema de vigilancia en salud que permitiera "una mayor integración de la información de la vigilancia y un mayor nivel de análisis y utilización de la información que ella genera, en aras de lograr un mayor beneficio para la toma de decisiones, con base científica a los diferentes niveles, acorde con la política de descentralización de las instancias de gobierno", (Rodríguez Milord 2001; Guerrero Pupo 2008). En este contexto se crearon las Unidades de Análisis y Tendencias en Salud (UATS), con una visión epidemiológica integral e intersectorial que partiendo del análisis y evaluación permanente de la Situación de Salud existente y prospectiva, actúa como instancia integradora de la información que la vigilancia genera en las diferentes áreas y niveles, (Rodríguez Milord 2001; Rodríguez Milord 2002).

En consonancia con todo lo antes expuesto, el sistema de vigilancia en salud se ha venido implementando en función de las realidades y necesidades de la Salud Pública actual, y se ha estructurado fundamentalmente sobre la base un componente de alerta - acción, un componente estratégico, un componente evaluativo y un grupo de información científica y de investigaciones, (Coutin Marie 2008):

a) **Componente Táctico:** A través de la implementación del SISTEMA ALERTA - ACCIÓN, se ofrece información a corto plazo sobre las principales incidencias que con carácter agudo y relevante ocurren en el orden de las enfermedades, daños o condiciones relacionadas, de los factores de riesgo y de los eventos en los servicios de salud. Ello permite el mantenimiento de un sistema sostenible y actualizado de información para la toma de decisiones oportunas en cada nivel del sistema de salud, a través de la emisión de informes sobre la situación de salud, con periodicidad diaria y semanal.

b) **Componente Estratégico:** Este componente integra la información obtenida de diferentes fuentes como las estadísticas de salud, los resultados de programas y de investigaciones, los estudios especiales, informes de los diferentes subsistemas de vigilancia, la información extrasectorial relacionada con la salud, los informes de países y de organismos internacionales, entre otras, para ser utilizada en la vigilancia a mediano y

a largo plazo. El mismo centra su actividad en la realización de análisis, síntesis e interpretación de la información sobre los perfiles de salud y enfermedad.

c) **Componente Evaluativo:** Se ha desarrollado un a través de un grupo de comisiones técnicas asesoras o equipos de expertos como instrumento dinamizador de las estrategias elaboradas por el Ministerio de Salud Pública en el abordaje de la problemática de salud, dirigidas a la reorientación del sistema hacia la atención primaria, la recuperación de la vitalidad de los hospitales, la consolidación del Programa Nacional de Medicamentos y el avance científico - técnico, además de otras prioridades.

La Vigilancia en Salud ha contribuido al fortalecimiento de las capacidades a nivel central y de las provincias para la utilización más efectiva de la información epidemiológica y del análisis, así como del seguimiento y control de los problemas vinculados a la salud. Con ello, el proceso de la toma de decisiones se ha visto beneficiado con una mayor oportunidad y eficiencia en la prevención y el control,(Batista Moliner 2001).

La importancia de los sistemas de vigilancia en salud radica en los principios siguientes,(Álvarez Sintet 2008):

- ✓ Permiten conocer la situación de los problemas de salud.
- ✓ Proveen de datos necesarios sobre diferentes factores complejos que intervienen en la ocurrencia de distintos problemas de salud.
- ✓ Facilitan la formulación de orientaciones sobre bases objetivas, con el fin de prevenir y controlar los problemas de salud.
- ✓ Permiten prevenir y actuar con más precisión y efectividad en las acciones frente a brotes epidémicos y epizooticos.
- ✓ Facilitan la toma de decisiones de los niveles ejecutivos de las instituciones de salud y del gobierno.

El sistema de vigilancia debe influir en la actividad de investigación, con propuestas permanentes de nuevas hipótesis, y ofrecer información útil para el establecimiento de las políticas y programas de salud, por lo que la interrelación INVESTIGACIÓN – VIGILANCIA - ACCIÓN constituye una tríada esencial para el propio desarrollo de la Salud Pública, (Guerrero Pupo 2008).

Una de las tareas mas importantes de la Vigilancia Epidemiológica es el análisis del comportamiento y tendencias de las enfermedades y/o daños sujetos a vigilancia y, dentro de ello, el análisis de los patrones epidémicos observados, que permitan la estimación numérica de parámetros de importancia médica, e incluso el tamaño de la población afectada. Las técnicas estadísticas pueden ayudar a los epidemiólogos a reconocer si la agrupación de casos fuera de lo usual es estadísticamente significativa, y puede corresponder al inicio de una epidemia, (Casas Cardoso 2003).

Para el análisis de la información a mediano y largo plazo se introducen metodologías y técnicas estadístico - matemáticas y epidemiológicas, con estructuración de bases de datos en cada nivel del sistema, descripción de variables en tiempo, lugar, personas y sus interacciones, análisis exploratorio de datos, métodos de análisis de series cronológicas con modelación matemática y sistemas gráficos de presentación de datos, entre otros. Estas técnicas y procedimientos están siendo difundidas para su empleo a los niveles locales y al resto del sistema de salud, (Rodríguez Milord 2002).

1.2 Técnicas de detección de conglomerados

Las modificaciones de los patrones epidemiológicos de numerosas enfermedades, así como la aparición de enfermedades emergentes y reemergentes han hecho imprescindible el perfeccionamiento de los sistemas de vigilancia y para ello las ciencias exactas han resultado de gran utilidad, (Toledo Curbelo 2007).

Desde el siglo pasado, la modelación matemática, o más específicamente la modelación estadística de las epidemias, ha venido enriqueciendo el conocimiento epidemiológico y ha sido de gran ayuda en la promoción efectiva de acciones de Salud Pública. Uno de los problemas al que se le presta especial atención en la actualidad es a la determinación de cuando una agregación de casos de enfermos en un área geográfica dada, en un período de tiempo limitado, o considerando ambos escenarios a la vez, es superior a lo esperado, (Casas Cardoso 1999) .

Los epidemiólogos tienen sus propios métodos de detección de epidemias, que de hecho han probado ser eficientes en numerosas ocasiones; les permiten detectar con cierta precisión la aparición de focos infecciosos, pero no son totalmente confiables y en

ocasiones conllevan a cometer errores. Los matemáticos están interesados en redefinir y hacer más precisos esos procedimientos mediante el uso de alguna prueba de significación, (Casas Cardoso 2002).

En la actualidad, el manejo de estas situaciones se ha comenzado a estudiar mediante las denominadas técnicas de *clustering* (del inglés *cluster* que significa aglomeración).

Un conglomerado se puede definir como la aparición de un número de casos de una enfermedad en una cantidad superior a la que se esperaría para un determinado grupo de población, un área geográfica y un período de tiempo determinado, (Lertxundi-Manterola 2005).

Para precisar esta definición es necesario dividirla en tres: se denomina conglomerado espacial o geográfico a un exceso de casos diagnosticados en un área geográfica que puede ir desde un pequeño poblado hasta todo un continente; un conglomerado temporal es un exceso de enfermos muy cercanos en el tiempo y un conglomerado espacio-temporal es un exceso de casos en ambos escenarios: espacio y tiempo. Este último difiere de las aglomeraciones simples en espacio y tiempo por separado, pues puede ocurrir que ellas sean independientes. La interacción describe el proceso en el cuál los pares de casos cercanos en tiempo lo están también desde el punto de vista espacial, por ejemplo en una enfermedad contagiosa, el proceso de infección requiere el contacto entre un individuo enfermo y uno o varios susceptibles, (Borracci 2005).

Esta metodología reúne a los individuos en conglomerados (*clusters*), de tal forma que los elementos de un mismo agrupamiento son más parecidos entre sí que a los elementos de otros conglomerados. Lo que se intenta de esta forma es maximizar la homogeneidad de los objetos dentro de los conjuntos mientras que a la vez se maximiza la heterogeneidad entre los agregados, (Borracci 2005).

En todas estas técnicas, la hipótesis nula o fundamental H_0 es la distribución aproximadamente uniforme de los casos diagnosticados o lo que es lo mismo, la no existencia de conglomerados. Pequeños valores de la probabilidad traen consigo el rechazo de H_0 y muestran evidencia de formación de aglomeraciones, (Casas Cardoso 2003).

Numerosas pruebas estadísticas para la detección de conglomerados se han reportado en la literatura en los últimos años. El pionero de toda esta teoría fue Knox, quién introdujo por primera vez la noción de interacción espacio-temporal. Knox partió de la definición de cierto criterio de adyacencia en espacio y en tiempo y mostró sus resultados finales en una tabla de contingencia 2 x 2. Según él, el número de pares de casos cercanos en espacio y tiempo sigue una distribución de Poisson. Los tests de Mantel y de Jacquez entre otros se utilizan también con el mismo fin aunque tienen una formulación matemática diferente. Entre las técnicas que detectan aglomeraciones inusitadas de enfermos en zonas geográficas se encuentran los tests de Cuzick y Edwards, y el test de Moran ajustado. Para detectar clusters en series de tiempo cortas se usan las pruebas de Larsen y la de Scan, por mencionar algunas. Existen también técnicas más generales que pueden emplearse en la detección de conglomerados en espacio, tiempo y en ambos: el método de Grimson constituye el mejor ejemplo en estos casos, (Casas Cardoso 2002).

Todos estos métodos tienen sus ventajas y limitaciones, los profesionales de la Salud Pública deben conocerlas con precisión a fin de no emitir conclusiones equivocadas.

No deben utilizarse solo métodos estadísticos para la detección de conglomerados de enfermos. Ante una posible agrupación no usual de casos, se deben evaluar la consistencia, la plausibilidad biológica y el efecto dosis respuesta, (Rodríguez Milord 2001).

La **consistencia**, en términos de las investigaciones de conglomerados, puede referirse a la observación de un patrón histórico en los casos notificados, o su compatibilidad con datos de la literatura; sin embargo, la consistencia se refiere generalmente a la homogeneidad de los casos. Los atributos personales, como edad, raza, sexo y ocupación, pueden reforzar un resultado estadístico sugestivo; además, ciertas características específicas de la enfermedad (como la estirpe celular en los tumores, las vías de exposición, la localización anatómica, entre otras) pueden ser importantes evidencias de consistencia. Dado que los eventos raros de salud se estudian mejor en conglomerados, la uniformidad dentro de la agregación aporta validez a las evidencias e interpretación estadística.

La **plausibilidad biológica** es siempre un aspecto importante en la interpretación epidemiológica de los datos. La presencia de un riesgo ambiental reconocido, o un factor de exposición importante, proporcionan plausibilidad biológica para una investigación; sin embargo, ésta no puede ser una base para descartar un conglomerado, aunque sí un criterio para reforzar y/o interpretar las pruebas estadísticas.

El **efecto dosis-respuesta** es uno de los criterios más fuertes para evaluar los datos epidemiológicos y en muchas ocasiones forma la base para detectar relaciones de causalidad.

1.2.1 Métodos para investigar conglomerados espaciales

Uno de los problemas que con mayor frecuencia se le presenta a los epidemiólogos es determinar si una enfermedad tiene una prevalencia mayor en una localidad o área específica que otras. Ante esta disyuntiva el investigador dispone de dos tipos diferentes de datos: **tasas** correspondientes a áreas de variado tamaño (barrios, por ejemplo) o **coordenadas de localización geográfica**, las cuales se obtienen, por ejemplo, a partir de la dirección de residencia de los casos.

En dependencia del tipo de dato disponible se explora cierto patrón espacial. Cuando se dispone de tasas se trata de determinar si las áreas con altas tasas forman *clusters* espaciales; si se dispone de coordenadas se busca determinar cuándo se agrupan localizaciones específicas de casos.

Existen muchos tests para la clusterización espacial. Uno de los más clásicos, conocido como **Test de Pearson**, trabaja con tasas sobre familias de una medida prefijada K y se ha generalizado para trabajar con "celdas", en vez de "familias", en las cuales la población de riesgo es homogénea. El método Pearson parte de dividir el área de estudio en celdas que tengan la misma población de riesgo. Determina la probabilidad de que las celdas espaciales sean consistentes con la hipótesis nula que plantea que los casos se distribuyen aleatoriamente entre ellas. Inicialmente daba una solución Chi-cuadrado aproximada; sin embargo el test estadístico razón de verosimilitud Chi-cuadrado se considera más apropiado para estos casos, (Casas Cardoso 1999).

1.2.2 Métodos para investigar conglomerados temporales

Otro de los problemas que se le pueden presentar al epidemiólogo es decidir si el número o proporción de casos de una enfermedad determinada que aparece en intervalos de tiempo consecutivos, suceden con una frecuencia diferente a la esperada debido a una distribución aleatoria, es decir, diferente a lo esperado si hubieran sido generadas por el azar. Las técnicas de *clustering* temporales pretenden dar una respuesta acertada, especialmente cuando los datos disponibles se refieren a series cortas de tiempo, que no pueden ser analizadas usando los métodos convencionales, (Casas Cardoso 2003) .

Aparecen en la literatura más de una docena de métodos que realizan ese tipo de análisis. Los más conocidos son el test de Grimson, la prueba de Scan, el test de Larsen, y los tests de las celdas vacías y de Dat, (Casas Cardoso 1999) .

En el caso temporal se han desarrollado técnicas que detectan conglomerados de enfermos en series de tiempo cortas, en las que no pueden aplicarse las técnicas tradicionales de modelación desarrolladas para series cronológicas. Estos tests trabajan mayormente con fechas de diagnóstico, de primeros síntomas o de muerte en aquellos casos en que se estudie la mortalidad. En ocasiones no se necesitan fechas propiamente dichas sino, por ejemplo, incidencia de casos diarios, semanales o mensuales. Al igual que en el espacio, existen métodos específicos que trabajan sobre unos u otros datos y que permiten determinar la presencia de una agrupación de casos superior a lo esperado, (Casas Cardoso 2003).

1.2.3 Métodos de detección de conglomerados espacio - temporales

El estudio de aglomeraciones espacio - temporales no es la presencia de clusters en espacio y tiempo a la vez. Ambos pueden existir por separado sin que aparezca la interacción. Ella supone que los casos cercanos en el espacio estén además cercanos en el tiempo y por ello este tipo de patrones con interacción es muy útil cuando se desean investigar enfermedades transmisibles, (Col 2000).

Los procesos espacio - temporales son relevantes en epidemiología debido a que incluyen aspectos como la infección o el contagio, las exposiciones ambientales que inducen

patrones espaciales de enfermedades, explican la distribución espacial de enfermedades transmitidas por vectores y otros procesos de difusión. A su vez, covariables como el nivel socio - económico y la etnicidad tienden a estar espacialmente estructurados y pueden inducir patrones espaciales a través de la asociación con resultados de salud, (Alegret Rodríguez 2008).

En el orden de la elaboración de pruebas para la significación de conglomerados hay que citar en primer lugar a Knox, (Knox 1959; Knox 1963), quién partió de examinar no exactamente los intervalos de tiempo entre dos eventos sucesivos, sino más bien de los intervalos de tiempo entre todas las parejas posibles de eventos. En los trabajos referidos estas ideas fueron aplicadas no sólo a distribuciones temporales, sino también a distribuciones espaciales donde la noción de eventos sucesivos no tiene sentido, pero sí toda pareja de eventos geográficos. La extensión de tales ideas al caso de la interacción espacio - tiempo constituyen un hito en toda esta teoría, (Casas Cardoso 2003).

1.2.4 Otros métodos de detección de conglomerados

Además de estos tres patrones, comienza a tener importancia otro que consiste en el análisis simultáneo de varias series temporales. Se trata en este caso de estudiar diversas series de tiempo “independientes” (correspondientes por ejemplo a varias localidades) y conjugarlas para determinar hasta que punto, estos datos completos (y no cada serie por separada), sugieren la formación de conglomerados en el tiempo, de acuerdo por ejemplo a cierto mecanismo de transmisión prepostulado. Con tal fin pueden utilizarse los métodos de Grimson y de Dat adecuados convenientemente.

En los últimos tiempos se ha prestado también especial atención al desarrollo de métodos efectivos para detectar aglomeraciones partiendo de localizaciones dudosas. Se han publicado varios trabajos sobre las estadísticas de detección de tales aglomeraciones para datos espacio - temporales imprecisos.

Los llamados métodos de detección de aglomerados focalizados pretenden investigar tasas inusitadas cerca de un foco, supuesta fuente de contaminación. Recientemente se realizó un estudio muy interesante acerca del daño al ADN que pueden sufrir las personas de acuerdo a su ocupación, (Casas Cardoso 2003).

El posible efecto de una intervención sobre conglomerados se estudia con ayuda de técnicas robustas de regresión que pretenden en definitiva no detectar inicialmente grupos, sino más bien estimar si ellos se alteran o desintegran producto de dicha interacción.

Las técnicas de exploración multivariada (análisis de *clusters* y escalamiento multidimensional) son métodos que permiten una reducción de dimensiones que simplifica complejos patrones de asociaciones entre muchas entidades. En el caso del análisis de *cluster* permite crear una representación bidimensional (dendrograma) de *clusters* de objetos diferentes cuyas relaciones se representan a través de matrices. El escalamiento multidimensional se usa para identificar las dimensiones que mejor explican las similitudes y diferencias entre las variables, (Sánchez Tarragó 2007).

1.3 Acerca de la metodología general

El Centro para Control de Enfermedades (CDC) formuló una guía o metodología general para investigar conglomerados de enfermedades, (Jacquez 1996). La misma recomienda un protocolo para manipular investigaciones de detección de aglomeraciones procediendo con cuatro pasos, desde un primer contacto hasta el diseño e implementación de una investigación etiológica. La metodología ha sido internacionalmente validada. A continuación aparece su esquema general:

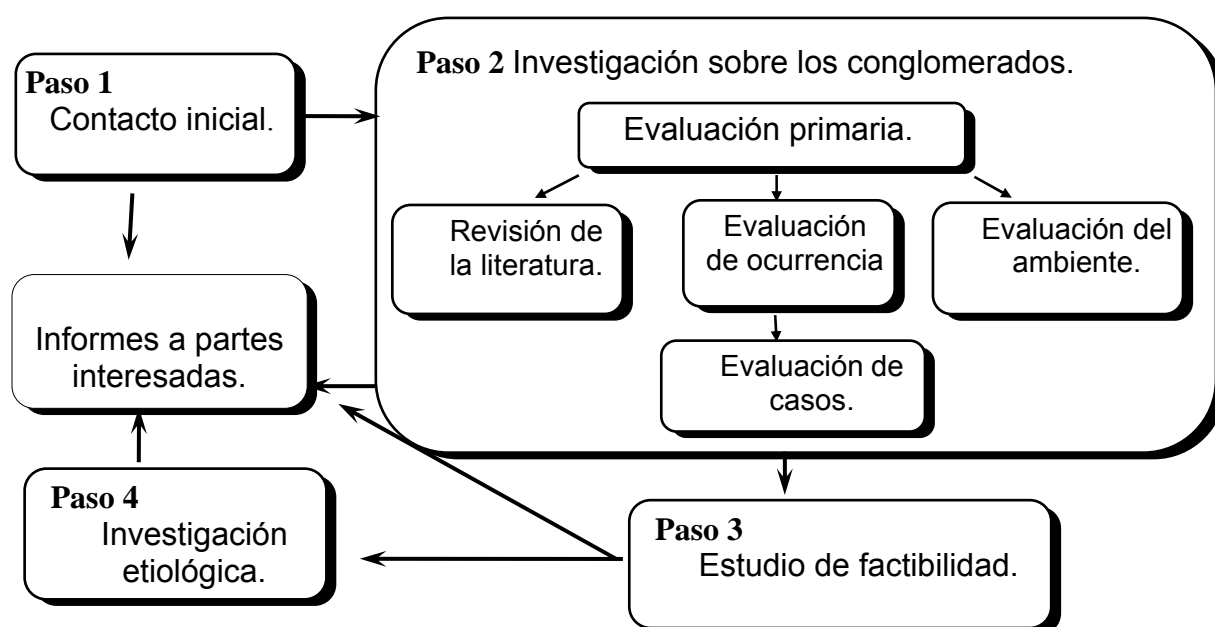


Figura 1.1 Metodología general para investigar conglomerados de enfermedades.

Esta guía plantea que las agencias locales de salud establecen un primer contacto con cualquier persona o institución que reporte un posible conglomerado de enfermos y lo primero que se hace es obtener información describiendo al informante (o a la información en sí). El propósito de este paso es determinar si se requiere una evaluación más profunda, (CDC. 1990), si se necesita tal investigación, se procede al análisis de conglomerados que debe proporcionar un estimado riguroso de cuán atípica es realmente la aglomeración de casos. Durante la evaluación primaria se define un área geográfica y un período de tiempo determinado y se asume que los casos reportados son reales. Nótese que la metodología propone esto último porque la evaluación detallada de casos sólo se emprende si se encuentra que el patrón representado de casos diagnosticados no es típico. En este mismo paso dos puede ser necesaria la revisión de la literatura, así como alguna investigación ambiental si se sospecha por ejemplo que los casos han estado expuestos a un factor de este tipo. Hecha la evaluación de la ocurrencia, se decide entonces hacer la evaluación de casos detallada y en esta etapa, como en la primera y en todas las restantes, se brinda información a las personas o instituciones autorizadas.

Si el número de casos reportados es suficiente, se estima la tasa que se compara con la de la población de referencia (digamos a través de índices estandarizados de mortalidad o morbilidad), para determinar cuán grave o marcado es el conglomerado, como un dato adicional a la comprobación de su carácter significativo. Si no hay disponibles tasas históricas, o no pueden calcularse las actuales, se trabaja con la cantidad física de casos.

Desde el punto de vista de las pruebas estadísticas, el tipo de dato disponible a veces determina los métodos que pueden usarse (espaciales, temporales, espacio - temporales), pero cada vez que sea posible deben ensayarse todos los que puedan resultar adecuados.

En la evaluación de casos se hace una verificación del diagnóstico inicial si la investigación preliminar ha evidenciado un exceso de casos y si los datos evidencian un problema de cierta importancia en la Salud Pública.

El estudio de mayor factibilidad se emprende en el paso tres sólo después que se han conformado los casos y los datos sugieren algún tipo de mecanismo biológico. El propósito de esta etapa es determinar si se necesita una investigación plena etiológica, en caso

afirmativo se diseña el plan de investigación detallado. Los procedimientos del estudio de factibilidad suponen:

- ✓ Conducir una revisión bibliográfica detallada para acceder al conocimiento de las posibles causas de la aglomeración.
- ✓ Generar un diseño de estudio apropiado y evaluar su costo y resultados esperados.
- ✓ Determinar los datos, la logística, y los métodos estadísticos y epidemiológicos requeridos para implementar el diseño del estudio.
- ✓ Asegurar el clima social y político y predecir el impacto de los posibles resultados del estudio.
- ✓ Evaluar el impacto sobre los recursos de salud pública, en caso de emprenderse el estudio.

Si un estudio de factibilidad demostrara que poco puede ganarse continuando la investigación, los resultados hasta ese momento se resumen y envían a las partes interesadas. En caso contrario se procede a la investigación etiológica.

Durante esta última se implanta el estudio concebido en la etapa anterior con el objetivo sobre todo de investigar la etiología de una supuesta relación exposición - enfermedad. Dicha investigación no necesariamente explica las causas subyacentes del conglomerado original (aun cuando puede estar entre sus objetivos); más bien ella debe evaluar posible hipótesis epidemiológicas generadas como explicación a la formación de conglomerados, para evitar su expansión o repetición.

1.4 La integración de los factores de riesgo a los métodos de detección de conglomerados

La presencia de factores de riesgo está relacionada con elementos individuales o ambientales que influyen en su aparición y una vez que están presentes actúan durante largo tiempo para desencadenar el efecto. Por tanto, su modificación (control o eliminación) requiere de acciones que deben aplicarse de forma sistemática para observar cambios en los parámetros o indicadores de su comportamiento, (Batista Moliner 2001).

Todos los seres vivos están expuestos constantemente a múltiples y diversos riesgos de enfermar y de morir. El hombre, que vive en un ambiente sociocultural artificial, es decir, creado y desarrollado históricamente por él mismo, tiene por razones ecológicas y sociales una diversidad grande de riesgos y una oportunidad también mayor de enfrentarse a ellos.

La noción de riesgo epidemiológico es función de la existencia del ser humano viviendo en un ambiente social en variación permanente, vale decir, los riesgos que afectan nuestra salud, dependen de la vida en sociedad, y más precisamente, del tipo estructural de la sociedad en que vivimos, (Duque de Estrada Riverón 2001) .

El concepto epidemiológico de riesgo implica la probabilidad que tiene un individuo (riesgo individual) de desarrollar una enfermedad determinada, un accidente o un cambio en su estado de salud en un período específico y en una comunidad dada, con la condición de que no enferme o muera de otra causa en ese período, .

Los factores de riesgo (o criterios de riesgo) son aquellas características y atributos (variables) que se presentan asociados diversamente con la enfermedad o el evento estudiado; ellos no son necesariamente las causas (o la etiología necesaria), solo sucede que están asociadas con el evento. Como constituyen una probabilidad medible, tienen valor predictivo y pueden usarse con ventajas tanto en prevención individual como en los grupos y en la comunidad total, (Narey B 1990; Rodríguez Calzadilla 1997).

Concretamente, entendemos por criterio o factor de riesgo toda característica y circunstancia determinada ligada a una persona, a un grupo de personas o a una población, la cual sabemos que está asociada con un riesgo de enfermedad, la posibilidad de evolución de un proceso mórbido o de la exposición especial a tal proceso, (Goldberg 2008).

Las técnicas de detección de *clusters* por sí solas se limitan a detectar el conglomerado en cuestión, no lo caracterizan. Un profesional de la Salud Pública necesitaría, además de detectar a tiempo un posible foco epidémico, conocer cuál o cuáles causas propiciaron en mayor medida la afección, (Casas Cardoso 2002).

La propia concepción general de los métodos de detección de aglomeraciones como técnicas que deben estar integradas en general a otros métodos epidemiológicos y no como técnicas aisladas y autosuficientes para la detección de focos epidémicos, la necesidad de que dichas técnicas no sólo deben ser capaces de ayudar en la detección de una congregación, sino también en el esclarecimiento de su naturaleza, el interés de fortalecer los métodos de distinción de conglomerados verdaderos de los falsos que tan frecuentemente aparecen en la práctica y concretamente validar consistencia, posibilidad biológica y evidencia de dosis - respuesta ponen de manifiesto la importancia de desarrollar una teoría matemática que facilite integrar el estudio de factores de riesgo a la detección de conglomerados, (Casas Cardoso 2003).

Existe un procedimiento matemático para adicionar factores sucesivos a los métodos clásicos de detección de *clusters*, (Casas Cardoso 2002); mediante la observación de la evolución de la significación del test aplicado pueden detectarse los principales riesgos que están afectando a la aglomeración y esto sirve de gran ayuda a los profesionales de la Salud Pública en la determinación de las causas que lo provocan.

Desde el punto de vista matemático se requiere dicotomizar los factores de riesgo, así $F = 1$ representa la presencia del factor: hábito de fumar por ejemplo, y $F = 0$ representa su ausencia. El número de individuos a los que se aplica el test se reduce a aquellos en los que el factor está presente ($F = 1$). Así, a medida que se incrementa el número de factores deberá disminuir la cantidad de individuos sobre los que se aplica el método. Si los factores seleccionados caracterizan bien a la enfermedad, este número no disminuirá demasiado pues debe existir coincidencia entre los pacientes enfermos y los pacientes con los factores. Si esto sucede, a medida en que se incorpore un nuevo factor al análisis, el valor de la probabilidad empírica de significación debe ser menor que su valor sin considerar al factor, ver (Casas Cardoso 2002).

1.5 Introducción a la lógica borrosa

Dos de los aspectos que contaminan normalmente la información en cualquier área del saber, son la **imprecisión** que tiene en su expresión y la **incertidumbre** que puede provocar la fuente que nos la proporciona. Ciertas personas tienen suficiente habilidad para

tomar decisiones correctas a partir de un conjunto de datos que viene expresados de forma vaga o imprecisa (borrosos) casi siempre utilizando adjetivos o adverbios como mucho, poco, alto, bajo, normal, muy, entre otros. Tales personas pueden controlar eficientemente un proceso tecnológico (en un Central azucarero el tradicional puntista que controla el proceso de cristalización del azúcar), diagnosticar enfermedades o una enfermedad a partir de síndromes y síntomas (el médico clínico) o tomar una decisión acertada en una determinada empresa e institución. El ser humano se desenvuelve con extraordinaria facilidad a la hora de manejar este tipo de información, sin embargo, cuesta explicar que procedimientos sigue para ello, (Calviño 2003).

Para hacerle frente a la información imprecisa han surgido diferentes teorías matemáticas: teorías como la de la probabilidad (Feller 1971), como la de la evidencia (Yager 2008), o como la de los Factores de Certeza (Shortliffe 1975). Estas teorías han despertado un creciente interés en la investigación científica. La herramienta por excelencia para modelar fenómenos en los que rige el principio de simultaneidad gradual es la Teoría de los Subconjuntos Borrosos, cuya base son las lógicas multivalentes desarrolladas en las primeras tres décadas del siglo XX (Lukasiewicz 1910). El concepto de conjunto borroso (que caracteriza de manera apropiada la imprecisión en la información) fue introducido en la década del 60 por Lofti A. Zadeh (Zadeh 1968), quien se considera el padre de la lógica borrosa.

En términos más rigurosos, la **teoría de conjuntos borrosos** parte de la teoría clásica de conjuntos, añadiendo una función de pertenencia al conjunto, definida como un número real entre 0 y 1. Así, se introduce el concepto de conjunto o subconjunto borroso asociado a un determinado valor lingüístico, definido por una palabra, adjetivo o etiqueta lingüística A, es decir podemos definir a un subconjunto borroso como $A = \{(x, \mu_A(x)) \mid x \in X\}$ siendo la función de pertenencia:

$$\begin{aligned}\mu_A : X &\rightarrow [0,1] \\ x \in X &\rightarrow \mu_A(x) \in [0,1]\end{aligned}$$

Donde 0 indica la no pertenencia al conjunto A y 1 la pertenencia absoluta. Evidentemente, existe una degradación del nivel de pertenencia de forma que si $\mu_A(x) = 0.9$, el nivel de pertenencia del elemento x es muy elevado, y si $\mu_A(x) = 0.1$ el nivel de pertenencia de x es muy bajo. Así puede interpretarse como el grado en que un elemento particular que se considera cumple con las especificaciones que definen a los elementos del conjunto en cuestión y no debe interpretarse como la probabilidad de pertenencia. Si la probabilidad de que un elemento x pertenece al conjunto A es de 0.9 y se afirma que x pertenece al conjunto A , tenemos un 90 % de probabilidad de acertar, pero el elemento intrínsecamente pertenece o no pertenece a A . Cuando se dice que la función de pertenencia de x es 0.9 se quiere decir que cumple en nuestro criterio con el 90% de las características que definen los elementos del conjunto A . En resumen, la probabilidad indica incertidumbre estadística mientras que la función de pertenencia indica vaguedad y subjetividad.

Desde la aparición de la lógica borrosa son incontables las aplicaciones que se han hecho de ella en el mundo de la investigación en general y en particular en la Ciencia Madre, “Las Matemáticas”, ellas de forma general siguen el esquema de la figura 1.2. No obstante, quizás la principal aplicación actual sean los sistemas de control borroso, que utilizan sus expresiones para formular reglas orientadas al control de sistemas (Brubaker 1992; Nass 1992). Dichos sistemas de control borroso pueden considerarse una extensión de los sistemas expertos, pero superando los problemas prácticos que éstos presentan en el razonamiento en tiempo real, causados por la explosión exponencial de las necesidades de cálculo requeridas para el análisis lógico completo de las amplias bases de reglas que manejan. Un ejemplo relevante es el Metro de la ciudad japonesa de Sendai inaugurado el 15 de julio de 1987, (Martín del Brio 2005).

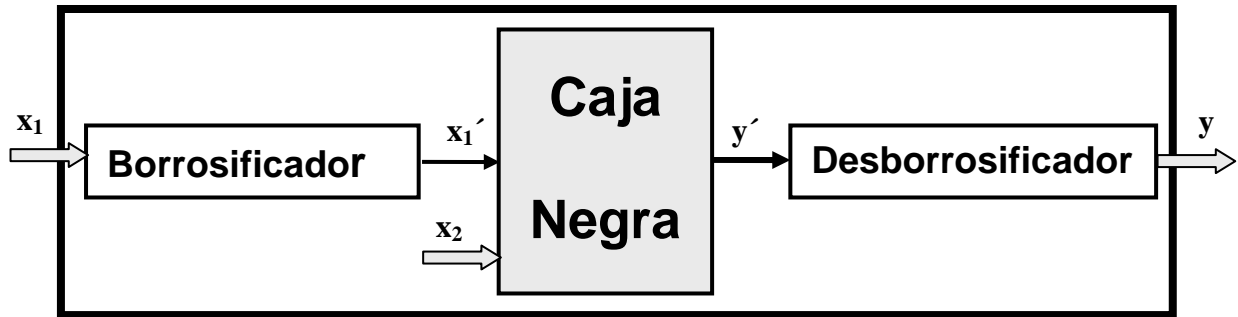


Figura 1.2 Funcionamiento de un sistema de control difuso.

1.5.1 Funciones de pertenencia

La función de inclusión o pertenencia (*membership function*) de un conjunto borroso consiste en un conjunto de pares ordenados $A = \{ (u, \mu_A(u)) / u \in U \}$ si la variable es discreta, o una función continua si no lo es. Como ya se ha comentado, el valor de $\mu_A(u)$ indica el grado en que el valor u de la variable U está incluida en el concepto representado por la etiqueta A . Para la definición de estas funciones de pertenencia se utilizan convencionalmente ciertas familias de formas estándar, en las que se encuentran las siguientes:

Función de pertenencia triangular

Se define por sus límites inferior a y superior b , y el valor modal m , tal que $a < m < b$.

$$A(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{(x - a)}{(m - a)} & \text{si } x \in (a, m] \\ \frac{(b - x)}{(b - m)} & \text{si } x \in (m, b) \\ 0 & \text{si } x \geq b \end{cases}$$

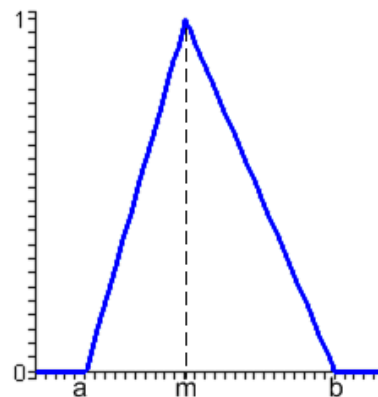


Figura 1.3 Función de pertenencia triangular.

Función de pertenencia trapezoidal

Definida por sus límites inferior a y superior d, y los límites de su soporte, b y c, inferior y superior respectivamente.

$$A(x) = \begin{cases} 0 & \text{si } (x \leq a) \text{ ó } (x \geq d) \\ \frac{(x - a)}{(b - a)} & \text{si } x \in (a, b] \\ 1 & \text{si } x \in (b, c) \\ \frac{(d - x)}{(d - c)} & \text{si } x \in (c, d) \end{cases}$$

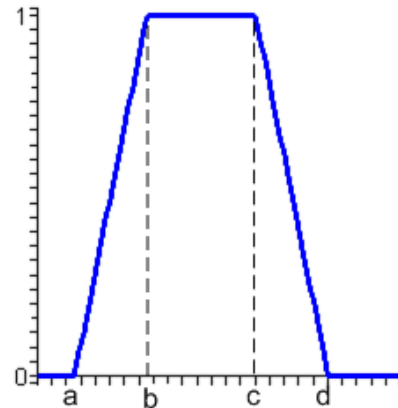


Figura 1.4 Función de pertenencia trapezoidal.

Función de pertenencia Gaussiana

Definida por su valor medio m y el valor $k > 0$. Es la típica campana de Gauss. Cuanto mayor es el valor de k, más estrecha es la campana:

$$A(x) = e^{-k(x - m)^2}$$

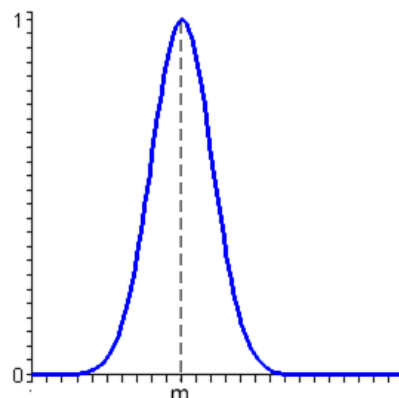


Figura 1.5 Función de pertenencia Gaussiana.

Función de pertenencia S

La Función S está definida por sus límites inferior a y superior b, y el valor m, o punto de inflexión tal que $a < m < b$. El valor típico es: $m = (a+b) / 2$. El crecimiento es más lento cuanto mayor sea la distancia a - b.

$$A(x) = \begin{cases} 0 & \text{si } x \leq a \\ 2 \left(\frac{(x-a)}{(b-a)} \right)^2 & \text{si } x \in (a, m] \\ 1 - 2 \left(\frac{(x-b)}{(b-a)} \right)^2 & \text{si } x \in (m, b) \\ 1 & \text{si } x \geq b \end{cases}$$

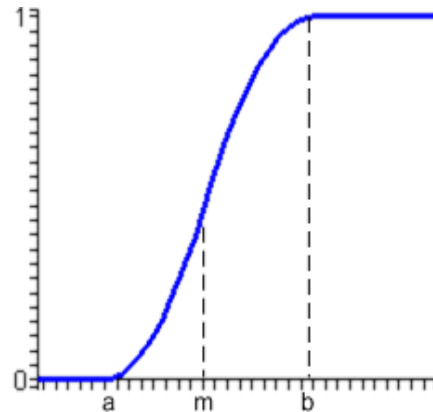


Figura 1.6 Función de pertenencia S.

1.5.2 Borrosificador

El borrosificador establece una relación entre los puntos, $x = (x_1, x_2, \dots, x_n)$, de entrada no borrosos del sistema, y sus correspondientes conjunto borroso A en U (las variables procedentes del exterior serán, en general, valores no borrosos y habrá que borrosificarlas previamente). Se pueden utilizar diversas estrategias de borrosificación: (Martín del Brio 2005)

Borrosificador Singleton

Es el método de borrosificación más utilizado, sistemas de control, y consiste en considerar los propios valores discretos como conjuntos borrosos. De otra forma, para cada valor de entrada x se define un conjunto A' que lo soporta, con función de pertenencia $\mu_{A'}(x')$, de modo que:

$$\mu_{A'}(x') = \begin{cases} 1 & \text{si } x' = x \\ 0 & \text{si } x' \neq x \end{cases} \quad \forall x' \in U$$

Borrosificador no Singleton

En este método de borrosificación se utiliza una función exponencial del tipo siguiente:

$\mu_{A'}(x') = a \cdot \exp\left[-\left(\frac{x' - x}{\sigma}\right)^2\right]$ función con forma de campana, centrada en el valor x de entrada, de ancho σ y amplitud a .

1.5.3 Desborrosificador

El desborrosificador es la función que transforma un conjunto borroso V , normalmente salida de un dispositivo de inferencia borrosa, en un valor no borroso $y \in V$, (Martín del Brio 2005). Para esta tarea se utilizan diversos métodos tales como:

Desborrosificador por máximo

Consiste en $y' = \arg.\sup_{y \in V} (\mu_{B'}(x))$ es decir, y' es el punto de V en que $\mu_{B'}(x)$ alcanza su valor máximo, donde $\mu_{B'}(x)$ es el conjunto de los grados de pertenencia de todas las etiquetas analizadas en el problema.

Desborrosificador por medida de centros de salidas:

$$y = \frac{\sum_{i=1}^M y(\mu_{B^i}^-(y))}{\sum_{i=1}^M (\mu_{B^i}^-(y))}$$

y^- representa el centro del conjunto borroso G^i , es decir, el punto en V donde $\mu_{G^i}^-(y)$ alcanza su valor máximo y $\mu_{B'}(y) = \sup_{x \in U} [\mu_{F_1^1 x \dots x F_n^1 \rightarrow G'}(x, y) * \mu_{A'}(x)]$

Desborrosificador por centro de área

$$y = \frac{\sum_{i=1}^M M^i(\mu_{B^i}^-(y))}{\sum_{i=1}^M A^i(\mu_{B^i}^-(y))}$$

donde M^l es el momento (entorno al eje y del universo de discurso de la salida V) de la función de inclusión del conjunto borroso G^l , A^l es el área, y $\mu_{B'}(y)$ se define como:

$$\mu_{B'}(y) = \sup_{x \in U} \left[\mu_{F_1^l x \dots x F_n^l \rightarrow G'}(x, y) * \mu_{A'}(x) \right]$$

1.6 Paquetes estadísticos para la detección de conglomerados

En el mundo existen varios paquetes específicos para detectar conglomerados de enfermos.

1.6.1 Paquetes generales

Uno de ellos es el denominado CLUSTER, (Aldrich 1993). Fue creado por Tim Aldrich, Wanzer Drane y otros colegas pertenecientes a *"The Agency for Toxic Substances and Disease Registry"*. El *software* cuenta con tres categorías principales, seis subcategorías y doce métodos estadísticos para la detección de aglomeraciones espaciales, temporales y espacio - temporales.

SaTScan es un *software* específico para la detección de conglomerados en espacio, tiempo y en espacio - tiempo basados solamente en métodos Scan. Está diseñado para determinar si un conjunto de casos de enfermos forman o no una aglomeración significativa, para comprobar por ejemplo que la enfermedad en estudio se distribuya aleatoriamente por todo el espacio, por todo el tiempo o por ambos a la vez, y para detectar áreas geográficas que tengan una incidencia de enfermos significativamente alta o significativamente baja.

La compañía BioMedware es también la autora de un nuevo producto de *software* al que denominó ClusterSeer. Su objetivo es en esencia el mismo que el de los dos sistemas anteriores, pero se diferencia en que incluye técnicas que permiten localizar la aglomeración, determinar cuándo surgió e intenta definir si el conglomerado se encuentra cerca de un presunto foco infeccioso.

Todos estos sistemas son útiles para la detección de aglomeraciones significativas de enfermos, pero ninguno de ellos brinda facilidades para realizar estudios combinados con factores de riesgo.

En Cuba ya se encuentra registrada la primera versión del sistema EpiDet (*Epidemic Detector*), que contiene varios métodos de detección de clusters. EpiDet permite además trabajar con factores de riesgo adicionales, (Casas Cardoso 2002) .

1.6.2 Sistema estadístico para la detección de conglomerados: EpiDet

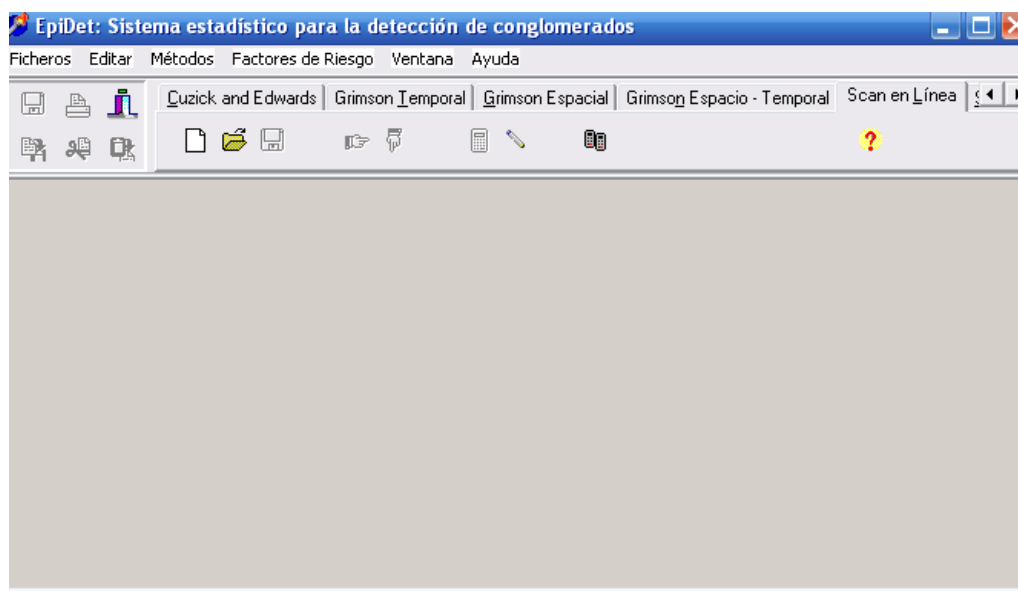


Figura 1.7 EpiDet: Sistema estadístico para la detección de conglomerados.

El sistema EpiDet (*Epidemic Detector*) elaborado en la Universidad Central de las Villas se utiliza para la detección de conglomerados (posibles epidemias en un ambiente epidemiológico) utilizando las técnicas descritas con anterioridad. Se ejecuta sobre *Windows* y brinda al usuario un ambiente cómodo. El sistema consta de dos ficheros: el ejecutable EpiDet.exe y su ayuda: EpiDet.hlp. Lee los datos a partir de ficheros textos con una determinada estructura en dependencia del tipo de método que se va a ejecutar.

El EpiDet trabaja con localizaciones exactas en tiempo (fecha desglosada en día, mes y año); en espacio con localizaciones exactas (coordenadas X e Y) e imprecisas (celda a la que pertenece el enfermo). Tiene incorporados los métodos seleccionados para la realización de este trabajo. Para detectar conglomerados temporales se implementó el método Scan en sus dos variantes: sobre una línea y sobre un círculo, los productos de *software* anteriores sólo contemplan el primero; el cálculo de las probabilidades también se hace de manera exacta y utilizando la versión sugerida por Nauss. Para detectar

aglomeraciones espaciales el sistema cuenta con el método de Pearson y con el de Cuzick and Edwards y para el análisis de la interacción se implementó, además del conocido método de Knox clásico, su perfeccionamiento según Abe para el caso de tablas 2 x 2, lo que ofrece la posibilidad de utilizar ambas técnicas, compararlas y llegar a conclusiones más fieles y seguras. El método de Grimson está presente en las tres variantes anteriores.

La importancia y novedad del EpiDet radica en la posibilidad que brinda de realizar estudios combinados con factores de riesgo para intentar esclarecer las causas de las aglomeraciones en aquellos casos en que no resulte evidente. En este sentido EpiDet tiene incorporado el principio general de inclusión de factores y los métodos de Knox y Grimson cuentan además con la vía particular correspondiente, (Casas Cardoso 2003).

1.7 Consideraciones finales

En los últimos años se ha perfeccionado el Sistema de Vigilancia Epidemiológica en Cuba, esto ha sido posible gracias a la interacción de múltiples disciplinas. El análisis del comportamiento y tendencias de las enfermedades se ha desarrollado notablemente con la introducción de metodologías y técnicas estadístico - matemáticas y epidemiológicas.

Las técnicas de detección de conglomerados han sido de gran utilidad en los estudios de posibles epidemias, existen en la literatura numerosos métodos de detección de conglomerados, siendo los métodos Scan uno de los más importantes para la detección de conglomerados temporales.

CAPÍTULO II. LOS MÉTODOS SCAN

En este capítulo se describen los fundamentos matemáticos de los métodos Scan clásicos y borrosos sobre una línea y sobre un círculo y se presenta de manera breve las funciones elaboradas usando el paquete Mathematica.

2.1 Los métodos Scan Clásicos

Los métodos Scan se utilizaron inicialmente para detectar aglomeraciones dentro de períodos de tiempo consecutivos, pues un conglomerado temporal puede extenderse por dos o más intervalos. Todos los casos diagnosticados deben estar ordenados cronológicamente de acuerdo con la fecha de primeros síntomas o de diagnóstico de la enfermedad, de muerte o cualquier otro evento de salud que se considere, (Casas Cardoso 2002).

Lo primero que se hace con el método es definir un intervalo o una ventana de tamaño fijo de acuerdo con la duración esperada de la epidemia. Esto debe hacerse antes de inspeccionar los datos recolectados. La ventana hallada se desplaza a lo largo de la línea del tiempo y se determinan en cada caso la cantidad de enfermos asociadas a ella. El método se basa en la idea de que si existe un conglomerado, el número máximo de casos hallados en la ventana debe ser grande. El test estadístico en esencia calcula la probabilidad p de que aparezcan n o más casos en una ventana. La significación p hallada, para un conjunto particular de casos, depende del ancho de la ventana y del paso del Scan seleccionado por el investigador. Como esto puede ser subjetivo resulta de gran utilidad hacer varias repeticiones utilizando amplitudes diferentes, especialmente cuando no se conoce la duración esperada de la supuesta epidemia, (Valdés Hernández 2009).

2.1.1 Método Scan Clásico en una línea para la detección de conglomerados temporales

Sean X_1, X_2, \dots, X_N variables aleatorias independientes e idénticamente distribuidas que denotan el tiempo de ocurrencia de N eventos en un intervalo $(0, T]$. El método Scan prueba la hipótesis nula de que los eventos están uniformemente distribuidos en todo el

período contra la alternativa de que forman al menos un conglomerado en algún subintervalo de $(0, T]$. Nauss originalmente calculó la probabilidad de que existiera un subintervalo de $(0, T]$ de tamaño fijo w que contuviera k o más de los N eventos. El estadígrafo SW se definió como el número máximo de eventos definidos en un intervalo o ventana de longitud w que se mueve de manera continua sobre todo $(0, T]$, (Rodríguez Corvea 2008).

Para la formulación sean:

t : amplitud de la ventana,

T : período de tiempo total que se analiza,

$L = T/t$: fracción que representa el período de tiempo total que se analiza con relación al ancho de la ventana,

n : cantidad de enfermos diagnosticados en T ,

λ : número esperado de casos por unidad de tiempo en un proceso de Poisson,

$w_{y, y+t}$: cantidad de enfermos en la ventana $[y, y+t)$.

Hipotéticamente el estadígrafo: $w = w_t(T) = \max_{0 \leq y \leq T-t} \{w_{y, y+t}\}$ representa el número de casos que aparecen en una ventana cuando se mueve continuamente a lo largo del tiempo. En la práctica, la ventana $[y, y+t)$ se mueve discretamente a partir de una sucesión de puntos equidistantes y_1, y_2, \dots, y_k que cubren todo el período de análisis de amplitud T . Se denomina paso del Scan o paso del desplazamiento a $\Delta y = y_k - y_{k-1}$.

Realmente, el estadígrafo anterior se estima por su versión discreta:

$$\bar{w} = \bar{w}_{t, \Delta t}(T) = \max_{1 \leq i \leq k} \{w_{y_i, y_{i+t}}\}$$

La idea del método es que, si existe un conglomerado, el número máximo de casos hallados en una ventana debe ser grande. El test estadístico depende de varios de los parámetros explicados con anterioridad y en esencia calcula la probabilidad p de que

aparezcan w o más casos en una ventana. La fórmula que se utilizó para p es la propuesta en (Naus 1982):

$$p = P^*(w, \lambda L, 1/L) = 1 - Q^*(w, \lambda L, 1/L) \quad (2.1)$$

donde Q^* puede ser aproximado para cualquier $L > 2$ a partir de sus valores con $L = 2$ y $L = 3$.

$$Q^*(w, \lambda L, 1/L) \approx Q^*(w, 2\lambda, 1/2) [Q^*(w, 3\lambda, 1/3) / Q^*(w, 2\lambda, 1/2)]^{L-2} \quad (2.2)$$

La aproximación (2.2) es fácilmente calculable usando una microcomputadora personal. El cálculo exacto de $Q^*(w, 2\lambda, 1/2)$ y $Q^*(w, 3\lambda, 1/3)$ se basa en un teorema demostrado también en (Naus 1982) y cuya esencia se resume aquí:

Para $w > 2$, $p_i = e^{-\lambda} \lambda^i / i!$, $F_w = \sum_{i=0}^w p_i$, $\lambda > 0$, se tiene que:

$$Q^*(w, 2\lambda, 1/2) = F_{w-1}^2 - (w-1)p_w p_{w-2} - (w-1-\lambda)p_w F_{w-3}$$

$$Q^*(w, 3\lambda, 1/3) = F_{w-1}^3 - A_1 + A_2 + A_3 - A_4$$

donde:

$$A_1 = 2p_w F_{w-1} ((w-1)F_{w-2} - \lambda F_{w-3})$$

$$A_2 = 0.5 p_w^2 ((w-1)(w-2)F_{w-3} - 2(w-2)\lambda F_{w-4} + \lambda^2 F_{w-5})$$

$$A_3 = \sum_{r=1}^{w-1} p_{2w-r} F_{r-1}^2$$

$$A_4 = \sum_{r=2}^{w-1} p_{2w-r} p_r ((r-1)F_{r-2} - \lambda F_{r-3})$$

donde $F_i = 0$ para todo $i < 0$.

La aproximación (2.2) puede calcularse para valores no enteros de L . Esto la diferencia de otras expresiones matemáticas que se usaban con estos fines anteriormente. Además de ser menos restrictiva, varios autores demuestran que (2.2) es mucho más precisa, (Naus 1982; Casas Cardoso 2003).

2.1.2 El método Scan sobre un círculo

Este método constituye una variación del anterior y se utiliza para problemas que tienen un comportamiento estacional. Los datos se encuentran ordenados a lo largo del eje de longitud y el círculo se forma uniendo el final con el inicial.

El algoritmo es en esencia, el mismo que el lineal. La ventana se desplaza sobre el círculo y se determina en cada una el número de enfermos asociados a ella. Con este desplazamiento circular se pretende incorporar al análisis la cercanía de posibles casos a finales del último período con los del principio del primer periodo, como si fueran períodos continuos.

Dado los detalles anteriores el algoritmo varía en el cálculo de la probabilidad de observar w o más casos en un intervalo o ventana de tamaño fijo ya que Q no se estima de igual forma, (Casas Cardoso 2003).

$$p = P_c^*(w, \lambda L, 1/L) = 1 - Q_c^*(w, \lambda L, 1/L)$$

donde ahora:

$$Q_c^*(w, \lambda L, 1/L) \approx Q^*(w, 4\lambda, 1/4) [Q^*(w, 3\lambda, 1/3)]^{L-2} [Q^*(w, 2\lambda, 1/2)]^{L-1}$$

Para hallar $Q^*(w, 4\lambda, 1/4)$ se utiliza $L=4$ en (2.2). Después de simplificar se obtiene:

$$Q^*(w, 4\lambda, 1/4) \approx [Q^*(w, 3\lambda, 1/3)]^2 / Q^*(w, 2\lambda, 1/2)$$

Luego $Q^*(w, 4\lambda, 1/4)$ también queda en función de $Q^*(w, 2\lambda, 1/2)$ y de $Q^*(w, 3\lambda, 1/3)$. Estos últimos valores se calculan en forma exacta a partir del mismo teorema anterior, (Naus 1982).

2.1.3 Consideraciones sobre los métodos Scan

Como se ha visto la significación p hallada para un conjunto particular de casos, depende del ancho de la ventana y del paso del Scan seleccionados por el investigador. Resulta imposible determinar los valores óptimos y por eso se recomienda realizar varias repeticiones del método utilizando amplitudes diferentes, especialmente cuando no se conoce la duración esperada de la supuesta epidemia. Además, esta información puede

ser muy útil para determinar algunos aspectos importantes de la naturaleza del conglomerado, en caso de existir, y de su etiología, (Rodríguez Corvea 2008).

Algunos autores han tratado de generalizar el método Scan a dos dimensiones para que pueda utilizarse en la detección de aglomeraciones espaciales de enfermos. En vez de un intervalo de longitud fija se mueve un rectángulo de dimensiones fijas sobre un área también rectangular, aunque se ha estado estudiando la posibilidad de que el conjunto a mover tenga una forma general cualquiera al igual que el área total a analizar. También se han hecho generalizaciones al caso espacio - temporal, (Casas Cardoso 2003) .

2.1.4 Limitaciones del método Scan

El método Scan asume que la población riesgo permanece constante a lo largo de todo el período de estudio, al igual que la probabilidad de detección de casos, que también se supone constante. Por tanto, el método no es válido en situaciones en las que haya algún cambio en la población o en los métodos diagnósticos de la enfermedad.

Debe tenerse especial cuidado en la selección de la ventana temporal y paso del método, pues dicha elección determina los resultados obtenidos. Cuando no se conoce la duración esperada de la supuesta epidemia, conviene aplicar varias veces el test utilizando diferentes amplitudes de ventana y paso.

El procedimiento no es útil cuando el número de casos analizados es extremadamente bajo, (Consellería 2002).

2.1.5 El Scan Lineal sobre el Mathematica

ScanValidation es el procedimiento general para el método Scan Lineal Clásico y tiene la responsabilidad dada la secuencia, el ancho de la ventana móvil y el paso de determinar:

1. Parámetro general de la distribución de Poisson (λ).
2. Estadígrafo del test
3. Fracción general de la cantidad de ventanas a formar teniendo en cuenta la longitud total de la secuencia y la longitud de la ventana móvil (L).
4. Cálculo de la significación del estadígrafo

ScanValidation[sec_,AnchoW_,Paso_] :=

```
CompoundExpression[ (* Buscar parámetro de la distribución de Poisson*)
W = Partition[sec,AnchoW,Paso]; (* Particiona la secuencia en ventanas*)
Win = N[Map[Function[lis,Plus@@lis], W], 8]; (* # de pacientes de cada ventana*)
media = N[Mean[Win]]; (* Promedio de pacientes por ventana "Landa de Poisson" *)
maximo = Max[Win];(*Print[Win];*) (*Ventana de mayor número de pacientes *)
L = N[Length[sec]/AnchoW, 9]; (* Fracción de ventanas a formar *)
signifs = N[pFinal[media,maximo,L],10]; (* Busca la significación del estadígrafo.*)
Return[signifs]
];
```

Para buscar la significación se utiliza el procedimiento pFinal donde están programadas las formulas aproximadas de (Naus 1982), donde es necesario hallar las constantes Q_2 y Q_3 , esta última utiliza las constantes A_1 , A_2 , A_3 y A_4 . Las funciones implementadas sobre el Mathematica se describen a continuación:

```
Fnn[m_, n_] := Module[{}, If[n<0,p:=0,p:=CDF[PoissonDistribution[m],n]];Return[N[p,10]]];
Psi[m_, i_] := Module[{},p:=PDF[PoissonDistribution[m],i];Return[N[p,10]]];
A1[m_, n_] := 2 Psi[m, n] Fnn[m, n - 1] ((n - 1) Fnn[m, n - 2] - m Fnn[m, n - 3])
A2[m_, n_] := 0.5 Psi[m, n]^2 ((n - 1) (n - 2) Fnn[m, n - 3] - 2 (n - 2) m Fnn[m, n - 4] + m^2 Fnn[m, n - 5])
A3[m_, n_] := Sum[Psi[m, 2 n - r] Fnn[m, r - 1]^2, {r, 1, n - 1}]
A4[m_, n_] := Sum[Psi[m, 2 n - r] Psi[m, r] ((r - 1) Fnn[m, r - 2] - m Fnn[m, r - 3]), {r, 2, n - 1}]
Q2[m_, n_] := Fnn[m, n - 1]^2 - (n - 1) Psi[m, n] Psi[m, n - 2] - (n - 1 - m) Psi[m, n] Fnn[m, n - 3]
Q3[m_, n_] := Fnn[m, n - 1]^3 - A1[m, n] + A2[m, n] + A3[m, n] - A4[m, n]
Q[m_, n_, L_] := Q2[m, n] (Q3[m, n] / Q2[m, n])^(L - 2)
pFinal[m_, n_, L_] := 1 - Q[m, n, L]
```

Donde: Fnn y Psi calcula la probabilidad acumulada y puntual, respectivamente, de un valor, dado el parámetro de Poisson.

2.2 Los métodos Scan Borrosos

En este epígrafe se propone una modificación de los métodos Scan clásicos siguiendo las ideas publicadas en (Rodríguez 2009).

2.2.1 El método Scan Borroso sobre una Línea

La idea consiste en modificar la ventana de tamaño fijo por una ventana que tenga en cada extremo una función de pertenencia, relacionada con la cantidad de enfermos por días próximos a la ventana móvil. De esa forma, se suavizan los extremos y surge el concepto de “ventana móvil borrosa”. (Rodríguez 2009).

Mediante este procedimiento se le asigna un peso a los casos que se encuentran añadidos en los extremos de la ventana móvil. Los métodos Scan borrosos (lineal y circular), utilizan esos valores de pertenencia para determinar su influencia en la salida borrosa que representa el número de casos detectados en la ventana, (Valdés Hernández 2009).

La ventana borrosa se define analíticamente de la siguiente forma:

$$\text{Ventana borrosa}_k = \begin{cases} (i - k + g + 1) * \frac{s_i}{(g + 1)} & i = k - g, \dots, k \\ s_i & i = k, \dots, k + t - 1 \\ (k + t + g - i) * \frac{s_i}{(g + 1)} & i = k + t, \dots, k + t + g - 1 \end{cases}$$

donde:

s_1, s_2, \dots, s_n : es la secuencia que representa la cantidad de enfermos por días,

t : longitud de la ventana fija,

g_i : longitud de la parte borrosa de la nueva ventana. A esta parte se le llamará “suavizado”.

La formulación matemática del test es esencialmente la misma: el método “escanea” los datos usando una ventana móvil borrosa. Ahora, el número máximo de casos reportados en una ventana no es necesariamente un valor entero, sino real. La figura 2.1 muestra una representación gráfica de ambas ventanas.

Método Scan (t=5)		
	Clásico	Fuzzy (g=1)
Secuencia:	0 3 1 2 4 0 1 3 0 8 1 4	0 3 1 2 4 0 1 3 0 8 1 4
Ventana:	<div style="border: 1px solid black; display: inline-block; padding: 2px;">10</div>	<div style="display: inline-block; padding: 2px;">.5 \ + 10 + / 0</div>
Estadígrafo:	$\eta_{\max} = 10$	$\eta^*_{\max} = 10.5$

Figura 2.1 Ventanas clásica y borrosa en el método Scan sobre una línea.

Obsérvese en el epígrafe 2.2.1 que el valor de la significación del método Scan sobre una Línea se basa en distribuciones de Poisson. Esta distribución está definida para variables aleatorias discretas, luego hay que realizar transformaciones al cálculo de la significación en el método clásico.

Concretamente se proponen tres formas diferentes de calcular la significación:

1. Aproximar el valor real al valor entero más próximo. Las distribuciones de probabilidad y de distribución de Poisson se utilizan con bastante frecuencia dentro de las expresiones que aparecen en el epígrafe 2.1.1 y que se utilizan para hallar el valor de p. De aquí se deduce que la propagación del error pudiera no ser tan pequeña. Nos referiremos a este método como aproximación borrosa 1, ver figura 2.2

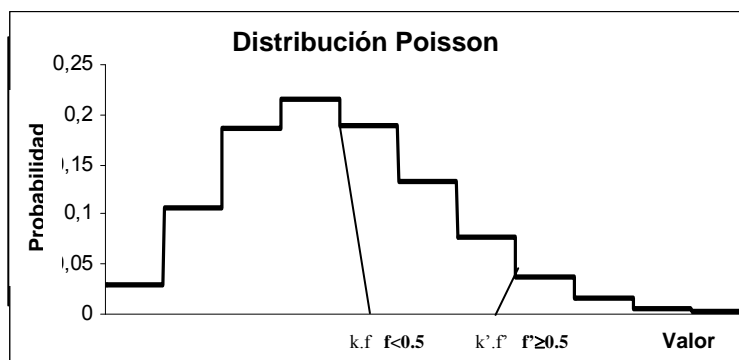


Figura 2.2 Función de probabilidad de Poisson ajustada usando la aproximación al valor entero más próximo, (aproximación borrosa 1).

2. Aproximar el valor real usando una combinación de dos distribuciones: Poisson hasta el valor entero inferior y uniforme para estimar la parte decimal. Nos referiremos a este método como aproximación borrosa 2, ver figura 2.3.

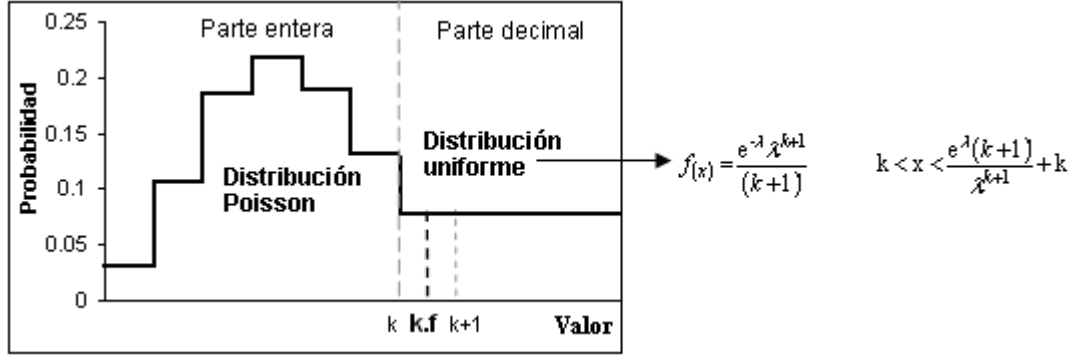


Figura 2.3 Función de probabilidad de Poisson ajustada usando la aproximación basada en dos distribuciones: Poisson y uniforme, (aproximación borrosa 2).

Las fórmulas originales de Naus (Naus 1982), necesitan ser modificadas de la manera siguiente:

- $$P[x \leq k.f] = \sum_{n=0}^k \frac{\lambda^n}{n!} e^{-\lambda} + f * \frac{e^{-\lambda} \lambda^{k+1}}{(k+1)!}$$
- $$P[x = k.f] = \frac{e^{-\lambda} \lambda^k}{k!} + f * \left(\frac{e^{-\lambda} \lambda^k}{k!} - \frac{e^{-\lambda} \lambda^{k+1}}{(k+1)!} \right)$$
- $$A3 = \sum_{r=1+f}^{k.f} P[x = 2 * k.f - r] * P[x \leq r-1]^2$$
- $$A4 = \sum_{r=2+f}^{k.f} P[x = 2 * k.f - r] * P[x = r] ((r-1)P[x \leq r-2] - \lambda P[x \leq r-3])$$

3. Aproximar el valor real utilizando funciones de interpolación. La interpolación es un método matemático de “construcción” de nuevos datos a partir de los ya existentes. En nuestro caso, los datos ya existentes se corresponden con las funciones de probabilidad y de distribución de Poisson respectivamente. Se utilizó un polinomio de interpolación de grado 4, nos referiremos a este método como aproximación borrosa 3, ver figura 2.4.

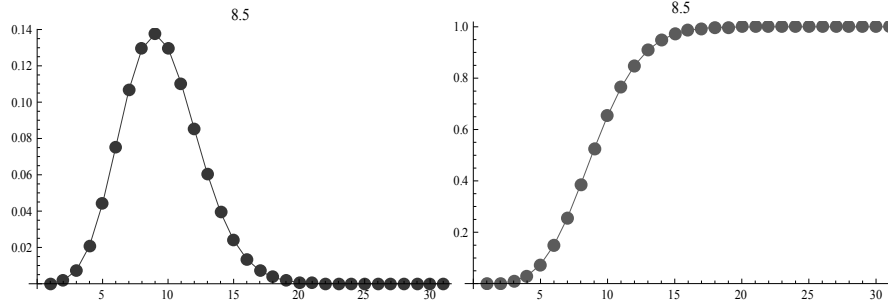


Figura 2.4 Funciones de interpolación para las funciones de probabilidad y de distribución de Poisson, (aproximación borrosa 3).

Finalmente, la respuesta del método se “particiona” en dos conjuntos borrosos con las etiquetas: “significativo” y “no significativo”. Cada uno de ellos tiene una función de pertenencia S de la manera siguiente:

$$\text{No significativo: } S_{(u, 0.05, 0.0625, 0.075)} = \begin{cases} 0 & u \leq 0.05 \\ 2 * \frac{u - 0.05}{0.025^2} & 0.05 < u < 0.0625 \\ 1 - 2 * \frac{u - 0.075}{0.025^2} & 0.0625 \leq u < 0.075 \\ 1 & u \geq 0.075 \end{cases}$$

$$\text{Significativo: } S_{(u, 0.075, 0.0875, 0.1)} = \begin{cases} 1 & u \leq 0.075 \\ 1 - 2 * \frac{u - 0.075}{0.025^2} & 0.075 < u < 0.0875 \\ 2 * \frac{u - 0.0875}{0.025^2} & 0.0875 \leq u < 0.1 \\ 0 & u \geq 0.1 \end{cases}$$

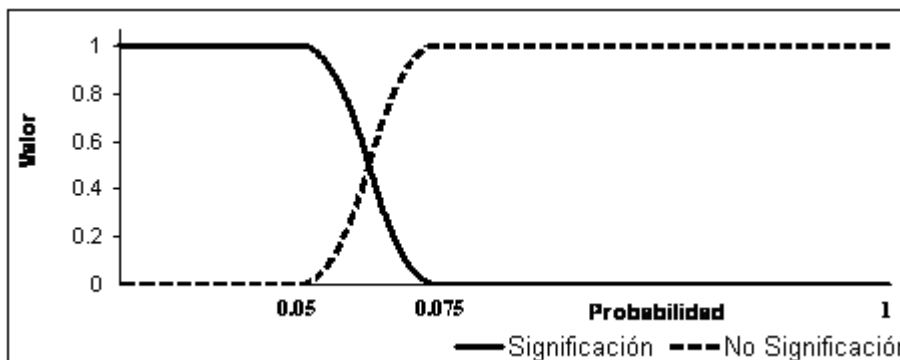


Figura 2.5 Funciones de pertenencia borrosas: “significativo” y “no significativo”.

Se aplica el método del máximo para eliminar el término borroso y obtener una respuesta dura (Martín del Brio 2005).

2.2.2 El método Scan Borroso sobre un círculo

El Scan Borroso sobre un círculo, se obtiene de una forma similar a su equivalente lineal

Al igual que su variante lineal, la ventana móvil se “suaviza” con una función de pertenencia en cada uno de sus extremos. El método borroso, usa estas funciones de pertenencia como factores de peso que influyen en la cantidad de casos reportados en cada día próximos a la ventana analizada. Analíticamente, la ventana se define como:

$$\text{Ventana borrosa}_k = \begin{cases} (i - k + g + 1) * \frac{S_i}{(g + 1)} & i = k - g, \dots, k \\ S_i & i = k, \dots, k + t - 1 \\ (k + t + g - i) * \frac{S_i}{(g + 1)} & i = k + t, \dots, k + t + g - 1 \end{cases}$$

donde:

$S_1, S_2, S_3, \dots, S_n, S_{n+1}, S_{n+2}, \dots, S_{n+t-1}$ es la secuencia binaria

$S_{n+i} = S_i$ para $i = 1$ to $t - 1$

- If $i < 1$ then $S_i = S_{n-i}$
- If $i > n + t - 1$ then $S_i = S_{n-i}$

g: longitud de la parte borrosa de la nueva ventana.

La formulación matemática del test es esencialmente la misma: el método “escanea” la secuencia circular y contabiliza la cantidad de enfermos en cada ventana suavizada. Como la ventana es ahora borrosa, el máximo número de casos reportados en una ventana no es necesariamente un valor entero, sino real.

De la misma forma que para el método lineal, se definen tres formas diferentes de calcular la significación del test:

- Aproximar el valor real al valor entero más próximo.

- Aproximar el valor real usando una combinación de dos distribuciones: Poisson hasta el valor entero inferior y uniforme para estimar la parte decimal.
- Aproximar el valor real utilizando funciones de interpolación.

La explicación de estas formas es básicamente la misma que se explicó en el epígrafe anterior.

Igualmente, la respuesta del método se “particiona” en dos conjuntos borrosos con las etiquetas: “significativo” y “no significativo”. Cada uno de ellos tiene una función de pertenencia S como muestra la figura 2.5.

Se desborrosifica aplicando el método del máximo para eliminar el término borroso y obtener una respuesta dura.

2.2.3 Algunas consideraciones sobre los métodos Scan Borrosos

En estudios de simulación realizados en (Rodríguez 2007) se ha demostrado la superioridad de los métodos Scan borrosos en relación con los clásicos.

No obstante, la detección del parámetro óptimo para el tamaño de la ventana, o al menos la detección de un parámetro adecuado, sigue siendo un problema no resuelto. En algunas aplicaciones epidemiológicas, en las que se conoce bien el comportamiento de una determinada enfermedad, la selección del ancho de la ventana puede no ser un problema tan grave. Sin embargo, en otras aplicaciones, como por ejemplo en estudios bioinformáticos, esta selección a priori no resulta ser tan sencilla. La selección de parámetros no adecuados, puede conllevar a falsas conclusiones.

2.2.4 El Scan Lineal Borroso sobre el Mathematica

En el Scan lineal Borroso modificamos el procedimiento general ScanValidation utilizado en Scan Lineal Clásico, es decir, necesita los parámetros secuencia, ancho de la ventana móvil, paso y es necesario añadir el parámetro incremento que significa la cantidad de elementos que suaviza las ventanas en ambos laterales, para mayor comprensión los aspectos similares al Scan Clásico se señalan con menor intensidad. En general se encarga de determinar:

- Parámetro general de la distribución de Poisson (λ),
- Suavizar cada una de las ventanas,
- Estadígrafo del test,
- Fracción general de la cantidad de ventanas a formar teniendo en cuenta la longitud total de la secuencia y la longitud de la ventana móvil (L),
- Cálculo de la significación del estadígrafo en dependencia del tipo de método.

ScanValidation[sec_,AnchoW_,Paso_, Incr_]:=

CompoundExpression[

(* Buscar parámetros de la distribución de Poisson*)

W = Partition[sec, AnchoW, Paso]; (* Particiona la secuencia en ventanas*)

Win = N[Map[Function[lis, Plus@@lis], W], 6]; (* # de pacientes de cada ventana*)

media = N[Mean[Win]]; (* Promedio de pacientes por ventana "Landa de Poisson" *)

W = If[Incr>0, Fuzzy[sec, AnchoW, Paso, Incr], Partition[sec, AnchoW, Paso]];

(* Fuzzifica o Calcula cada una de las ventanas suavizadas o normales *)

Win = N[Map[Function[lis, Plus@@lis], W], 6]; (* Recalcula # de pacientes x ventanas*)

maximo = Max[Win];(*Print[Win];*) (*Ventana de mayor número de pacientes *)

L = N[Length[sec]/AnchoW,9]; (* Fracción de ventanas a formar *)

signifs = If[Incr == 0, {N[Pfinal[media,maximo,L],10]}; (*Significación Scan Clásico.*)

signifs = If[Incr <> 0

{DesFuzzificacion[N[Pfinal[media,Round[maximo],L],10]],

DesFuzzificacion[N[Pfinal[media,maximo,L],10]],

DesFuzzificacion[N[NausSignif[media,maximo,L],10]]}];

(*Significación Scan Borroso, por las tres vías permitidas.*)

Return[**signifs**]

];

En el cálculo de la significación en Scan Lineal Borroso, se utilizan las vías anteriormente explicadas. Veamos:

Primer procedimiento

```
Fnn[m_, n_] := Module[{}, If[n<0, p := 0, p :=  
CDF[PoissonDistribution[m], Floor[n]] +  
PDF[PoissonDistribution[m], Ceiling[n]]*FractionalPart[n]];  
Return[N[p,10]];
```

(* Calcula probabilidad acumulada. Si estadígrafo es entero el segundo sumando es nulo *)

```
Psi[m_, i_] := Module[{}, p = Psi1[m,i] - (Psi1[m,i] -Psi2[m,i]) *  
FractionalPart[i]; Return[N[p,10]]]
```

(*Calcula probabilidad puntual. Si estadígrafo es entero el segundo miembro es nulo *)

```
Psi1[m_, i_] := Module[{}, p := PDF[PoissonDistribution[m], Floor[i]];  
Return[N[p,10]]]
```

(*Calcula probabilidad puntual, del entero más pequeño del estadígrafo *)

```
Psi2[m_, i_] := Module[{}, p := PDF[PoissonDistribution[m], Ceiling[i]];  
Return[N[p,10]]]
```

(*Calcula probabilidad puntual, del entero más grande del estadígrafo *)

$A1[m_, n_] := 2 \text{ Psi}[m, n] \text{ Fnn}[m, n-1] ((n-1) \text{ Fnn}[m, n-2] - m \text{ Fnn}[m, n-3])$

$A2[m_, n_] := 0.5 \text{ Psi}[m, n]^2 ((n-1) (n-2) \text{ Fnn}[m, n-3] - 2 (n-2) m \text{ Fnn}[m, n-4] + m^2 \text{ Fnn}[m, n-5])$

$A3[m_, n_] := \sum_{r=1+\text{FractionalPart}[n]}^{n-1} \text{ Psi}[m, 2n-r] \text{ Fnn}[m, r-1]^2$

$A4[m_, n_] := \sum_{r=2+\text{FractionalPart}[n]}^{n-1} \text{ Psi}[m, 2n-r] \text{ Psi}[m, r] ((r-1) \text{ Fnn}[m, r-2] - m \text{ Fnn}[m, r-3])$

$Q2[m_, n_] := \text{ Fnn}[m, n-1]^2 - (n-1) \text{ Psi}[m, n] \text{ Psi}[m, n-2] - (n-1-m) \text{ Psi}[m, n] \text{ Fnn}[m, n-3]$

$Q3[m_, n_] := \text{ Fnn}[m, n-1]^3 - A1[m, n] + A2[m, n] + A3[m, n] - A4[m, n]$

$Q[m_, n_, L_] := Q2[m, n] (Q3[m, n] / Q2[m, n])^{L-2}$

$p\text{Final}[m_, n_, L_] := 1 - Q[m, n, L]$

Segundo procedimiento

```
FPsi1[n_, flpdf_] := Module[{}, If[n<0, p=0, FPsi=Interpolation[flpdf]; p=FPsi[n]]; Return[N[p,10]]]
```

(*Calcula probabilidad puntual usando una función de interpolación *)

```
FFnn1[n_, flcdf_] := Module[{}, If[n<0, p=0, FFnn=Interpolation[flcdf]; p=FFnn[n]]; Return[N[p,10]]]
```

(*Calcula probabilidad acumulada usando una función de interpolación *)

NausSignif[m_, n_, L_] :=

Module[{},

lp = Table[{k,PDF[PoissonDistribution[m],k]},{k,-1,2 n+1}};

(*Función de interpolación de probabilidades*)

lc = Table[{k,CDF[PoissonDistribution[m],k]},{k,-1,2 n+1}};

(*Función de interpolación de probabilidades acumulada*)

FA1 = 2 **Fpsi1**[n,**lp**] **FFnn1**[n-1,**lc**] ((n-1) **FFnn1**[n-2,**lc**]-m **FFnn1**[n-3,**lc**]);

FA2 := .5 (**Fpsi1**[n,**lp**])²((n-1) (n-2)**FFnn1**[n-3,**lc**]-2 (n-2)m **FFnn1**[n-4,**lc**]+m² **FFnn1**[n-5,**lc**]);

FA3 :=
$$\sum_{r=1+\text{FractionalPart}[n]}^{n-1} \text{Fpsi1}[2\ n-r,\mathbf{lp}] \text{FFnn1}[r-1,\ \mathbf{lc}]^2 ;$$

FA4 :=
$$\sum_{r=2+\text{FractionalPart}[n]}^{n-1} \text{Fpsi1}[2\ n-r,\mathbf{lp}] \text{Fpsi1}[r,\mathbf{lp}] ((r-1) \text{FFnn1}[r-2,\ \mathbf{lc}]-m \text{FFnn1}[r-3,\mathbf{lc}])$$

FQ2 := **FFnn1**[n-1,**lc**]² -(n-1) **Fpsi1**[n,**lp**] **Fpsi1**[n-2,**lp**]- (n-1-m) **Fpsi1**[n,**lp**] **FFnn1**[n-3,**lc**]

FQ3 := **FFnn1**[n-1,**lc**]³ - -**FA1**+**FA2**+**FA3**-**FA4**;

FQ := **FQ2** (**FQ3** / **FQ2**)^{L-2};

Pfin := 1-**FQ**;

Return[N[**Pfin**,10]];

]

Para suavizar las ventanas se necesitan tres procedimientos:

- **Fuzzy**: Procedimiento general borroso, que dirige los siguientes procedimientos:
 - **IncremTamWindows**: Suaviza todas las ventanas de la secuencia, es decir, añade los elementos adecuados por la izquierda de cada ventana y posteriormente hace el mismo procedimiento por la derecha.
 - **Fuzzificacion**: Pesa los elementos suavizados de cada ventana en dependencia de su valor y posición dentro de la ventana.

Procedimiento General

Fuzzy[sec_, AnchoW_, Paso_, Incr_] :=

CompoundExpression[

W = IncremTamWindows[sec, AnchoW, Paso, Incr]; (* Suaviza todas las ventanas *)

Inc1 = 1/(Incr+1);

(* Fracción general que aporta al peso cada elemento suavizado de una ventana *)

TW = Length[W];

W2 = W; (* W, W2 lista con las ventanas suavizadas *)

Map[Function[x, **W2** = Fuzzificacion[W2, TW, Inc1*x, x]], Range[Incr]];

(* Valoriza la parte izquierda de cada ventana *)

L2 = AnchoW+2*Incr+1; (* L2 cantidad de elementos de una ventana suavizada *)

Map[Function[x, **W2** = Fuzzificacion[W2, TW, Inc1*x, L2-x]], Range[Incr]];

(* Valoriza la parte derecha de cada ventana *)

Return[**W2**]

]

Procedimiento que permite suavizar cada ventana de la secuencia

IncremTamWindows[sec_, AnchoW_, Paso_, Incr_] :=

CompoundExpression[

W = Partition[sec, AnchoW, Paso];

TW = Length[W];

sec1=PadLeft[sec,Length[sec]+Incr]; (* Inserta ceros a la izquierda de la secuencia *)

W1 = Map[Function[z,Join[Take[sec1,{(z-1)*Paso+1,(z-1)*Paso+Incr}],W[[z]]],Range[TW]];

(* Suaviza parte izquierda de las ventanas *)

sec1 = PadRight[sec,Length[sec]+Incr]; (* Inserta ceros a la derecha de la secuencia *)

W1 = Map[Function[z,Join[W1[[z]],Take[sec1,{(z-1)*Paso+1+AnchoW,(z-1)*Paso+AnchoW +Incr}]]],Range[TW]];

(* Suaviza parte derecha de las ventanas *)

Return[**W1**];

]

Procedimiento que pesa los elementos suavizados de cada ventana

Fuzzificacion[W_, L1_, Valor_, Pos_] :=

CompoundExpression[

KK=Map[Function[z,If[(W[[z, Pos]]!=0) ,
ReplacePart[W[[z]],Valor*W[[z,Pos]],Pos],W[[z]]], Range[L1]];

(* Dada la posición de un elemento móvil lo pesa según su valor en todas las ventanas *)

Return[KK]

];

El valor borroso se desborrosifica utilizando la variante que toma como resultado final el conjunto borroso de mayor valor.

DesFuzzificacion[x1_] :=

CompoundExpression[(* x valor de la significación fuzzificada *)

Which[

$x1 \geq 0.075$, **gs** = 0,

$x1 \leq 0.05$, **gs** = 1,

$x1 < 0.0625$, **gs** = $1 - 2 * ((x1-0.05) / 0.025)^2$

$x1 < 0.075$, **gs** = $2 * ((x1-0.075) / 0.025)^2$

]; (* Calcular grado de pertenencia de x al conjunto borroso significativo *)

Which[

$x1 \leq 0.05$, **ns** = 0,

$x1 \leq 0.075$, **ns** = 1,

$x1 < 0.0625$, **ns** = $2 * ((x1-0.05) / 0.025)^2$,

$x1 < 0.075$, **ns** = $1-2 * ((x1-0.075) / 0.025)^2$

]; (* Calcular grado de pertenencia de x al conjunto borroso no significativo *)

DF1 = If[**gs** ≥ **ns**, "Si", "No"]; (* Calcula definitivamente el conjunto al cual pertenece *)

Return[**DF1**];

]

2.3 Metodología para la aplicación de los métodos Scan en la detección de conglomerados de enfermos

Como parte de este trabajo, se decidió formalizar un conjunto de pasos que sirven de guía a los epidemiólogos para la correcta aplicación de los métodos Scan en la detección de conglomerados de enfermos. A continuación se describen y se comentan cada uno de ellos:

1. Recopilación de datos (selección de las enfermedades, afecciones, trastornos de la conducta etc. a evaluar)
2. Determinación de los valores de los parámetros del método Scan (Se recomienda que sean varios valores)
3. Aplicar el método Scan clásico. Si los resultados coinciden para todos los valores de los parámetros seleccionados, concluir.
4. Si “hay dudas” (no coincidencia de los resultados para todos los valores de los parámetros seleccionados), entonces aplicar el método Scan Borroso. En base a los resultados que arroje este último método, concluir.

Para realizar el paso 1 debe consultarse las bases de datos de mortalidad y morbilidad existentes en los departamentos de estadísticas de salud. La información relacionada con la mortalidad es recogida en bases de datos informatizadas en la dirección provincial de salud, sin embargo el trabajo con la morbilidad es más complejo pues para obtener estos datos es necesario consultar las tarjetas de Enfermedades de Declaración Obligatoria que registran esta información pues la misma no está informatizada, en este caso se debe hacer un trabajo cuidadoso para evitar errores.

A pesar de que esta información puede ser obtenida en los municipios, es recomendable obtener la misma de los registros provinciales pues en ellos encontramos datos más completos y confiables, para ello es necesario contar con la autorización del Jefe del Departamento Provincial de Estadísticas.

Los datos utilizados para aplicar el método Scan son las fechas de diagnóstico, de primeros síntomas o de defunción en el caso que se estudie la mortalidad, los mismos deben ser almacenados en ficheros tipo texto.

En el paso 2, recuérdese que los parámetros a los que nos referimos son el ancho de la ventana móvil y el paso del desplazamiento. Generalmente estos valores están relacionados con la forma en la que se recoge la información: semanal, quincenal, mensual, etc., aunque los paquetes de *software* están preparados para aceptar cualquier valor. Es recomendable siempre consultar el criterio de los especialistas. Estos valores no tienen que ser los mismos para todas las enfermedades estudiadas. Debido a que la selección de estos parámetros es subjetiva y puede variar en dependencia de los criterios que tengan los epidemiólogos (no siempre se ponen de acuerdo), se recomienda probar con varias configuraciones. Es importante mencionar que los valores de los parámetros deben elegirse sin haber revisado previamente los datos.

El paso 3 se refiere concretamente a la aplicación de los métodos clásicos. Para ello resulta cómodo utilizar las facilidades del sistema EpiDet por su simplicidad y sencillez. Se recomienda aplicar las siguientes reglas:

- Si para todos los valores de los parámetros previamente seleccionados, los resultados son significativos, se concluye que existen conglomerados de enfermos. Terminar.
- Si para todos los valores de los parámetros previamente seleccionados, los resultados son no significativos, se concluye que no existen conglomerados de enfermos. Terminar.

Al paso 4 se llega si existen dudas, es decir si los resultados no coincidieron para todas las configuraciones de parámetros seleccionadas. En estos casos se debe aplicar el método Scan Borroso. Recuérdese que este método tiene un parámetro adicional: la longitud de la parte borrosa de la ventana móvil. Se recomienda que este valor no sea demasiado grande.

Al aplicar el método Borroso los resultados pueden seguir discrepando unos con otros. En este paso es crucial realizar el análisis con los especialistas. Sólo una opinión conjunta de los resultados estadísticos unido a los criterios de epidemiólogos será definitiva.

2.4 Consideraciones finales

Los métodos Scan resuelven el problema de la detección de conglomerados en secuencias de fechas de diagnóstico o de primeros síntomas. Las variantes borrosas de estas técnicas ofrecen mejores resultados que las clásicas.

Se muestran los fragmentos más importantes de los programas elaborados en el paquete Mathematica.

El capítulo culmina con la presentación de una metodología para la utilización de estos métodos por especialistas de salud.

CAPÍTULO III. APLICACIONES

En este capítulo se describen brevemente las enfermedades seleccionadas y se muestran los resultados obtenidos con la aplicación de los métodos Scan. En todos los casos se hizo una discusión con los especialistas en Higiene y Epidemiología de las posibles causas que determinaron la existencia de conglomerados de enfermos.

3.1 Recopilación de los datos

Los datos utilizados para la realización de este trabajo fueron obtenidos de las bases de datos de mortalidad y morbilidad de la dirección Provincial de Salud en Villa Clara, de ellas se tomaron los correspondientes al municipio de Cifuentes. Se seleccionó la mortalidad por Tumores malignos, las Enfermedades Cerebrovasculares, las Enfermedades del Corazón, los Suicidios y los Accidentes, que constituyen las cinco primeras causas de muerte en el territorio. Se utilizaron además, los datos de morbilidad por Hepatitis A, Meningoencefalitis Viral e Intentos Suicidas por ser las enfermedades transmisibles que incrementaron notablemente su incidencia en los últimos diez años.

Los Suicidios e Intentos Suicidas no son enfermedades como tal, sino que se definen como “trastornos de la conducta”. En cualquier caso, resultó muy interesante para los médicos especialistas que participaron en esta investigación su inclusión en el estudio. En lo adelante se utilizará el término “enfermedades” de una forma general, para referirse también a ellos, sin que eso afecte la claridad del objetivo que se pretende en este capítulo.

En el caso de la morbilidad se realizó un trabajo mucho más intenso pues estos datos no están informatizados, sólo se encuentran archivadas sus tarjetas de Enfermedades de Declaración Obligatoria (EDO). Esta información fue recopilada y tabulada utilizando las facilidades del tabulador electrónico Microsoft Excel. Posteriormente se transformaron estos ficheros a ficheros tipo texto que fueran compatibles con el sistema EpiDet.

Se utilizó el método Scan sobre una línea para la detección de conglomerados temporales de enfermos en esas enfermedades en los períodos de estudio. Se realizaron modificaciones en los parámetros para la búsqueda de valores de significación,

auxiliándonos del criterio de especialistas en Higiene y Epidemiología de la Unidad de Higiene de Cifuentes. En todos los casos se aplicó el método sobre la totalidad de casos recopilados y posteriormente de forma independiente para cada uno de los sexos por ser este factor de interés para los epidemiólogos.

3.1.1 Procedimiento de recolección de datos sobre el Mathematica

El procedimiento ScanInicio recibe una lista con las fechas en cualquier orden y cuenta los enfermos de cada día.

```
ScanInicio[Camino_, Fact_, AnchoW_, Paso_, Incr_] :=
CompoundExpression[
  Datos = ReadList[Camino, Number, RecordLists → True];
  (* Lee la fecha de inicio de la enfermedad y factor de riesgo de cada paciente *)
  Datos1 = {}; (* Crea una lista vacía *)
  If[Fact != -1, Map[Function[i, If[i[[4]] == Fact, Datos1 = Append[Datos1,i]], Datos],
    Datos11=Datos]; (* Duplica los datos *)
  If[Fact!= -1,Datos11=Map[Function[x,Delete[x,4]], Datos1]]; (* Datos11 solo
  fecha *)
  Datos1 = Map[Reverse,Datos11]; (* Datos1→fecha con formato año, mes y día *)
  Datos = Sort[Datos1]; (* Datos posee fechas ordenadas *)
  VI = Map[Function[i, DaysBetween[Datos[[i-1]], Datos[[i]] ], Range[2, Length[Datos]]];
  VI = Insert[VI, 1, 1];
  (* Lista de la diferencia en día entre la aparición de dos enfermos consecutivos *)
  AuxVI =
  Split[
    Flatten[{
      Table[If[VI[[i]]> 0, Flatten[{Table[0,{j,VI[[i]]-1}], If[VI[[i+1]]>0,1,a}],a},{i,Length[VI[[ ]]]-
      1}], If[VI[[Length[VI[[ ]]]]>0,Flatten[{Table[0,{j,VI[[Length[VI[[ ]]]]-1}],1}],a]
    }
  ];
  (* Lista de 0, 1 y a pacientes x días *)
  (* # Consecutivos de a indica # de enfermos en un día *)
  Valor = Flatten[Table[If[AuxVI[[i,1]]===a ,Length[AuxVI[[i]],AuxVI[[i]],{i,Length[AuxVI[[ ]]]}];
  (* Lista de # pacientes enfermos por días *)
  bbb=ScanValidation[Valor,AnchoW,Paso,Incr]
  (* Buscar si hay clusters de enfermos *)
]
```

Los epígrafes que aparecen a continuación describen los resultados obtenidos de la aplicación del método Scan Clásico. En dependencia de las conclusiones que se extrajeron, se formaron tres grupos, que se corresponden con los tres epígrafes siguientes:

- **Resultados no significativos:** aquí se resumen la información de aquellas enfermedades en las que no se demostró la presencia de conglomerados.
- **Resultados significativos para todos los valores de los parámetros:** este epígrafe resume la información de aquellas enfermedades en las que se demostró la presencia de conglomerados para todos los valores de los parámetros considerados.
- **Resultados significativos para algunos valores de los parámetros:** en esta sección se resume la información de aquellas enfermedades en las que el método Scan arrojó dudas. Para algunos valores de los parámetros los resultados fueron significativos mientras que para otros no. En los valores dudosos, se decidió aplicar además el método Scan Borroso para llegar a conclusiones más certeras.

3.2 Resultados no significativos

Los valores de significación obtenidos por ambos métodos, demuestran la no existencia de conglomerados en las Enfermedades Cerebrovasculares, Accidentes y Suicidios en el período comprendido entre 1997-2007. Esto puede ser corroborado al analizar los gráficos correspondientes.

3.2.1 Enfermedades Cerebrovasculares

Se entiende por Enfermedad Cerebrovascular (ECV) aquellos trastornos en los que hay un área cerebral afectada de forma transitoria o permanente por isquemia o hemorragia y/o cuando uno o más vasos sanguíneos cerebrales están afectados por un proceso patológico, (Rodríguez Rueda 2006). Las Enfermedades Cerebrovasculares constituyen uno de los problemas de salud más importantes en todos los países desarrollados y en nuestro país, (MINSAP 2000).

La inactividad física, el hábito de fumar y la ingestión de bebidas alcohólicas son los factores de riesgo más frecuentes. La Hipertensión Arterial, la Diabetes Mellitus y la Cardiopatía Isquémica constituyen las patologías asociadas más frecuentes, (Expósito Carralero 2004).

En Cuba, y los países desarrollados constituyen la tercera causa de muerte y la primera de discapacidad, observándose en las últimas tres décadas un ascenso de las mismas en tasas brutas, detectándose una tendencia al desplazamiento de la muerte por estas causas hacia edades más tempranas de la vida, (Díaz Chalala 2008), superadas sólo por las enfermedades cardiovasculares y el cáncer; por otra parte, entre el 50 y el 70 % de los casos quedan con secuelas, y aproximadamente el 75% pierde sus facultades para reincorporarse al trabajo, considerándose la enfermedad neurológica invalidante más prevalente de la población adulta mayor de 65 años, (Valdés Gutiérrez 2006).

La tabla 3.1 muestra los resultados de la aplicación del método Scan usando el sistema EpiDet y el paquete Mathematica, (ambos resultados coinciden). Según el criterio de especialistas en el tema, se consideraron tres valores para el ancho de la ventana móvil: 60 días (dos meses), 30 días y 15. Los valores del parámetro “paso” también pueden observarse en la tabla. Ellos fueron de siete días (una semana), quince días (aproximadamente dos semanas) y de 30 (un mes). Debe destacarse que el valor del paso para el desplazamiento debe ser siempre menor que el tamaño de la ventana móvil, pues de lo contrario quedarían casos sin analizar. El sistema *EpiDet* controla que el valor del paso esté entre 1 y la longitud de la ventana. Todas las enfermedades consideradas se analizaron de esa misma manera.

La tabla 3.1 muestra el resultado del análisis general y después desglosado por sexo. Puede observarse que ningún caso evidencia la presencia de conglomerados.

Tabla 3.1 Resultados obtenidos con los métodos Scan para las Enfermedades Cerebrovasculares.

Vent. M.	Paso	General		Factores (Sexo)			
		Scan		Masculino		Femenino	
		Est	p.	Scan		Scan	
				Est.	p.	Est.	p.
60	30	11	0.7445	7	0.7301	7	0.8362
	15	12	0.4173	7	0.7239	7	0.8343
	7	13	0.1870	7	0.7258	7	0.8380
30	30	9	0.2494	5	0.8503	4	0.9998
	15	9	0.2469	5	0.8463	5	0.9061
	7	9	0.2350	5	0.8324	5	0.9226
15	15	6	0.6172	3	0.9999	3	0.9999
	7	6	0.6121	3	0.9999	3	0.9999

La figura 3.1 muestra el gráfico de la serie de tiempo asociada a la enfermedad. Puede apreciarse que en ningún lugar hay picos que sobrepasen en demasía al resto de los valores. Este hecho corrobora los resultados estadísticos anteriormente discutidos.

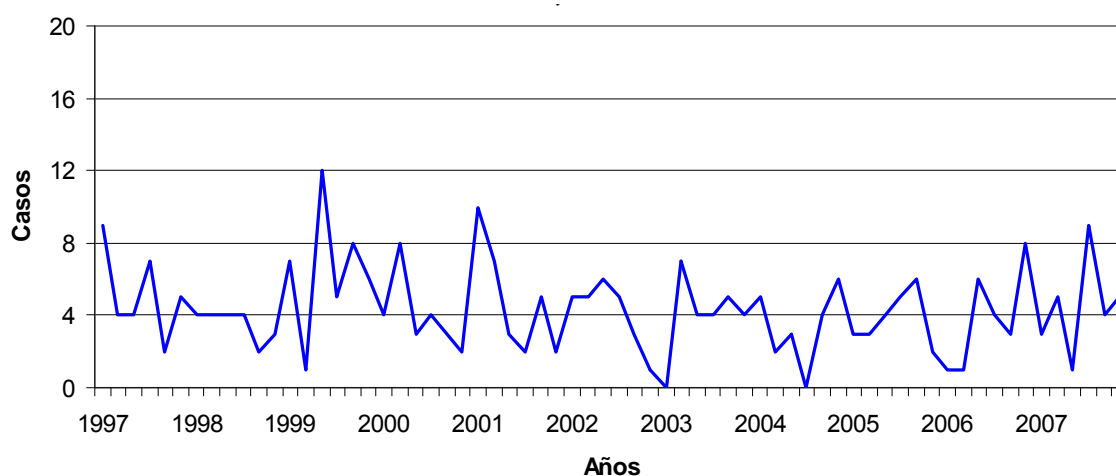


Figura 3.1 Distribución de la mortalidad por ECV en Cifuentes en el período 1997-2007.

3.2.2 Accidentes

La Organización Mundial de la Salud (OMS) considera accidente un acontecimiento fortuito, generalmente desgraciado o dañino, independientemente de la voluntad humana, provocado por una fuerza exterior que actúa rápidamente y que se manifiesta por la aparición de lesiones orgánicas o trastornos mentales.

Los accidentes se ubican entre las cinco primeras causas de muerte en niños y ancianos de 20 países de América Latina y el Caribe, incluyendo a Cuba. Su importancia está dada por la elevada morbilidad e invalidez que causan al paciente, al gran número de crisis familiares no transitorias que originan y al elevado costo económico que directa o indirectamente representan, (Álvarez Sintés 2001).

Constituyen un serio problema no solo de salud, sino también económico y social, esto permite considerarlos como un problema prioritario, y se pueden calificar como una enfermedad social de proporciones epidémicas, (Mastrapa Reyes 2008).

Tabla 3.2 Resultados obtenidos con los métodos Scan para los Accidentes.

Vent. M.	Paso	General		Factores (Sexo)			
		Scan		Masculino		Femenino	
		Est	p.	Scan		Scan	
				Est.	p.	Est.	p.
60	30	7	0.6401	6	0.5207	4	0.5278
	15	7	0.6365	6	0.5136	4	0.5024
	7	8	0.2489	6	0.5034	4	0.4949
30	30	4	0.9977	4	0.8864	3	0.7443
	15	5	0.7732	4	0.8806	3	0.7042
	7	5	0.7503	4	0.8525	3	0.6681
15	15	4	0.7382	2	1.0000	3	0.3433
	7	5	0.1405	3	0.9557	3	0.2750

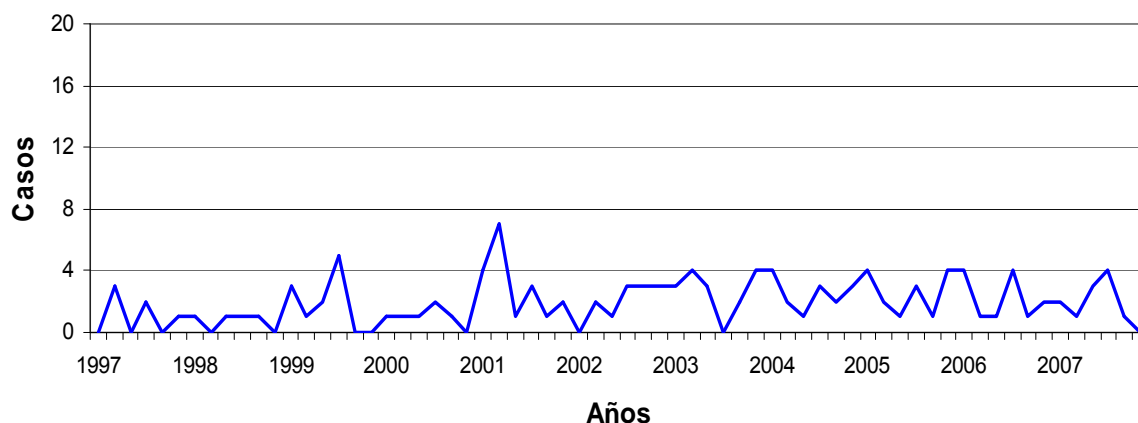


Figura 3.2 Distribución de la mortalidad por Accidentes en Cifuentes en el período 1997 - 2007.

Como puede apreciarse en el caso de la mortalidad por Accidentes el resultado fue el mismo que al analizar la mortalidad por ECV. El gráfico 3.2 igualmente ilustra lo que se demostró estadísticamente en la tabla 3.2

3.2.3 Suicidios

El Suicidio es el acto humano de autolesionarse con la intención de acabar con la propia existencia e implica una variedad de estados y motivaciones que, a veces, pueden ser conscientes o inconscientes, (Guibert Reyes 2003).

Este fenómeno va en aumento en todos los países según datos de la OMS, sobre todo los países de alto desarrollo tecnológico, donde las causas más importantes de muerte son las llamadas "enfermedades del desarrollo" (Cardiovasculares, Cerebrovasculares, Accidentes, etc.) con incremento del suicidio en un 15 % en dichos países, convirtiéndose en segunda causa la muerte entre individuos menores de 35 años, (Martínez Jiménez 2008).

En Cuba el suicidio ocupa la cuarta causa de mortalidad general y la segunda en las edades entre 15 - 49 años, (Torres Lio-Coo 2006; Díaz Chalala 2007). El índice de suicidio es también elevado en personas de más de 70 años, con el máximo en varones solteros. Algunos de los factores de riesgo son la falta de empleo, el aislamiento, la mala salud, el dolor, los sentimientos de rechazo, los antecedentes de enfermedades mentales y los

intentos previos de suicidios. El antecedente de alcoholismo es un factor mórbido concomitante elevado, (Rodríguez López 2003).

Tabla 3.3 Resultados obtenidos con los métodos Scan para los Suicidios.

Vent. M.	Paso	General		Factores (Sexo)			
		Scan		Masculino		Femenino	
		Est	p.	Scan		Scan	
				Est.	p.	Est.	p.
60	30	6	0.2497	5	0.3607	2	0.9493
	15	6	0.2466	5	0.3600	2	0.9493
	7	6	0.2461	5	0.3625	2	0.9375
30	30	3	0.9966	3	0.9496	1	1.0000
	15	3	0.9965	3	0.9477	1	1.0000
	7	3	0.9966	3	0.9517	1	1.0000
15	15	3	0.8633	3	0.6232	1	1.0000
	7	3	0.8581	3	0.6169	1	1.0000

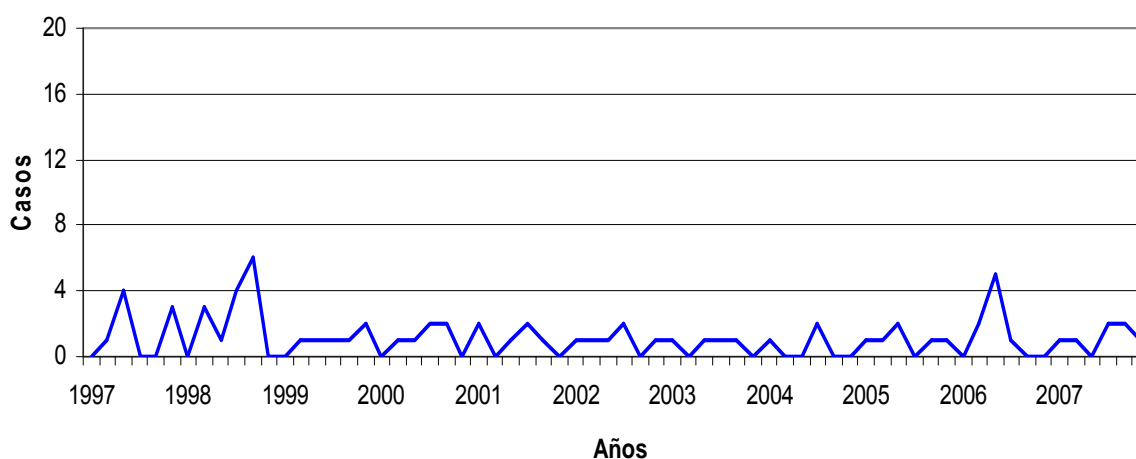


Figura 3.3 Distribución de la mortalidad por Suicidios en Cifuentes en el período 1997-2007.

3.3 Resultados significativos para todos los parámetros

Analizamos los datos de morbilidad por Meningoencefalitis Viral y Hepatitis A correspondientes al municipio de Cifuentes en los años 1999 - 2008 y se obtuvieron valores significativos para todos los parámetros lo cual nos indica la existencia de conglomerados de enfermos para ambas patologías.

3.3.1 Meningoencefalitis Viral

Es un evento infeccioso producido por un número importante de virus, que dependiendo de su tropismo, tienen la habilidad de penetrar al sistema nervioso central (SNC) y causar infección dando lugar en ocasiones a sintomatología leve y en otras a una disfunción neurológica grave, que puede causar la muerte, (Coria 2001).

Las infecciones del SNC constituyen un problema de salud en todas las regiones del mundo y representan un reto para el médico de asistencia, el epidemiólogo y el microbiólogo en relación al manejo adecuado del paciente. La población infantil es la más afectada. Las meningoencefalitis virales tienen la característica de sus cuadros clínicos benignos, no obstante se realiza vigilancia epidemiológica a fin de conocer su comportamiento, (Albert Cabrera 2007) .

Las manifestaciones clínicas dependen de la localización anatómica o región afectada del SNC, la virulencia del agente, la respuesta del huésped y la edad del niño. El síndrome encefalítico aparece habitualmente con rapidez y hasta con brusquedad, aunque existen casos de instauración lenta y progresiva. La fiebre no falta prácticamente nunca, la intranquilidad, llanto de tono agudo y agitación son frecuentes, así como el cambio de conducta, alteraciones de la conciencia, convulsiones, y también las parálisis, (Guirola León 2009).

Tabla 3.4 Resultados obtenidos con los métodos Scan para la Meningoencefalitis Viral.

Vent. M.	Paso	General		Factores (Sexo)			
		Scan		Masculino		Femenino	
		Est	p.	Scan		Scan	
				Est.	p.	Est.	p.
60	30	31	0.0000	17	0.0000	17	0.0000
	15	31	0.0000	17	0.0000	17	0.0000
	7	31	0.0000	17	0.0000	17	0.0000
30	30	18	0.0000	9	0.0019	9	0.0000
	15	25	0.0000	13	0.0000	13	0.0000
	7	27	0.0000	13	0.0000	13	0.0000
15	15	15	0.0000	8	0.0002	8	0.0002
	7	18	0.0000	8	0.0002	8	0.0002

La tabla 3.4 muestra resultados altamente significativos para todos los valores de los parámetros. Al segmentar la muestra analizada por sexo, este resultado se mantiene lo que indica que existen conglomerados de enfermos en ambos sexos.

Las figuras 3.4 y 3.5 muestran las series de tiempo asociadas que evidencian la existencia de un pico en los años 2000 - 2001. Los especialistas desconocen las causas que pudieron motivar el aumento de la incidencia de la enfermedad en este período pues la misma no se asocia a ningún factor de riesgo.

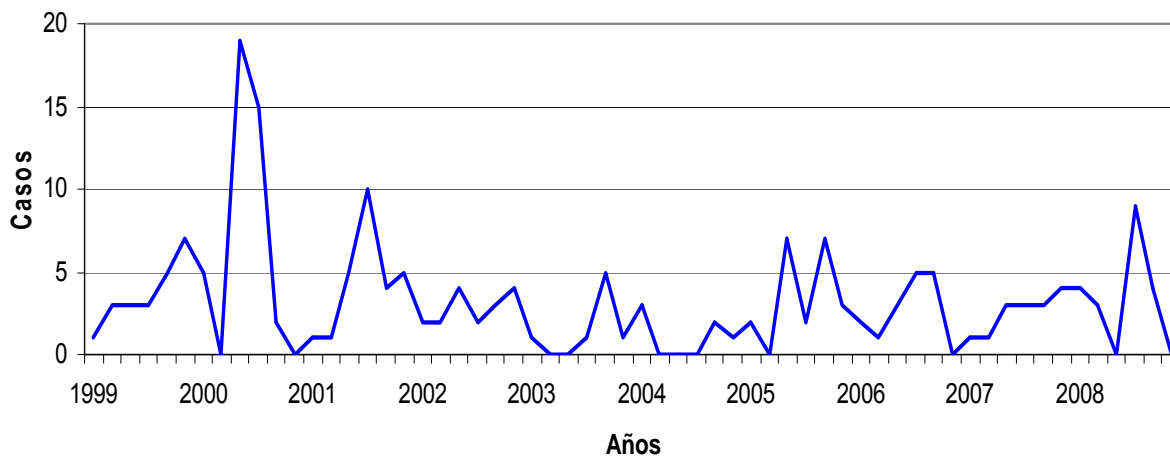


Figura 3.4 Distribución de la morbilidad por Meningoencefalitis Viral en Cifuentes en el período 1999 – 2008.

La figura 3.5 Muestra el mismo gráfico anterior, pero desglosado por sexo. Como dato curioso obsérvese que, a pesar de ser ambos significativos, se reportan más casos de enfermos por Meningoencefalitis Viral en hombres que en mujeres.

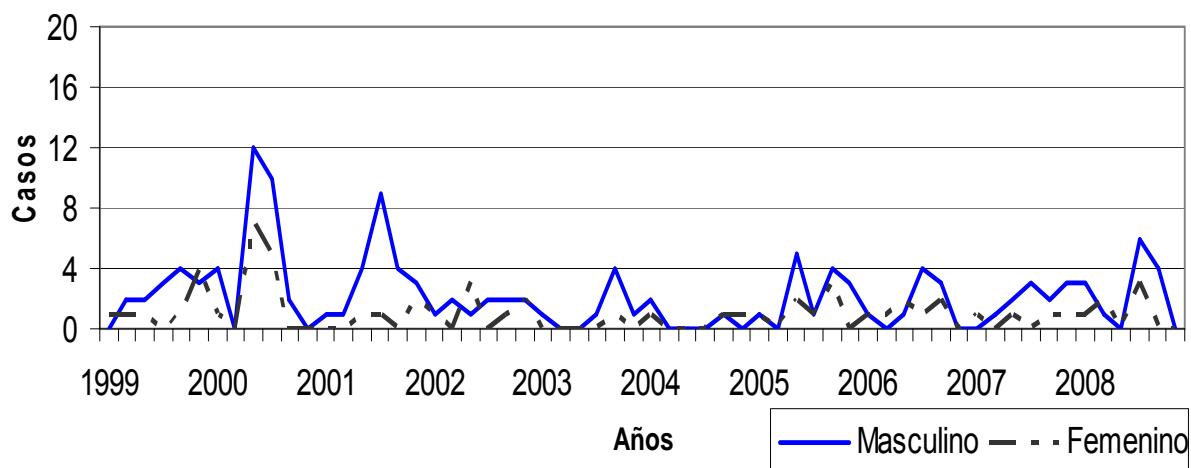


Figura 3.5 Distribución de la morbilidad por Meningoencefalitis Viral, según sexo, en Cifuentes en el período 1999 – 2008.

3.3.2 Hepatitis A

La Hepatitis A es una enfermedad caracterizada por la inflamación aguda del hígado, como consecuencia de la infección por el virus de la Hepatitis A. En general puede considerarse una enfermedad benigna - fundamentalmente cuando ocurre en niños - autolimitada, que cura sin dejar secuelas, y es la más frecuente entre los distintos tipos de hepatitis, (Más 1999).

El virus de la Hepatitis A se extiende principalmente por contacto fecal oral y, por tanto, se pueden producir en regiones con mala higiene. Se producen epidemias transmitidas por el agua o los alimentos, especialmente en países subdesarrollados. Los casos esporádicos también son frecuentes, generalmente como resultado de un contacto directo entre personas, (Beers 2007)

Las manifestaciones clínicas de la Hepatitis A tiene una marcada tendencia a variar con la edad del paciente. Mientras que en los niños menores de 2 años son muy frecuentes las infecciones inaparentes, las manifestaciones clínicas de hepatitis son más marcadas cuanto mayor es la edad del paciente. En el adulto, desde el punto de vista clínico, se destacan síntomas como, prurito, fiebre, diarrea y pérdida de peso, (Arias 1999).

Tabla 3.5 Resultados obtenidos con los métodos Scan para la Hepatitis A.

Vent. M.	Paso	General		Factores (Sexo)			
		Scan		Masculino		Femenino	
		Est	p.	Scan		Scan	
				Est.	p.	Est.	p.
60	30	53	0.0000	27	0.0000	25	0.0000
	15	53	0.0000	28	0.0000	25	0.0000
	7	54	0.0000	28	0.0000	25	0.0000
30	30	40	0.0000	15	0.0000	14	0.0000
	15	40	0.0000	16	0.0000	15	0.0000
	7	40	0.0000	19	0.0000	16	0.0000
15	15	23	0.0000	10	0.0000	8	0.0000
	7	24	0.0000	12	0.0000	10	0.0000

La tabla anterior muestra resultados altamente significativos para todos los valores de los parámetros analizados. Al segmentar la muestra por sexo, el resultado se mantiene lo que indica que existen conglomerados de enfermos de Hepatitis A en hombres y en mujeres.

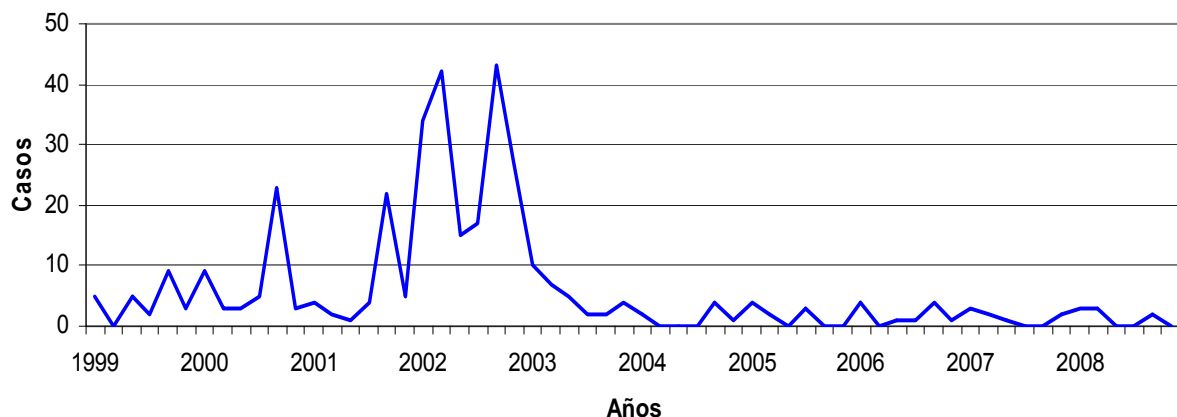


Figura 3.6 Distribución de la morbilidad por Hepatitis A en Cifuentes en el período 1999 – 2008.

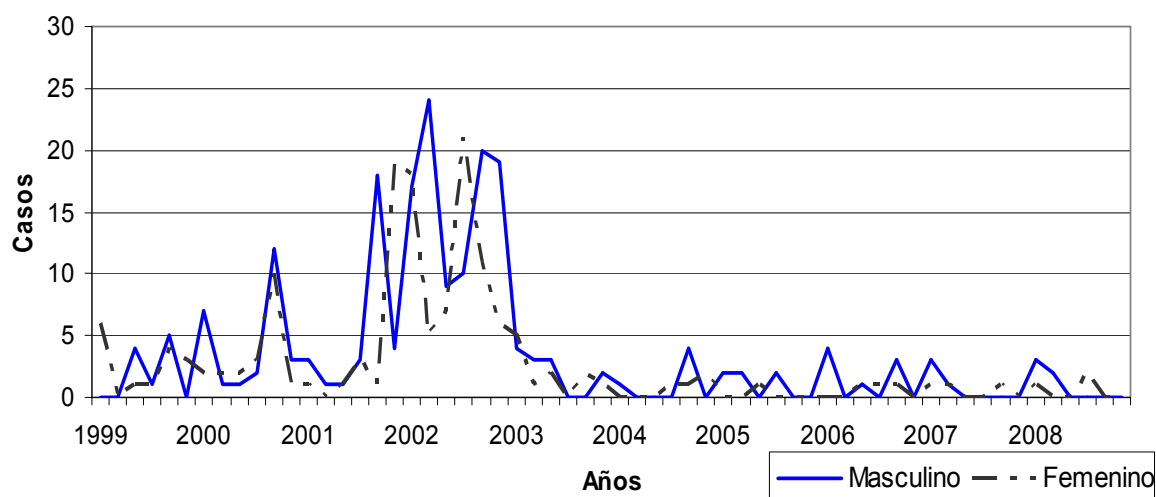


Figura 3.7 Distribución de la morbilidad por Hepatitis A, según sexo, en Cifuentes en el período 1999 – 2008.

Las figuras 3.6 y 3.7 Muestran el comportamiento de la incidencia de Hepatitis, general y desglosado por sexo.

En el período comprendido entre el 2001 y el 2003 se produjo un notable incremento de los casos de Hepatitis A en el municipio de Cifuentes. Esta enfermedad no depende del

sexo, por lo que no se aprecian diferencias en el número de casos reportados según esta variable.

Según estudios realizados en la Unidad de Higiene y Epidemiología del municipio, las mayores tasas de morbilidad por esta causa se presentaron en San Diego del Valle. Esta área de salud ha tenido una situación desfavorable en su red hidráulica, la cual se encontraba en mal estado y era muy común la contaminación del agua de consumo, la calidad del agua que reciben estos pobladores es también mala porque no se clora al no existir equipo clorador en la fuente de abasto, además las normas de higiene individual y colectiva no se cumplían adecuadamente. Todo esto propició que al ser la hepatitis una enfermedad transmisible se produjera un aumento del número de casos en un breve periodo de tiempo, por esta razón, por parte de la Vicedirección de Higiene y Epidemiología del área se incrementaron las acciones sanitarias en el Consejo Popular, lo que ha logrado disminuir la incidencia de esta enfermedad.

3.4 Resultados significativos para algunos parámetros

En este epígrafe se incluyó el análisis de la mortalidad por Enfermedades del Corazón y Tumores Malignos y el análisis de la morbilidad por Intentos Suicidas.

A diferencia de los otros procesamientos realizados, los métodos Scan aquí no fueron concluyentes. En los casos en los que hubo dudas, se aplicó la variante borrosa del método Scan para llegar a conclusiones.

3.4.1 Enfermedades del Corazón

Las Enfermedades del Corazón son la primera causa de muerte en Cuba. Son producidas por un desbalance entre la oferta y la demanda de oxígeno al miocardio, debido a lesiones orgánicas (aterosclerosis) o funcionales (espasmo) y que provocan varios cuadros, desde fenómenos asintomáticos (isquemia silente, disfunción diastólica) hasta cuadros de necrosis miocárdica extensa, (Penichet Montoto 2007).

La forma clínica más grave de estas enfermedades es el Infarto Agudo del Miocardio (IMA), esta entidad se caracteriza por un fuerte dolor precordial o retroesternal, que puede irradiarse a la axila, ambos brazos o el izquierdo, o al cuello, acompañado de sudoración

profusa, vómitos y mareos. El dolor habitualmente dura más de 10 minutos y requiere, con frecuencia, el uso de opiáceos para su alivio, (Toledo Curbelo 2007).

Tabla 3.6 Resultados obtenidos con los métodos Scan para las Enfermedades del Corazón.

Vent. M.	Paso	General				Factores (Sexo)							
		Scan		R. Borroso		Masculino				Femenino			
		Est.	p.	Inc	Sigf	Scan		R. Borroso		Scan		R. Borroso	
						Est.	p.	Inc	Sigf	Est.	p.	Inc	Sigf
60	30	32	0.0004			18	0.0587	1	Si	14	0.2890		
	15	32	0.0004			18	0.0600	1	Si	14	0.2936		
	7	32	0.0004			18	0.0580	1	Si	14	0.2979		
30	30	16	0.2456			11	0.3015			10	0.2335		
	15	20	0.0036			13	0.0289			10	0.2313		
	7	24	0.0000			14	0.0064			12	0.0163		
15	15	12	0.0835	2	Si	9	0.0684	1	Si	7	0.3850		
	7	14	0.0047			9	0.0631	4	Si	8	0.0872	4	No

Como puede observarse en el análisis general de todos los casos que se muestra en la tabla 3.6, la técnica Scan brinda resultados concluyentes acerca de la presencia de conglomerados para casi todos los valores de los parámetros considerados. Nótese que al analizar un ancho de ventana y un paso de 15 días, los resultados fueron medianamente significativos (0.0835). Este valor para el ancho de la ventana puede considerarse pequeño. En (Rodríguez 2009) se demuestra que para tamaños de ventanas pequeños la técnica Scan Borrosa produce mejores resultados que la clásica. Es por ello que a la tabla se le agrega otra columna en la que se pone el resultado de su aplicación. El método Scan Borroso demuestra la existencia de conglomerados para esos valores de los parámetros.

La figura 3.8 muestra una representación gráfica de los datos procesados. Pueden apreciarse picos con una incidencia más elevada de la enfermedad alrededor de los años 1998 -1999 y 2005 - 2006.

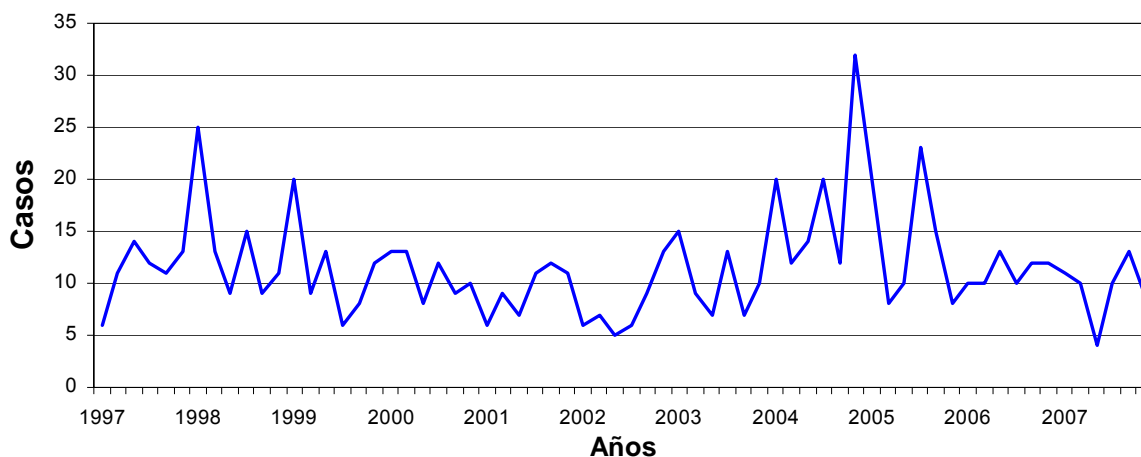


Figura 3.8 Distribución de la mortalidad por Enfermedades del Corazón en Cifuentes en el período 1997 – 2007.

En los años 1998 y 1999, se incrementó la mortalidad por Enfermedades del Corazón en el municipio, según los especialistas en Higiene y Epidemiología del municipio este aumento coincide con la etapa del período especial, donde se modificaron los estilos de vida de la población por la difícil situación económica que existió en el país durante esa fecha, se incrementó el consumo de grasa de origen animal, disminuyó la realización de ejercicios físicos, y aumentó el estrés, todo esto condujo a un aumento de la incidencia de hipertensión arterial, que constituyen los principales factores de riesgo de las Enfermedades del Corazón.

El mayor número de fallecidos por enfermedades del corazón se produjo en los años 2005 y 2006, debemos tener en cuenta que la edad es uno de los principales factores de riesgo de estas patologías y nuestra provincia y en particular nuestro municipio presenta una de las poblaciones más envejecidas del país, el grupo de edad de 65 años y más representa el 21 % de la población total de estos años. Además se incrementaron los hábitos tóxicos como el consumo de café, tabaco y alcohol fundamentalmente en la población masculina, existe un mal seguimiento en consulta de la hipertensión arterial y hay una tendencia al abandono del tratamiento por parte de los pacientes, todo esto pudo contribuir al incremento de la mortalidad por esta causa.

En la tabla 3.6 se hace también el análisis separado para ambos sexos. Se sigue la misma metodología: en los casos en los que el Scan clásico no brinda resultados concluyentes,

se aplica el método Scan borroso. De manera descriptiva puede apreciarse que esta enfermedad atacó más al sexo masculino que al femenino en el municipio de Cifuentes. El gráfico de las incidencias por sexo que aparece en la figura 3.9 así lo demuestra.

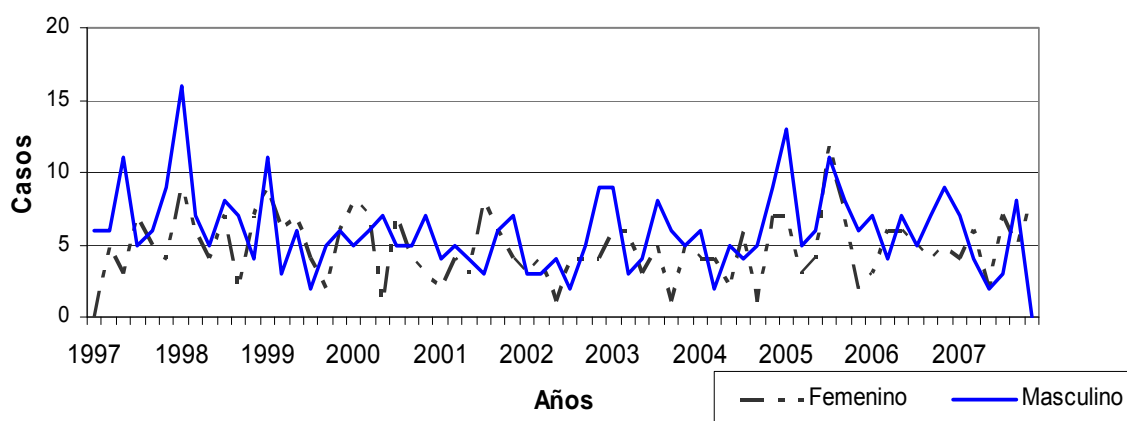


Figura 3.9 Distribución de la mortalidad por Enfermedades del Corazón, según sexo, en Cifuentes en el período 1997 – 2007.

3.4.2 Tumores Malignos

El Cáncer es una proliferación celular desordenada debido a la pérdida de los controles normales, que da lugar a crecimiento desordenado, ausencia de diferenciación, invasión tisular local y, con frecuencia, metástasis. El Cáncer puede aparecer en cualquier tejido u órgano, a cualquier edad. Con frecuencia hay una respuesta inmunitaria frente a los tumores.

Las neoplasias malignas pueden producir dolor, adelgazamiento, neuropatías, náuseas, anorexia, convulsiones, hipercalcemia, hiperuricemia y obstrucción. La muerte se produce típicamente como consecuencia de la insuficiencia súbita o progresiva de uno o más sistemas de órganos, (Beers 2007).

Se estima que el Cáncer provoca la pérdida de más de cuatro millones de personas anualmente, lo que lo convierte en la tercera causa de muerte a escala mundial. Constituye un problema de salud especialmente relevante en los países desarrollados, en los cuales se ha logrado controlar otras causas de muerte, pero esto se ha convertido en un verdadero flagelo. En Europa, uno de cada cuatro ciudadanos muere por esta causa;

en España se ha convertido en la segunda causa de muerte, y se conoce que un elevado porcentaje de tales pacientes (50-90 %) padece dolor. En Cuba, constituye la segunda causa de muerte desde 1958, y de años de vida potencialmente perdidos. Se estima que al iniciarse el siglo XXI haya superado a la cardiopatía, (Lovellette Jorge 2007)

La tabla 3.7 muestra los resultados de la aplicación de los métodos Scan para la detección de conglomerados a los casos de mortalidad por Tumores Malignos en el período comprendido entre los años 1997 y 2007. Puede observarse la presencia de conglomerados.

Tabla 3.7 Resultados obtenidos con los métodos Scan para la mortalidad por Tumores Malignos.

Vent. M.	Paso	General				Factores (Sexo)							
		Scan		R. Borroso		Masculino				Femenino			
		Est.	p.	Inc	Sigf	Scan		R. Borroso		Scan		R. Borroso	
						Est.	p.	Inc	Sigf	Est.	p.	Inc	Sigf
60	30	27	0.0012			16	0.0928	4	Si	13	0.0616	1	Si
	15	28	0.0004			18	0.0113			13	0.0632	4	No
	7	28	0.0004			18	0.0115			13	0.0622	1	Si
30	30	15	0.0916	4	Si	9	0.7482			8	0.3651		
	15	15	0.0916	4	No	12	0.0320			8	0.3681		
	7	16	0.0318			12	0.0336			8	0.3665		
15	15	9	0.5936			7	0.5145			5	0.9188		
	7	9	0.6320			7	0.5525			5	0.9212		

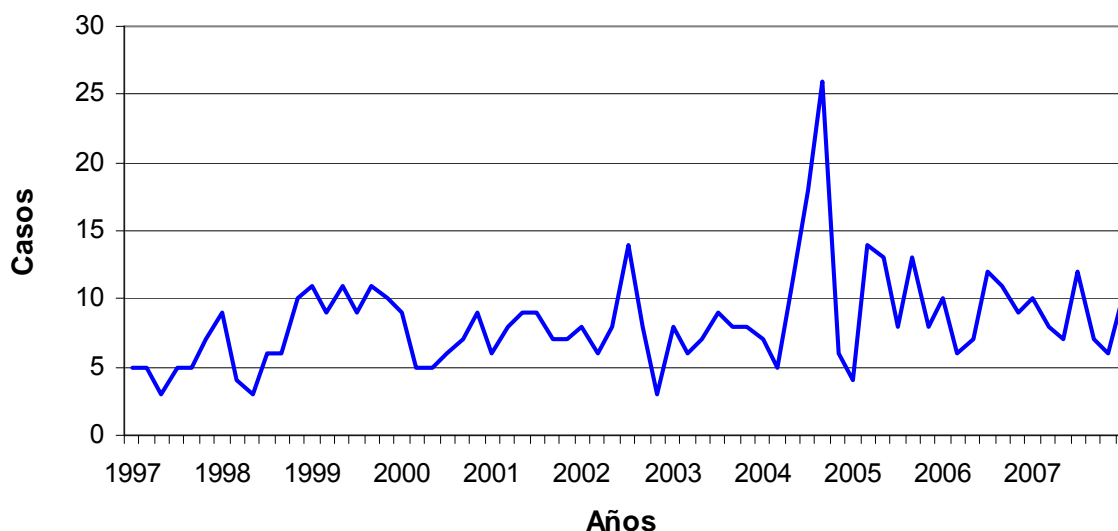


Figura 3.10 Distribución de la mortalidad por Tumores Malignos en Cifuentes en el período 1997 – 2007.

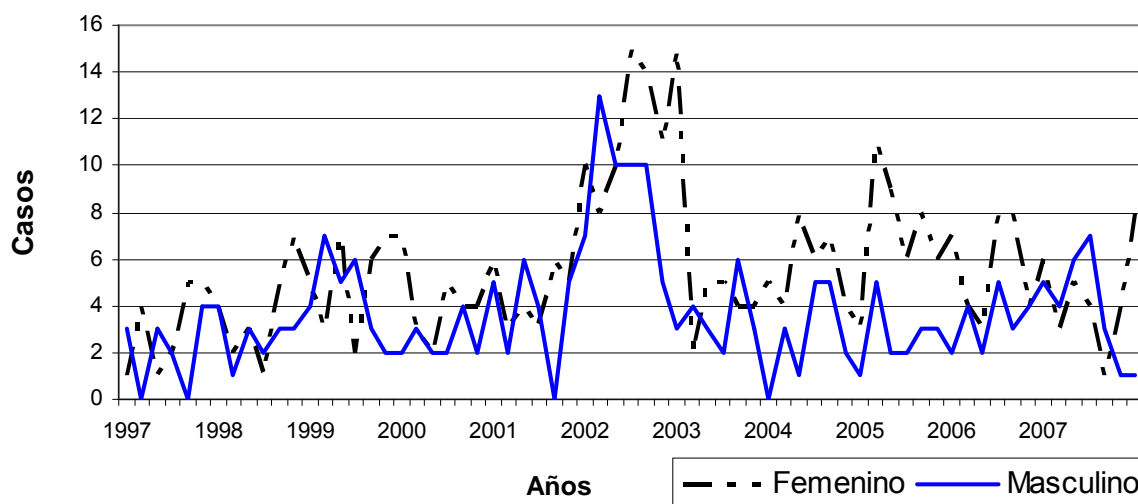


Figura 3.11 Distribución de la mortalidad por Tumores Malignos, según sexo, en Cifuentes en el período 1997 – 2007.

Según el criterio de especialistas del municipio, la mortalidad por Tumores Malignos aumentó durante los años, 2002, 2003 y 2005. Esta es una enfermedad de etiología desconocida, cuya aparición se asocia a factores de riesgo. Según estudios realizados (Beers 2007) los factores ambientales constituyen un riesgo a largo plazo en la aparición de estas enfermedades, coincidiendo con el uso de productos químicos para la

maduración de las frutas y el abuso de insecticidas en la agricultura durante el período especial, además se incrementó el consumo de café y alcohol, sobre todo de bebidas de fabricación casera con alto grado de sustancias tóxicas, y en estos años se reporta que el 23.6% de la población del municipio es fumadora.

3.4.3 Intentos Suicidas

El Intento Suicida o parasuicidio es definido por la OMS, "como un acto con una consecuencia no fatal en la cual el individuo realiza deliberadamente una conducta no habitual con amenaza de muerte, que sin la intervención de otros le causará autodaño, o ingiere una sustancia superior a las dosis terapéuticas generalmente reconocidas y cuyo objetivo es producir cambios que él o ella desean a través de las consecuencias físicas y psíquicas reales o esperadas cercanas a la muerte", (Guibert Reyes 2001).

Lo intentan más los adolescentes, en especial el sexo femenino, mediante el uso de los métodos más suaves como la ingestión de tabletas, aunque esto está relacionado con los medios disponibles a su alcance en el momento de la crisis, (Rodríguez Méndez 2006).

Los factores psicosociales de riesgo individuales que, de acuerdo con las investigaciones científicas más actuales en concepción son esenciales en la ocurrencia de un intento suicida son: presencia generalizada de sentimientos de desesperanza y culpa, presencia de depresión mayor, personas que han sobrevivido al intento suicida, personas que han llamado la atención por presagiar o amenazar con el suicidio (proyecto suicida), antecedentes familiares de suicidio o de intento, personas sin apoyo social y familiar y presencia de impulsividad o de ansiedad y hostilidad, (Guibert Reyes 2003).

Tabla 3.8 Resultados obtenidos con los métodos Scan para la morbilidad por Intentos Suicidas.

Vent. M.	Paso	General				Factores (Sexo)			
		Scan		R. Borrosos		Masculino		Femenino	
		Est.	p.	Inc	Nuevo Result	Scan		Scan	
						Est.	p.	Est.	p.
60	30	14	0.3459			5	0.8535	11	0.5995
	15	14	0.3428			5	0.8535	11	0.5931
	7	14	0.3610			5	0.8560	12	0.2951
30	30	10	0.2632			4	0.8062	8	0.5016
	15	10	0.2590			4	0.8003	8	0.4946
	7	10	0.2733			4	0.8300	8	0.5149
15	15	8	0.0982	4	No	2	1.0000	6	0.4967
	7	8	0.0972	1	Si	3	0.9651	6	0.4878

La mayoría de los resultados que aparecen en la tabla 3.8 muestran la no existencia de diferencias significativas, o lo que es lo mismo la no presencia de conglomerados de Intentos Suicidas en el municipio de Cifuentes en el período de 1997 a 2008.

Sólo dos combinaciones de parámetros ofrecen dudas, pues los resultados del valor de p son medianamente significativos. Ello ocurre al considerar la ventana móvil de tamaño 15 y el paso del desplazamiento de 15 y 7 días respectivamente. Siguiendo la metodología propuesta en esta tesis, se aplica para esos valores el método Scan Borroso. Con la primera combinación de valores, los resultados no fueron significativos por lo que se concluye la no presencia de conglomerados de Intentos Suicidas, tal y como sugiere la representación gráfica mostrada en la figura 3.12.

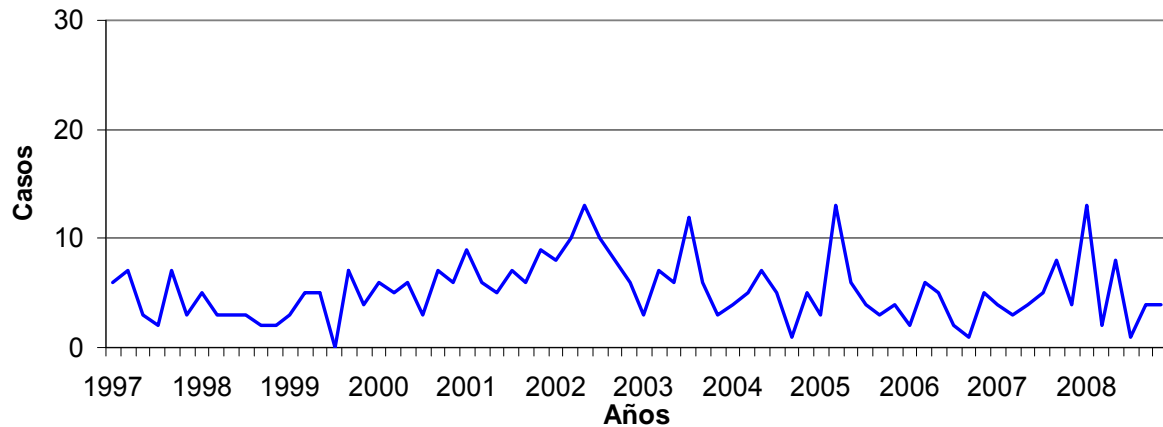


Figura 3.12 Distribución de la morbilidad por Intentos Suicidas en Cifuentes en el período 1997 – 2008.

Sin embargo, al analizar la otra combinación: tamaño de ventana 15 y paso del desplazamiento de una semana, el método Scan Borroso afirma que existe al menos un conglomerado. Se decidió entonces volver a graficar los datos de la incidencia de intentos suicidas, mostrando las cantidades de casos reportados cada 15 días. La figura 3.13 muestra el resultado gráfico:

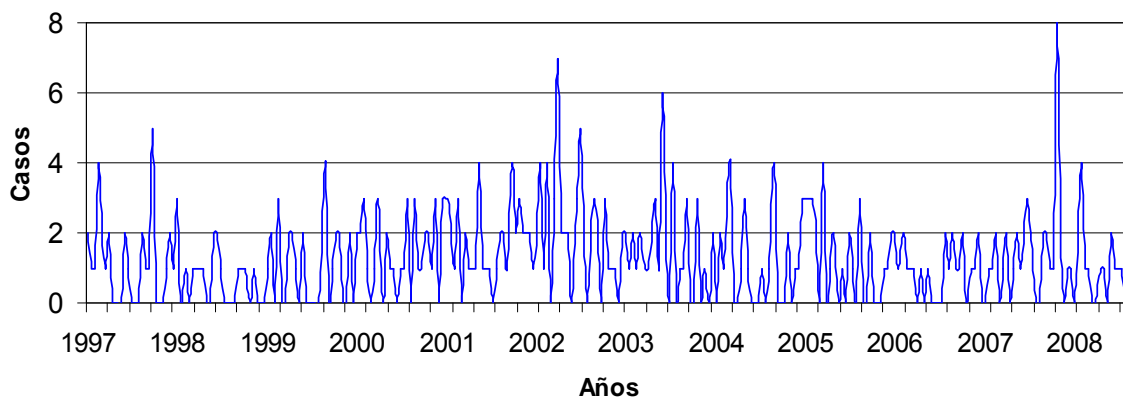


Figura 3.13 Distribución de la morbilidad quincenal por Intentos Suicidas en Cifuentes en el período 1997 – 2008.

Alrededor de los años 2002 y 2008 existen efectivamente dos picos notables. Esos son los que detecta el método borroso.

De manera general los picos no son tan elevados. Ello unido a los resultados no significativos del método Scan para las otras configuraciones de los parámetros, y a los criterios de los epidemiólogos, permitieron concluir que no existían *clusters* de enfermos en el período analizado.

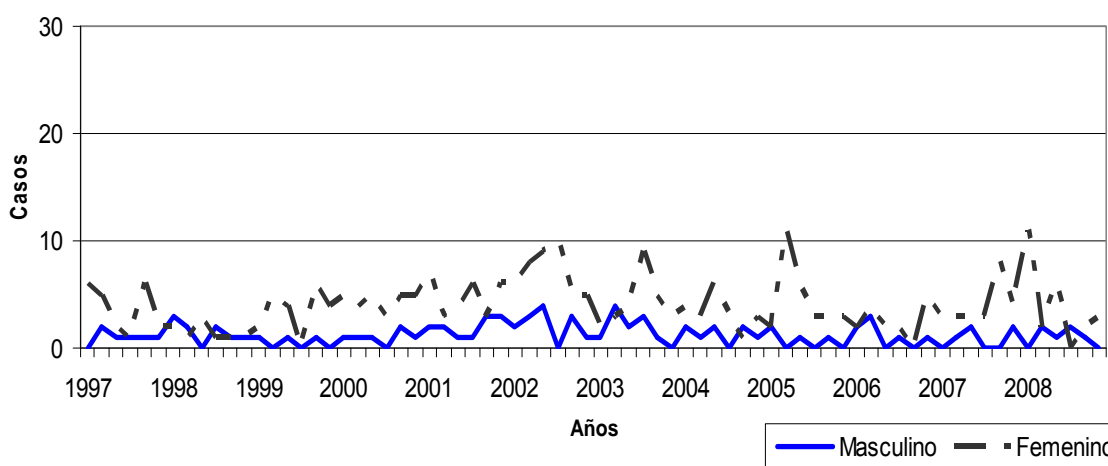


Figura 3.14 Distribución de la morbilidad por Intentos Suicidas, según sexo, en Cifuentes en el período 1997 – 2008.

Los Intentos Suicidas han tenido un comportamiento habitual en los años analizados, un ligero incremento de estos casos estuvo relacionado con los síndromes depresivos en el anciano que vive solo, siendo más evidente en el sexo femenino pues está demostrado por estudios realizados (Díaz Chalala 2007) que en las mujeres son más frecuentes los intentos y en los hombres el suicidio.

3.5 Consideraciones finales

En el presente capítulo se muestra una metodología para realizar análisis de detección de conglomerados de enfermos con ayuda de los métodos Scan. Se aplicó de manera satisfactoria a un conjunto de enfermedades del municipio de Cifuentes, considerando un período de tiempo de aproximadamente 10 años. La elección de las enfermedades de interés se realizó siguiendo sugerencias de epidemiólogos expertos en el tema.

Los métodos Scan (clásico y borroso) probaron ser herramientas útiles para este tipo de análisis. Los parámetros para la búsqueda de conglomerados se seleccionaron siguiendo sugerencias de expertos en el tema. Los resultados obtenidos fueron discutidos en todos los casos.

CONCLUSIONES

1. Se estudiaron los registros de enfermedades presentes en el municipio de Cifuentes para seleccionar las más importantes desde el punto de vista epidemiológico. Ellas fueron:
 - Enfermedades Cerebrovasculares
 - Accidentes
 - Suicidios
 - Meningoencefalitis Viral
 - Hepatitis A
 - Enfermedades del Corazón
 - Tumores Malignos
 - Intentos Suicidas
2. Se propuso una metodología general para la aplicación correcta de los métodos Scan.
3. Se realizó el procesamiento a las enfermedades seleccionadas según la metodología propuesta. En todos los casos se llegó a conclusiones estadísticas y epidemiológicas que explican el comportamiento de tales enfermedades.
4. Se realizó el procesamiento a las enfermedades con todos los casos y según el sexo del paciente.

RECOMENDACIONES

Consideramos que la metodología propuesta ayuda a los epidemiólogos a la correcta aplicación de los métodos Scan en la detección de conglomerados de enfermos. El sistema EpiDet sirve de gran ayuda en esta labor, pero él no tiene incorporado los métodos borrosos. Es por ello que se recomienda incorporárselos para que de esta manera sean factibles de utilizar por las UATS de la provincia y de toda Cuba.

REFERENCIAS BIBLIOGRÁFICAS

- Albert Cabrera, M. (2007). "Elementos básicos a conocer sobre las meningoencefalitis para los estudiantes de pregrado." Revista Electrónica de PortalesMedicos.com
- Aldrich, T. a. W., D. (1993). "Cluster", The agency for Toxic Substances and Disease Registry Division of Health Studies.
- Alegret Rodríguez, M., Grau Abalo, R., Rodríguez , M. (2008). "El enfoque espacio-temporal-contextual en el estudio del VIH-SIDA " Rev Cubana Salud Pública **Vol. 34 No. 3**.
- Álvarez Sintés, R., Díaz Alonso, G., et. al. (2001). Accidentes. Temas de Medicina General Integral. C. Médicas. La Habana. **II**: 379-384.
- Álvarez Sintés, R., Hernández Cabrera, G., Báster Moro, J., García Núñez, R., et. al. (2008). Vigilancia epidemiológica y vigilancia en salud. Medicina general integral. La Habana, Ed. Ciencias Médicas. **1**: 591-607.
- Arias, E. (1999). Clínica y terapia de las hepatitis virales. Bases moleculares para el estudio de las hepatitis virales. E. Scientiae. La Habana: 43-60.
- Batista Moliner, R., González Ochoa, E. y Feal Cañizares, P. (2001). "Evaluación de la vigilancia en salud en algunas unidades de atención primaria en Cuba." Revista Española de Salud Pública **5**: 443-458.
- Beers, H., Porter, R., Jones, T. (2007). Hematología y oncología. El manual Merck. E. española. **V**: 1119.
- Beers, H., Porter, R., Jones, T. (2007). Hepatitis. . El manual Merck. E. española. **I**: 237-249.
- Berdasquera Corcho, D., Mirabal López, María del C., Suárez Larreinag, C. (2000). "La integración de la Higiene y la Epidemiología a la Atención Primaria. Otro avance de la Salud Pública Cubana."
- Borracci, R. A., Arribalzaga Eduardo B. (2005). "Aplicación de análisis de conglomerados y redes neuronales artificiales para la clasificación y selección de candidatos a residencias médicas." Educación Médica **Vol. 8 No. 1**.
- Brubaker, D. a. S. C. (1992). "Fuzzy-logic system solves control problem." EDN **18**: 121-127.
- Calviño, M. H. (2003). "Aclarando la Lógica borrosa (Fuzzy Logic)." Revista Cubana de Física **20(2)**: 5.
- Casas Cardoso, G. (2003). Técnicas de detección de conglomerados incluyendo factores de riesgo. Centro de Estudios de Informática. Santa Clara, Universidad Central de las Villas.
- Casas Cardoso, G., Grau Abalo, R. y Cardoso Romero, G. (2002). "Validación de dos métodos de detección de conglomerados temporales usando un modelo de epidemia simple." Revista Investigación Operacional **Vol. 24, No. 1**.
- Casas Cardoso, G., Grau Abalo, R., Alegret Rodríguez, M. (1999). "Técnicas de clustering para la detección de epidemias." Unidad de Análisis y Tendencias en Salud. Ministerio de Salud Pública **Vol. 4, No. 7**
- Casas Cardoso, M., Cardoso Romero, G., Guerra Morales, V. y Herrera Jiménez, L. (2002). "Técnicas de detección de clusters aplicadas a la investigación psicológica." Revista Cubana de Psicología **Vol. 19, No. 1**.
- CDC. (1990). "Centers for Disease Control. Guidelines for Investigating Clusters of Health Events. ." Morbidity and Mortality Weekly Report, No. RR-11: 1-23.
- Col, L. A. (2000). "Approaches to space-time modelling of infectious disease behavior. IMA. ." Journal of Mathematics Applied in Medicine and Biology. **Vol 17**: 1-13.
- Consellería, S. (2002). Vigilancia en salud pública. X. d. galicia.

- Coria, L., Juárez, E., Velazco, Á. (2001). "Meningoencefalitis viral. Enfoque clínico " Revista Mexicana Pediatría **68** (6): 252-259
- Coutin Marie, G. (2008). "Pronósticos de mortalidad por enfermedades no transmisibles seleccionadas."
- Díaz Chalala, J., García Rivero, L., Yera Jaramillo, B. (2007). "Valoración de la problemática del suicidio en el municipio de Cifuentes. 2002-2006." Monografias.com
- Díaz Chalala, J., García Rivero, L., Yera Jaramillo, B. (2008). "Comportamiento, estratificación y tendencia de la mortalidad por Enfermedades Cerebrovasculares en el municipio Cifuentes.1997-2008." Monografias.com
- Duque de Estrada Riverón, J. y. R. C., A. (2001). "Factores de riesgo en la predicción de las principales enfermedades bucales en los niños " Rev Cubana Estomatología **39**(2): 111-9.
- Expósito Carralero, A., Camejo Roviralta, L., Mastrapa Reyes, C., Paredes Hernández, A. (2004). "Enfermedades cerebrovasculares. Prevalencia y factores de riesgo. Policlínica "Pedro del Toro". Holguín, enero 2003-enero 2004. ."
- Feal Cañizares, P., Batista Moliner, R., Rodríguez Milord, D. (1999). Vigilancia en la atención primaria de salud. La Habana.
- Feller, W. (1971). An Introduction to Probability Theory and Its Applications. . Reino Unido (INGLATERRA).
- Goldberg, M. (2008). La epidemiología sin esfuerzo. La Habana, Ciencias Médicas.
- Guerrero Pupo, J., Cordon Senra, Y., Paredes Hernández, A., Campins Ali, J. (2008). "Metodología de vigilancia en salud del trabajador en la atención primaria." Revista electrónica Ciencias Holguín.
- Guibert Reyes, W., Torres Miranda, N. (2001). "Intento suicida y funcionamiento familiar." Rev Cubana Med Gen Integr **17**(5): 452-60.
- Guibert Reyes, W. y. D. C. d. I., E. (2003). "Factores psicosociales de riesgo de la conducta suicida." Revista Cubana Medicina General Integral **5**: 19.
- Guirola León, I. (2009). "Revisión bibliográfica sobre meningoencefalitis con predominio de síntomas encefálicos. Encefalitis." Revista Electrónica de PortalesMedicos.com
- Jacquez, G. y. c. (1996). "Infection Control and Hospital Epidemiology." The analysis of Disease Clusters, Part I: Stat of the Art **17**(5): 319-27.
- Knox, E. (1959). "Secular pattern of congenital esophageal atresia." Brit. J. Prev. Soc. Med. **13**: 222-226.
- Knox, E. (1963). "Detection of low epidemicity: application of cleft lip and palate." Brit. J. Prev. Soc. Med. **17**: 121-127.
- Lertxundi-Manterola, A., Saez, M., Marcos-Gragera, R. y col. (2005). "Análisis de conglomerados de cáncer. El caso del barrio de Campdorà, Girona." Revista Española de Salud Pública **4**: 443-452.
- Lovelle Jorge, I., Cordero Gutiérrez, N., Álvarez Donis, A., Gutiérrez Gamazo, J., Méndez Rodríguez, M. y Rodríguez Jiménez, I. (2007). "Comportamiento de la mortalidad por tumores malignos." Revista Medicentro **11**(2).
- Lukasiewicz, J. (1910). "O zasadzie wyłączonego srodka." Przegl'd Filozoficzny **13**: 372-373.
- Martín del Brio, B. a. A. S. (2005). "Redes Neuronales y Sistemas Difusos. ." México, Alfaomega.
- Martín del Brio, B. y. A. S. (2005). "Redes Neuronales y Sistemas Difusos " México, Alfaomega.

- Martínez Jiménez, A., Moracén Disotuar, I., Madrigal Silveira, M., Almenares Aleaga, M. (2008). "Comportamiento de la conducta suicida infanto-juvenil." Revista Cubana Medicina General Integral **14(6)**: 554-9
- Más, P., Sariol, C.A., Pentón, E. (1999). El virus de la hepatitis A. Bases moleculares para el estudio de las hepatitis virales. La Habana, Elfos Scientiae: 43-60.
- Mastrapa Reyes, C., Paredes, A., Expósito Carralero, A., et. al. (2008). "Caracterización epidemiológica de los accidentes en niños menores de quince años." Correo Científico Médico de Holguín **12(4)**.
- MINSAP (2000). Programa nacional de prevención y control de las Enfermedades Cerebrovasculares. La Habana.
- Narey B, A. J. (1990). Medicina social y salud pública en Cuba. La Habana, Editorial Pueblo y Educación.
- Nass, R. (1992). "Fuzzy logic jnally gains acceptance in ihe U.S." EDN **25**: 37-40.
- Naus, J. I. (1982). ""Approximations for distributions of Scan statistics."" Journal of the American Statistical Association **77, No. 377** 177-183.
- Penichet Montoto, M., Pérez Pérez, R., Triolet Gálvez, A. (2007). Cardiopatía isquémica. Medicina Interna. Diagnóstico y tratamiento. C. Médicas. La Habana: 114.
- Rodríguez Calzadilla, A. (1997). "Enfoque de riesgo en la atención primaria de Estomatología." Rev Cubana Estomatología **18(1)**: 40-9.
- Rodríguez Corvea, L., Casas Cardoso, G., Grau Abalo, R., Pupo Meriño, M (2008). "Generalización de dos métodos de detección de conglomerados. Aplicaciones en bioinformática." Revista de Matemática: Teoría y Aplicaciones **15(1)**: 27-40
- Rodríguez, L., G. Casas, et al. (2007). "Validación del método Scan Generalizado con verdaderos y falsos conglomerados." X Congreso Nacional de Matemática y Computación. Holguín.
- Rodríguez, L., G. Casas, et al. (2009). ""Approximations for the distribution of Fuzzy Scan Statistics."" Investigación Operacional **30, No. 2** 131-139.
- Rodríguez López, L., Enríquez Govea. C., Gámez Pérez, A. et. al. (2003). "Intento suicida. Su comportamiento en la Unidad de Cuidados Intensivos." CIGET Pinar del Río **Vol.5 No.4**
- Rodríguez Méndez, O. e. a. (2006). Conducta suicida. Salud Mental Infanto - Juvenil. E. C. Médicas. La Habana: 182.
- Rodríguez Milord, D., Feal Cañizares, P., Sánchez Tarragó, N. (2001). "Las Unidades de Análisis y Tendencias en Salud dentro del Sistema de Vigilancia en Salud."
- Rodríguez Milord, D., Sánchez Tarragó, N., Feal Cañizares, P. (2002). "La práctica de la vigilancia en Salud pública. Un nuevo enfoque en la República de Cuba
- ".
- Rodríguez Rueda, J., Polanco Rodríguez, F., Olivera Bacallao, L., et. al. (2006). "Comportamiento de la enfermedad cerebrovascular en un período de dos años " Revista Cubana Medicina Militar **35(4)**
- Sánchez Tarragó, N. (2007). "Aproximación al análisis del dominio higiene y epidemiología en Cuba a través de técnicas conexionistas y multivariantes." VI Congreso Internacional de Informática en Salud.
- Shortliffe, E. H. a. B. G. B. (1975). A model of inexact reasoning in medicine. Mathematical Biosciences. **23**: 351-379.

- Toledo Curbelo, G. (2007). Fundamentos de salud pública. La Habana, Ciencias médicas.
- Torres Lio-Coo, V., Josué Díaz, L., Font Darias, I., Rodríguez Valdés, N., Urrutia Zerquera, E. y Gómez Arias, M. (2006). "Valoración de algunos aspectos relacionados con el intento suicida en pacientes jóvenes con trastornos de personalidad " Revista Cubana Medicina Militar **1**: 35.
- Valdés Gutiérrez, O., Díaz Castellón, M., Chávez Rivera, E. et al. (2006). "Comportamiento de algunos factores de riesgo asociados a las enfermedades cerebrovasculares."
- Valdés Hernández, E., Rodríguez Corvea, L., Casas Cardoso, G. (2009). Herramienta computacional para la detección de conglomerados en secuencias de ADN usando los métodos Scan. Informática 2009, La Habana.
- Yager, R. R. (2008). "Entropy and Specificity in a Mathematical Theory of Evidence, ." Springer Berlin / Heidelberg.
- Zadeh, L. A. (1968). "Fuzzy algorithms." Information and Control **12**.