

VALIDACION DE DOS METODOS DE DETECCION DE CONGLOMERADOS TEMPORALES USANDO UN MODELO DE EPIDEMIA SIMPLE

Gladys Casas Cardoso, Ricardo Grau Abalo y Gladys Cardoso Romero
Facultad de Matemática, Física y Computación, Universidad Central "Marta Abreu" de Las Villas

RESUMEN

La medicina moderna ha obtenido grandes avances en la cura de numerosas enfermedades, pero sus resultados más notables han estado en el campo de la prevención, por ejemplo resulta sumamente importante la detección temprana de focos epidémicos para evitar su propagación. Ninguno de los métodos de detección pueden aplicarse en todas las situaciones, por lo que resulta imprescindible conocer con exactitud cuando se puede confiar en sus resultados. En este trabajo se presenta la validación de dos métodos de detección de conglomerados temporales con el modelo de epidemia simple, se estudia además el comportamiento de ambos cuando se les incluyen factores de riesgo.

Palabras clave: conglomerados temporales, factores de riesgo, modelo de epidemia simple, Scan, Grimson.

ABSTRACT

Modern medicine has obtained many discoveries in healing of many diseases, however its most relevant results have been done in the field of prevention, for example it is very important to detect early an outbreak to avoid its propagation. There are many epidemic detection methods, but all of them have both vantages and limitations, it is fundamental to know them in order to avoid applying tests incorrectly. A validation of two temporal disease clusters tests with the simple epidemic model is present in this paper. The behavior of tests when risk factors are added to analysis is also presented.

Key words: temporal clusters, risk factors, simple epidemic model, Scan, Grimson.

MSC: 62H30

1. INTRODUCCION

En abril de 1790 el físico y matemático suizo Daniel Bernoulli presentó un trabajo investigativo a la Academia de Ciencias de París en el que, por primera vez se utilizaba un modelo matemático para el estudio de la dinámica de una enfermedad infecciosa. Bernoulli investigaba en particular la mortalidad causada por viruela. En su discusión matemática formuló y resolvió varias ecuaciones diferenciales y evaluó sus resultados en términos de la variación de las medidas de control de la enfermedad involucradas. Su trabajo teórico no sólo se basaba en un problema de salud real, sino que sus resultados estaban encaminados a ofrecer acciones prácticas para su solución. Más de dos siglos han pasado y las investigaciones de esa rama tienen tanta validez en nuestros tiempos como lo tuvieron en los días de Bernoulli, Bailey [1975].

En la actualidad la medicina moderna ha obtenido importantes avances en la cura de múltiples enfermedades, muchas de ellas pueden curarse e incluso evitarse mediante la utilización de vacunas. A pesar de esto las afecciones con carácter infeccioso constituyen una de las principales causas de muerte en el mundo entero; en los países en vías de desarrollo estos números alcanzan valores astronómicos. En países desarrollados por su parte las causantes de la mayoría de muertes se deben a enfermedades no transmisibles como los cánceres. Recientemente ha aparecido en el mundo una nueva enfermedad infecciosa que ataca por igual a países desarrollados o subdesarrollados y que en estos momentos ha cobrado un número inimaginable de vidas: el Síndrome de Inmunodeficiencia Adquirida (SIDA).

La epidemiología es una importante rama de la medicina que se ocupa de estudiar todo lo relacionado con las epidemias, en Cuba se estudia tanto las enfermedades infecciosas como las no infecciosas. La epidemiología por sí sola no puede dar todas las respuestas que se espera de ella, por tanto 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 frecuencia mayor. Una de las tareas más importante de los epidemiólogos es detectar en poco tiempo pequeños focos epidémicos para evitar su propagación, que puede ser rápida en ciertas enfermedades

infecciosas. En este trabajo se presentan dos métodos de detección y se validan con datos simulados del modelo más simple de epidemia.

2. LA DETECCION DE UNA INFECCION. ESTUDIOS DE CONGLOMERADOS

La mayoría de las investigaciones matemáticas sobre epidemias asumen que ya esta fue reconocida como tal y comienzan sus estudios buscando un modelo matemático suficientemente sencillo que no se aleje demasiado de la realidad epidemiológica. Por lo general para probar si un modelo es adecuado se realizan pruebas con diversos tests de bondad de ajuste, ver Bailey. [1975]

Los epidemiólogos por su parte tienen sus propios métodos de detección de epidemias, que de hecho han probado ser eficientes en numerosas ocasiones; les permiten detectar con cierta precisión el surgimiento 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 algún test de significación.

Se denomina conglomerado o “*cluster*” de enfermos a un exceso de casos diagnosticados. 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, Jacquez. [1996 a]

Numerosas pruebas estadísticas para la detección de conglomerados se han reportado en la literatura en los últimos 40 años, Jacquez, [1996 b]. El pionero de toda esta teoría fue Knox, quién introdujo por primera vez la noción de interacción espacio-temporal. Knox [1964 a,b] partió de la definición de cierto criterio de adyacencia en espacio y en tiempo y mostró sus resultados finales en una tabla de contingencia 2x2. Según él, el número de pares de casos cercanos en espacio y tiempo sigue una distribución de Poisson. Los tests de Mantel [1967] y de Jacquez [1996 c] 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 [1990], y el test de Moran ajustado, ver Oden, [1996] entre muchos otros, ver Marshall [1991]. Para detectar *clusters* en series de tiempo cortas se usan las pruebas de Larsen [1973] y la de Scan, ver Nauss [1982] y Nargawilla, [1966] 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 [1991] constituye el mejor ejemplo en estos casos.

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. Con este trabajo se pretende validar dos de los test de detección de conglomerados temporales con el modelo más simple de una epidemia. Los tests seleccionados son el Scan y el de Grimson, a continuación se explicarán muy brevemente los aspectos esenciales de su formulación matemática.

2.1. Método Scan 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 $(0, T]$. Los eventos pudieran ser por ejemplo las fechas de diagnóstico o de primeros síntomas de una determinada enfermedad o la fecha de fallecimiento en caso de que se quiera estudiar la mortalidad. 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 [1982] 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 estadístico S_w 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 $(0, T]$.

Nauss encontró $P(S_w \geq k)$ con bastante precisión, aunque ella conduce a fórmulas que son un poco engorrosas. Numerosos autores han estudiado sus propiedades hallando otras aproximaciones, para algunos casos específicos, que de cierta forma la mejoran y las han tabulado, ver Glaz [1993], Nargawilla [1996].

2.2 Método de Grimson para la detección de conglomerados temporales

El test de Grimson parte de dividir el tiempo a estudiar en intervalos más pequeños denominados celdas. Por lo general las celdas se forman por períodos regulares de tiempo: meses, semanas, días, etc. Las celdas, cuya cantidad de casos excede el valor esperado de una distribución de Poisson, se denominan celdas marcadas y la hipótesis nula del método se describe en términos de la aleatoriedad de la distribución de las celdas marcadas en todo el período de tiempo considerado. La hipótesis alternativa enuncia que existe un gran número de celdas marcadas adyacentes.

Sean:

x: número total de celdas,

y_i : número de celdas adyacentes a la celda i , $i=1,2,\dots, x$.

Para completar la formulación es necesario definir un criterio de adyacencia: dos celdas son adyacentes si ellas representan unidades continuas de tiempo, o sea:

$$y_i = \begin{cases} 1 & i = 1 \text{ o } i = x \\ 2 & i = 2, \dots, x-1 \end{cases}$$

El estadístico "A" de Grimson representa el número observado de celdas marcadas adyacentes. Bajo la hipótesis nula el valor esperado de las celdas marcadas adyacentes y su varianza pueden calcularse según fórmulas específicas detalladas en Grimson [1991].

El estadístico "A" puede estandarizarse en la forma usual y se puede aplicar un test normal clásico para determinar si "A" es significativamente mayor que su valor esperado.

3. INTRODUCCION DE FACTORES A LOS METODOS DE DETECCION DE CONGLOMERADOS

Los métodos clásicos de detección de conglomerados no ofrecen ninguna información adicional acerca de sus posibles causas, ellos solo detectan la aglomeración. Según el Centro para el Control de Enfermedades debe llevarse a cabo una investigación epidemiológica intensiva y muy costosa para tratar de detectar dichas causas.

Por lo general son muchos los factores de riesgo asociados a una patología particular y los epidemiólogos necesitan conocer cuáles de ellos son los que más están influyendo en el conglomerado en cuestión. Existe un procedimiento matemático para adicionar factores sucesivos a los métodos clásicos de detección de *clusters*, Casas [1999 a]; 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 [1999 a].

4. SOFTWARE ESPECIALIZADO PARA LA DETECCION DE CONGLOMERADOS

A nivel mundial existen paquetes específicos para la detección de aglomeraciones espaciales, temporales y espacio-temporales, tal es el caso del CLUSTER, Aldrich [1993] y el Stat!, Jacquez [1996 d], ninguno de ellos brinda facilidades de incorporación de factores a los métodos clásicos de detección de aglomeraciones inusitadas de enfermos. En Cuba ya se encuentra registrada la primera versión del sistema EpiDet (Epidemic Detector), Casas [1999 b] que contiene varios métodos de detección de *clusters* entre los que se encuentran los dos analizados con anterioridad. EpiDet permite además trabajar con factores de riesgo adicionales.

En muchas ciencias es posible realizar experimentos para obtener información adicional y probar hipótesis, pero realizar experimentos relacionados con enfermedades infecciosas en poblaciones humanas carece de ética, es totalmente imposible. Los datos que se tienen disponibles son los de epidemias ya ocurridas y con frecuencia estos están incompletos; por otra parte tampoco se pueden repetir experimentos en Epidemiología. Debido a estas dificultades resulta imprescindible el uso de modelos matemáticos y de simulaciones por computadoras para realizar experimentos teóricos. Ello posibilita además la realización de una gran cantidad de cálculos para diferentes valores de los parámetros y para distintos juegos de datos.

Los métodos que contiene el EpiDet han sido probados con datos reales de epidemias ya ocurridas y se han obtenido buenos resultados, pero resulta obvio que cada método en particular tiene sus ventajas y sus limitaciones. Es muy importante conocer dichas limitaciones para evitar aplicaciones incorrectas de los tests y arribar a conclusiones erróneas que pueden ser sumamente perjudiciales en un ámbito epidemiológico.

Con el objetivo de simular epidemias fue creado otro software: el EpiSim (Epidemic Simulator), Casas [2000 a]. Hasta el momento permite simular datos de epidemias simple y estimar los parámetros de este y otros modelos.

5. BASES DE LA SIMULACION REALIZADA

Se presentará un ejemplo para ilustrar el modelo más sencillo de epidemia en el cual la infección se propaga por contacto entre los miembros de una comunidad. En este modelo no hay eliminación de circulación de los individuos por muerte, recuperación o aislamiento, al final todos los susceptibles se infectan; puede aplicarse en situaciones en las que la enfermedad es altamente infecciosa pero no lo suficientemente seria como para que los individuos infectados mueran o haya que aislarlos, todos los infectados permanecen inmunes durante la mayor parte de la epidemia. El modelo pudiera ajustarse a los catarras comunes y a otras infecciones del tracto respiratorio, por sólo mencionar algunos ejemplos, ver Bailey [1975].

Supongamos que se tiene un grupo homogéneo de “n+a” individuos. La epidemia comienza en el instante “t = 0”, con “a” individuos infectados y “n” susceptibles; en general en el instante de tiempo “t” hay “X” susceptibles y “Y” infectados, así “X + Y = n + a”. La razón de ocurrencia de una nueva infección es proporcional al número de infectados y al número de susceptibles. El número de nuevos infectados en el intervalo de tiempo “Δt” puede escribirse como “βX YΔt”, donde “β” es la razón de infección. Existen modelos determinísticos y estocásticos para epidemias simples reportados en Bailey [1975], pero para simular la epidemia más que el modelo se necesita conocer la distribución del tiempo que transcurre entre la aparición de dos infectados consecutivos.

Si en el instante de tiempo “t” hay “r” susceptibles, habrá “n + a - r” infectados, y la probabilidad de que ocurra una nueva infección en “Δ t” será “r(n + a - r) Δt”. Esto sugiere que el intervalo de tiempo “t” que ocurre entre la infección “r” y la “r + 1” sigue una distribución exponencial negativa dada por:

$$f(t) = r(n + a - r) e^{-r(n+a-r)t} r.$$

El sistema EpiSim utiliza esta ecuación para simular epidemias simples. El obtiene números con distribución exponencial negativa y los convierte al formato de Día, Mes y Año que necesita el sistema EpiDet.

Para generar los factores dicotómicos el EpiSim utiliza una distribución de Bernoulli en la que el usuario entra la probabilidad asociada a cada uno de ellos. EpiSim genera una secuencia de 1 y 0 de acuerdo con el parámetro especificado.

6. VALIDACION CON DATOS SIMULADOS: DOS EJEMPLOS

Se simularon epidemias simples para analizar las posibilidades de los dos métodos de detección de conglomerados temporales descritos anteriormente ante situaciones de este tipo. Se eligieron cuatro valores para el tamaño de la población que van desde los más pequeños (n = 10, 20) hasta otros un poco más grandes, (n = 50, 100). Se seleccionaron tres valores para la razón de infección, considerando igualmente infecciones altas, (β = 0.01) y otras un poco menores, (β = 0.001, 0.0001). La cantidad de infectados inicial

“a” se consideró 1 en todos los casos. Para cada combinación de los valores anteriores se realizaron 1000 simulaciones.

Se aplicaron los métodos de Scan y Grimson sobre los datos simulados. Si el valor de la posibilidad empírica de significación obtenida es menor que $\alpha = 0.05$, los métodos detectan correctamente la epidemia, un valor entre 0.05 y 0.1 se considera dudoso y se le llamó medianamente significativo, y uno superior a 0.1 ofrece un resultado negativo: el test en cuestión no fue capaz de detectar el foco epidémico. Las Tablas 1 y 2 contabilizan la cantidad de juegos de datos en los que los resultados de los métodos fueron significativos, medianamente significativos o no significativos para todas las combinaciones de parámetros probadas.

Tabla 1. Resultados del método Scan.

β	n	Sig.	M. Sig.	No sig.
0.01	10	88	106	806
	20	68	90	842
	50	820	90	90
	100	1000	0	0
0.001	10	1000	0	0
	20	1000	0	0
	50	974	16	10
	100	762	23	215
0.0001	10	1000	0	0
	20	1000	0	0
	50	1000	0	0
	100	1000	0	0

Tabla 2. Resultados del métodos de Grimson.

β	n	Sig.	M. Sig.	No sig.
0.01	10	817	1	182
	20	918	38	44
	50	902	11	87
	100	701	19	280
0.001	10	777	0	223
	20	999	1	0
	50	1000	0	0
	100	999	1	0
0.0001	10	746	0	254
	20	1000	0	0
	50	1000	0	0
	100	1000	0	0

Cuando la razón de infección es alta ($\beta = 0.01$) y el tamaño de la población pequeño ($n = 10, 20$) el método Scan no puede utilizarse para la detección de epidemias simples, obsérvese que en 806 y 842 casos, respectivamente, los resultados fueron completamente negativos, un resultado similar aparece en Casas [2000 b]. Para estos casos el método de Grimson ofrece mejores resultados, ver Tabla 2. Cuando el tamaño de la población aumenta ($n = 50, 100$) el método Scan comienza a ofrecer resultados más alentadores. Para epidemias con razones de infección menores ($\beta = 0.001, 0.0001$) ambos métodos: Scan y Grimson pueden emplearse con independencia del tamaño de la población.

De manera general Grimson ofrece resultados más confiables que Scan, sus peores valores con “ β ” pequeño ($\beta = 0.001, 0.0001$) aparecen en poblaciones muy chicas ($n = 10$), para otros parámetros se comporta muy satisfactoriamente.

Tres factores de riesgo dicotómicos se añadieron al análisis anterior, el primero con probabilidad 0.7 y los dos restantes con probabilidades de 0.6. Se seleccionaron probabilidades relativamente altas porque se supone que los primeros factores que se tengan en cuenta por los profesionales de la Salud Pública sean aquellos con una mayor frecuencia. Las combinaciones en las que se obtuvieron malos resultados en las tablas anteriores se excluyeron del análisis. Nuevamente se realizaron 1000 simulaciones para cada combinación de valores de la razón de infección, el tamaño de la población y los factores considerados, un resumen de los resultados obtenidos se muestra en las Tablas 3 y 4.

Tabla 3. Resultados del método Scan con factores.

β	n	F {}	F {.7}	F {.7, .6}	F {.7, .6, .6}
0.01	10	88	-	-	-
	20	68	-	-	-
	50	820	513	210	55
	100	1000	1000	999	967
0.001	10	1000	735	313	79
	20	1000	962	639	346
	50	974	795	446	210
	100	662	318	139	67
0.0001	10	1000	863	555	185
	20	1000	998	873	560

50	1000	1000	973	832
100	1000	1000	976	762

Tabla 4. Resultados del método Grimson con factores.

β	n	F {}	F {.7}	F {.7, .6}	F {.7, .6, .6}
0.01	10	817	252	36	10
	20	918	843	359	82
	50	902	863	884	814
	100	701	661	639	602
0.001	10	777	248	40	2
	20	999	737	299	53
	50	1000	997	657	210

	100	999	988	981	809		50	1000	985	612	161
0.0001	10	746	255	41	3		100	1000	1000	865	333
	20	1000	744	232	43						

Cuando se añade un nuevo factor al análisis disminuye el número de casos en los que ambos métodos detectan un conglomerado, esto es un aspecto positivo, pues se debe a que el valor de la probabilidad empírica de significación aumenta como pronostica la teoría. Por ejemplo el método Scan con razón de infección $\beta = 0.01$ y tamaño de la población $n = 50$ detecta epidemias en 820 casos de 1000 simulados, al incluir un factor con probabilidad 0.7; este número se reduce a 513 lo que significa que en al menos 307 casos hubo una disminución de la significación, al incluir un segundo factor con probabilidad 0.6 se produce una disminución de la significación en al menos 303 casos y al incluir el tercer factor en 155. Después de incluir los tres factores Scan sigue detectando epidemias simples solo en 55 juegos de datos de los 1000 generados. A medida que aumenta el tamaño de la población ($n = 50, 100$) el decrecimiento es más lento, en estos casos hacen falta más factores para caracterizar completamente la aglomeración detectada. El mismo análisis se realiza en las dos tablas, los resultados obtenidos para ambos métodos son similares.

7. CONCLUSIONES

La detección a tiempo de una epidemia es un aspecto sumamente importante porque tiene una amplia repercusión social, los epidemiólogos necesitan para ello instrumentos matemáticos y computacionales precisos para arribar a resultados certeros. Resulta de vital importancia conocer con precisión en que momento se puede aplicar una herramienta determinada sin temor a equivocaciones que puedan ocasionar pérdidas lamentables.

Cuando ocurren procesos infecciosos simples en poblaciones pequeñas que se desencadenan con rapidez, el método Scan no es el adecuado para detectarlas, el test de Grimson ofrece resultados mucho más alentadores. Con razones de infección pequeñas ambos métodos conducen a resultados favorables.

Al incluir factores adicionales en ambos test hay un decrecimiento de los valores de la significación, lo que se corresponde con lo enunciado en la teoría.

8. RECOMENDACIONES

Se recomienda realizar el mismo estudio con otros modelos matemáticos de epidemias y ampliarlo en el análisis otros tests de detección de conglomerados temporales, espaciales y espacio-temporales.

REFERENCIAS

- ALDRICH, T. and D. WANZER (1993): "Cluster", The Agency for Toxic Substance and Disease Registry Division of Health Studies. User Manual.
- BAILEY, N.T.J. (1975): **The mathematical theory of infectious diseases and it's applications**, Charles Griffin & Company LTD, Second edition.
- CASAS, G. y R. GRAU (1999a): "Un enfoque diferente de las técnicas de *clustering* para el estudio de epidemias", **Revista de Matemática: Teoría y Aplicaciones** 6(2), 175-187.
- CASAS, G.; R. GRAU y Y. MARTINEZ (1999b): "EpiDet: Sistema Estadístico para la Detección de Conglomerados incluyendo Factores Adicionales". Manual de Usuario.
- CASAS, G.; R. GRAU y J. REYES (2000a): "EpiSim: Sistema para la Simulación de Epidemias". Manual de Usuario.
- CASAS, G. y R. GRAU (2000b): "Sistema estadístico para la detección de epidemias. Validación de un método", Memorias del III Congreso Internacional de Informática Médica de La Habana, ISBN.
- CUZICK, J. and R. EDWARDS (1990): "Spatial clustering for inhomogeneous populations" (with discussion), **Journal of the Royal Statistics Society**, Series B; 52, 73-104.

- GLAZ, J., (1993): "Approximations for the tail probabilities and moments of the Scan statistics", **Statistics in Medicine**, 12, 1845-1852.
- GRIMSON, R. and R. ROSE (1991): "A versatile test for clustering and a Proximity Analysis of Neurons", **Methods of Information in Medicine**, 30, 299-303.
- Jacquez, G. **et al.** (1996a): "The analysis of Disease Clusters, Part I: Stat of the Art", **Infection Control and Hospital Epidemiology**, 17(5), 319-327.
- _____ (1996b): "The analysis of Disease Clusters, Part II: Introduction to techniques" **Infection Control and Hospital Epidemiology** 17(6) 17, 385-97.
- _____ (1996c): "A k-nearest neighbor test for space-time interaction", **Statistics in Medicine** 15, 1934-1949.
- _____ (1996d): "Stal: Statistical Software for the Clustering of Health Eventual Bio-Medware", User Manual.
- KNOX, E. (1964a): "The detection of space-time interactions", **Applied Statistics** 13, 25-29.
- _____ (1964b): "Epidemiology of childhood leukemia in Northumberland and Durham", **British Journal of Preventative and Social Medicine**, 18, 17-24.
- LARSEN, R.; C. HOLMES and C. HEATH (1973): "A statistical test for measuring unimodal clustering, a description of the test and of it's applications of cases of acute leukaemia in metropolitan Atlanta, Georgia", **Biometrics**, 29, 301-09.
- MANTEL, N. (1967): "The detection of disease clustering and a generalized regression approach", **Cancer Research**, 27, 209-220.
- MARSHALI, R. (1991): "A review of methods for the statistical analysis of spatial patterns of disease", **Journal of the Royal Statistical Society Association**, 154, 421-441.
- NAGARWILLA, N. (1996): "A Scan statistic with a variable window", **Statistics in Medicine**, 15, 845-850.
- NAUSS, J. (1982): "Approximations for distributions of Scan statistics", **Journal of the American Statistics Association**, 77, 377.
- ODEN, N. and G. JACQUEZ (1996): "Realistic Power Simulations compare Point-and Area-Based Disease Cluster Test", **Statistics in Medicine**, 15, 783-806.