

Introducción.

La complejidad de los modelos medioambientales es que virtualmente cualquier proceso físico lleva consigo una variabilidad en el espacio y tiempo.

Definición: Si se toman al azar dos sucesos arbitrarios en tiempo y espacio y existe una relación entre ellos de poca distancia espacial y temporal, se dirá que existe una **interacción espacio-tiempo**.

¿Cómo se modelizan las interacciones espacio-tiempo?

Si **sólo** se considera la **componente temporal** usaremos modelos autoregresivos que son dinámicos en el sentido que explotan la componente unidireccional del flujo del tiempo. Se consideran los datos como series temporales separadas pero correlacionadas en el espacio (i.e., un modelo multivariado de **series temporales**). Tienen el inconveniente que estos procedimientos no permiten una predicción espacial directa en aquellas localizaciones para las que no hay datos.

Si **sólo** se considera la **componente espacial** los métodos geoestadísticos son métodos descriptivos en el sentido que no hay ninguna interpretación causativa asociada.

Una **metodología** que modelice ambas componentes sería la **adecuada**, buscando un modelo estadístico que fuera temporalmente dinámico y espacialmente descriptivo.

¿En qué contextos son útiles estos modelos?

El estudio de aglomeraciones en dominios espacio-temporales va más allá de la detección de conglomerados (clusters) en el espacio, por un lado, y en el tiempo, por otro; de hecho, ambos pueden existir por separado sin que haya interacción. La interacción supone que los casos cercanos en el espacio son, además, cercanos en el tiempo, de modo que la localización de un caso depende de la localización del caso que lo precede. Este tipo de patrones es muy útil cuando se desea investigar **enfermedades transmisibles**. También se pone de manifiesto la interacción cuando la causa de la enfermedad es la exposición a un agente geográficamente localizado (una sustancia tóxica, ciertos tipos de radiaciones, etc.).

Así consideraremos una **epidemia, desde el punto de vista temporal**, una función del número de eventos en una unidad de tiempo elegida. La epidemia se traza en un mapa con respecto al tiempo, se observa que los sucesos se agregan discretamente y son las columnas del histograma las que indican una epidemia.

Desde el punto de vista espacial, las enfermedades epidémicas se pueden extender en el espacio y si a menudo se ignora este hecho es porque las condiciones de un estudio están frecuentemente limitadas por límites geográficos dentro de lo que es un modelo bastante homogéneo y simultáneo en el tiempo.

Inconveniente la claridad en una dimensión exige la restricción en la otra

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.

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.

Teniendo en cuenta la gran cantidad de datos y el gran número de contrastes de hipótesis a realizar se utilizará, como soporte informático, el software libre Epidat 3.1. Este podrá ser descargado en la dirección de internet <http://dxsp.sergas.es> , aunque viene adjunto en la publicación.

Los métodos estudiados para la detección de agregaciones temporales, casos agrupados:

- Método Scan
- Método Texas
- Método Poisson
- Método Cusum

1. Agregaciones temporales.

En este tipo de agregaciones el investigador busca 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 si se tratara de una distribución aleatoria.

Las técnicas de agregaciones 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.

1.1. Método Scan

El test Scan, desarrollado por Nauss¹⁷⁻¹⁸, es una prueba estadística para detectar agrupaciones temporales 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 una componente espacial, es adecuado para estudios de enfermedades en los que la población se distribuye de forma desigual en el área de estudio.

El estadístico Scan, en el que se basa la prueba, es el número máximo de casos observados en una ventana móvil de longitud fija, que se mueve de forma continua a lo largo del tiempo. Si los casos se agrupan en el tiempo, entonces el número máximo de casos en la ventana temporal será grande. El ancho de la ventana se basa en la duración esperada de una epidemia.

Los **datos** necesarios para aplicar la prueba son:

- La serie temporal de casos observados (en días, semanas, cuatrisesmanas, meses o años): $X_i, i=1, \dots, T$.
- w : la amplitud de la ventana, expresada en las mismas unidades de tiempo. Si $X_{t,t+w}$ denota el número de casos en el período $(t, t+w]$, el estadístico Scan se expresa como:

$$S_w = \max_{0 \leq t \leq T-w} Y_{t,t+w}$$

Se rechaza la hipótesis nula, al nivel de significación α , si la probabilidad de que el estadístico sea mayor que el valor observado en la serie es mayor o igual que α :

$$\text{Valor } p = \Pr(S_w \geq n) \geq \alpha$$

Este p-valor depende de 3 datos:

- El número total de casos observados (N).
- La amplitud relativa de la ventana respecto al período total de estudio ($r=w/T$).
- El número máximo de casos observados en todas las ventanas temporales de amplitud w (n). Se denota como

$$\text{Pr}(n, N, r)$$

Matemáticamente, es la probabilidad de que si N puntos están distribuidos aleatoriamente a lo largo de una línea de longitud 1, exista un intervalo de longitud r que contenga al menos n de ellos. Como la distribución del estadístico Scan bajo la hipótesis nula es desconocida, el valor p debe calcularse con aproximaciones o por simulación; Epidat utiliza una aproximación basada en la distribución binomial.

Las **limitaciones** del método Scan son:

- El método Scan asume que la población a 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.

- No puede distinguir entre un cluster debido a un cambio en el patrón de los factores de riesgo conocidos de la enfermedad, y un cluster debido a otras causas.

- Hay que tener especial cuidado en la selección de la ventana temporal, 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.

- El procedimiento no es útil cuando el número de casos detectados es extremadamente bajo (<10).

Ejercicio: En un período de 28 años, entre 1975 y 2002, se notifican 255 casos de cierto tipo de cáncer en una región. La serie anual, que puede verse en la Tabla 2, ¿Hay evidencia de alguna agregación de casos en el lapso de un año? ¿y en 3 años?

Tabla 2. Serie anual de casos observados de 1975 a 2002.

AÑO	CASOS
1975	15
1976	12
1977	13
1978	11
1979	15
1980	12
1981	9
1982	12
1983	8
1984	4

AÑO	CASOS
1985	17
1986	13
1987	15
1988	12
1989	9
1990	12
1991	6
1992	7
1993	8
1994	4

AÑO	CASOS
1995	6
1996	5
1997	3
1998	7
1999	8
2000	2
2001	4
2002	6

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Excel o Access (entrada automática). En ambos casos, el usuario debe seleccionar en primer lugar la unidad de medida del tiempo entre días, semanas, cuatrisesemanas, meses y años. A continuación, si la entrada es manual, se introducen el período inicial (por ejemplo, el mes y año cuando la serie de casos es mensual) y el número total de períodos (en ese caso, número de meses); entonces, el programa genera una tabla con tantas filas como períodos y sólo hay que introducir el número de casos en cada fila.

Si la entrada es automática, Epidat 3.1 necesita que la tabla de datos tenga una estructura determinada, con una variable que contenga la serie temporal de casos observados, y una o dos variables que identifiquen el período temporal de ocurrencia, según la unidad de medida seleccionada (días, semanas, cuatrisesemanas, meses o años). Por ejemplo, si el período elegido es días, sólo se necesita una variable temporal que contenga la fecha de ocurrencia de los casos. Si la serie de casos es semanal, mensual o cuatrisesemanal, la tabla debe contener una variable para identificar el año y otra para el número de semana, mes o cuatrisesemana, respectivamente. Por último, en el caso de series anuales, sólo es necesario el año de ocurrencia (como ejemplo, véase la Tabla 3).

Para analizar si hay agregaciones anuales de casos debe tomarse una ventana de amplitud 1.

Para ese valor, el número máximo de casos en una ventana de amplitud 1 es el máximo de los casos anuales, que corresponde a los 17 casos del año 1985 (Tabla 2).

Tabla 3. Formato de tabla preparada para importar datos desde Epidat 3.1 para la aplicación del método Scan.

AÑO	CASOS
1975	15
1976	12
1977	13
1978	11
...	...
2001	4
2002	6

Resultados con Epidat 3.1 para una ventana de amplitud 1 año:

```
Vigilancia: Detección de clusters, agregaciones temporales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \SCAN.xls

Campo que identifica:
  Casos observados: CASOS
  Años: AÑO

Método          : Scan
Tipo de período: Años
Ventana temporal en Años: 1

Nº máximo de casos en la ventana temporal: 17
Probabilidad {Casos esperados >= 17} = 0,7699

Intervalo 1
  Año      Casos
-----
  1985     17
-----
Total     17
```

En este caso, ya la probabilidad de observar un número de casos igual o mayor que $n=17$ es tan elevada (0,7699), no hay evidencias para rechazar la hipótesis nula con un nivel de significación del 5% y, por tanto, se admite que no existe ninguna agrupación anual de casos.

Si tomáramos una ventana de longitud 3 años, los resultados serían diferentes. En primer lugar, habría que calcular el número de casos en todas las posibles ventanas de 3 años:

Ventana de 3 años			Casos
1975	1976	1977	40
1976	1977	1978	36
1977	1978	1979	39
1978	1979	1980	38
1979	1980	1981	36
1980	1981	1982	33
...
2000	2001	2002	12

Resultados con Epidat 3.1 para una ventana de amplitud 3 años:

```

Vigilancia: Detección de clusters, agregaciones temporales

Archivo de trabajo: C:\Archivos de programa\Epidat 3.1\Ejemplos
\Vigilancia\SCAN.xls

Campo que identifica:
  Casos observados: CASOS
  Años: AÑO

Método           : Scan
Tipo de período: Años
Ventana temporal en Años: 3

Nº máximo de casos en la ventana temporal: 45
Probabilidad {Casos esperados >= 45} = 0,0457

Intervalo 1
-----
      Año      Casos
-----
      1985      17
      1986      13
      1987      15
-----
      Total      45

```

Ahora, en cada intervalo de 3 años, comenzando en 1975, se busca la mayor agrupación de casos que, en efecto, se encuentra en la ventana formada por los años 1985, 1986 y 1987. La probabilidad de observar 45 casos o más en tres años es un valor pequeño ($p=0,0457 < 0,05$), por lo que hay evidencia estadística de que la agrupación observada no es producto de la mera casualidad.

Hay que tener en cuenta que muchas veces se desconoce la duración esperada de la enfermedad, de modo que en la selección de la amplitud de la ventana puede haber un grado de subjetividad elevado; por esta razón, es aconsejable probar con varios valores.

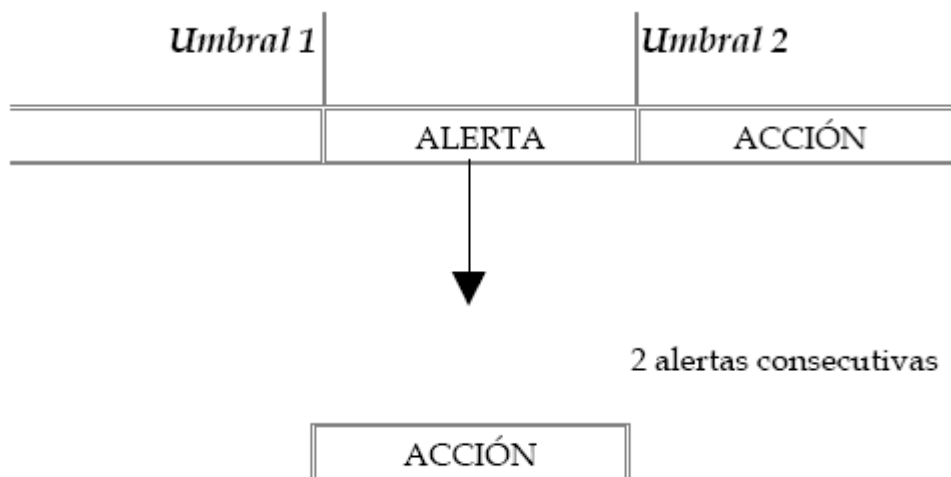
1.2. Método Texas

El método Texas es un procedimiento estadístico que permite la detección de agregaciones temporales en la incidencia de una enfermedad de baja frecuencia y ha sido desarrollado en el contexto de una comunidad potencialmente expuesta a una fuente contaminante. Por ejemplo, ha sido utilizado para el estudio de efectos adversos sobre la salud provocados por emisiones tóxicas de industrias locales.

Esta metodología se basa en una regla de decisión, en dos etapas, que utiliza la razón estandarizada de mortalidad (RME o SMR en inglés) para identificar un aumento significativo en la mortalidad o morbilidad de una región con respecto a lo esperado. El método establece dos valores umbrales ($umbral1 < umbral2$), que definen una *fase de alerta* cuando los datos indican un posible problema de salud, o una *fase de actuación* cuando hay evidencias de un problema de salud grave o prolongado y es necesario intervenir.

La regla de decisión para definir los estados de alerta o actuación es la siguiente: si se supera el *umbral2* se declara el estado de actuación, es decir, se identifica una agregación temporal inusual; si ese nivel no es excedido pero sí el *umbral1*, entonces se declara el estado de alerta. Además, dos alertas consecutivas definen también un estado de actuación (Figura 1).

Figura 1.- Esquema de la regla de decisión en el método TEXAS.



Los **datos** necesarios para aplicar el método son:

- La serie temporal de casos observados (en semanas, cuatrisesmanas, meses o años): O_i .
- La serie temporal de casos esperados (en semanas, cuatrisesmanas, meses o años): E_i .
- P1: la probabilidad de alerta.
- P2: La probabilidad de acción ($P2 < P1$).

El cálculo de los valores umbrales y la regla de decisión se basan en la distribución de Poisson si el número de casos esperado es menor o igual que 30, o en una aproximación a la distribución normal en caso contrario.

Si $E_i > 30$, el procedimiento tiene los siguientes **pasos**:

1. Se calculan los valores umbrales de cada período:

$$Umbral1 = 1 + \frac{\phi^{-1}(1 - P_1)}{\sqrt{E_i}} \quad Umbral2 = 1 + \frac{\phi^{-1}(1 - P_2)}{\sqrt{E_i}}$$

donde ϕ es la función de distribución de la Normal estándar.

2. Se calcula el SMR de cada período como el cociente entre el número de casos observado y el esperado.

3. Se aplica la siguiente regla:

- ALERTA: $Umbral1 < SMR_i < Umbral2$
- ACTUACIÓN:
 - $SMR_i \geq Umbral2$
 - $Umbral1 < SMR_i < Umbral2$ y $Umbral1 < SMR_{i-1} < Umbral2$ (2 alertas consecutivas)

Por el contrario, si $E_i \leq 30$, los **pasos** son:

1. Se calculan los valores umbrales:

$$Umbral1 = P_1 \quad Umbral2 = P_2$$

2. Se calcula, para cada período, la probabilidad de observar al menos los O_i casos observados a partir de una distribución de Poisson de media E_i :

$$P_i = \Pr[\text{Poisson}(E_i) \geq O_i]$$

3. Se aplica la siguiente regla:

- ALERTA: $Umbral1 < P_i < Umbral2$
- ACTUACIÓN:
 - $P_i \geq Umbral2$
 - $Umbral1 < P_i < Umbral2$ y $Umbral1 < P_{i-1} < Umbral2$ (2 alertas consecutivas)

Las **limitaciones** del método Texas son las siguientes:

- Para la estimación del número de casos esperado (generalmente se ajusta por edad, sexo, raza, etc) se requiere conocer las características demográficas de la población. También es necesario conocer las tasas específicas de mortalidad de una población de referencia y ésta última determina los resultados del método.

- La razón estandarizada de mortalidad es apropiada para determinar los estados de alerta o actuación si la población que se controla es pequeña.

- La determinación de las probabilidades de alerta y actuación, P1 y P2 respectivamente, es arbitraria.

Ejercicio: En una población cercana a una fuente contaminante se establece un sistema de vigilancia de abortos espontáneos. Los casos observados mensualmente durante el año 1999 se presentan en la Tabla 4, y se estima que el número esperado de abortos espontáneos al mes es de 10 casos. El ejemplo se ha tomado del manual del CLUSTER13, y los datos se encuentran en el archivo TEXAS.xls distribuido con Epidat 3.1. ¿En qué meses se producen alertas si se fijan las probabilidades de alerta y actuación en el 5% y 1%, respectivamente?

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Excel o Access (entrada automática). En ambos casos, el usuario debe seleccionar en primer lugar la unidad de medida del tiempo entre semanas, cuatrisesmanas, meses y años. A continuación, si la entrada es manual, se introducen el período inicial (por ejemplo, el mes y año cuando la serie de casos es mensual) y el número total de períodos (en ese caso, número de meses); entonces, el programa genera una tabla con tantas filas como períodos y sólo hay que introducir el número de casos observados y esperados en cada período.

Si la entrada es automática, Epidat 3.1 necesita que la tabla de datos tenga una estructura determinada, con dos variables que contengan, respectivamente, las series de casos observados y esperados, y una o dos variables que identifiquen el período temporal de ocurrencia, según la unidad de medida seleccionada (semanas, cuatrisesmanas, meses o años).

Por ejemplo, si la serie de casos es semanal, mensual o cuatrisesmanal, la tabla debe contener una variable para identificar el año y otra para el número de semana, mes o cuatrisesmana, respectivamente (como ejemplo, véase la Tabla 4); en el caso de series anuales, sólo es necesario el año de ocurrencia.

En cualquiera de los dos casos deben introducirse, además, las probabilidades de alerta y actuación, en porcentaje, que en el ejemplo son 5% y 1%, respectivamente.

Tabla 4. Casos de abortos espontáneos observados y esperados mensualmente durante el año 1999.

AÑO	MES	OBSERVADOS	ESPERADOS
1999	1	8	10
1999	2	10	10
1999	3	9	10
1999	4	11	10
1999	5	14	10
1999	6	16	10
1999	7	15	10
1999	8	13	10
1999	9	17	10
1999	10	14	10
1999	11	23	10
1999	12	13	10

Resultados con Epidat 3.1

```
Vigilancia: Detección de clusters, agregaciones temporales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \TEXAS.xls

Campo que identifica:
  Casos observados: OBSERVADOS
  Casos esperados: ESPERADOS
  Meses: MES
  Años: AÑO

Método                : Texas
Tipo de período       : Meses
Probabilidad de alerta (en %): 5
Probabilidad de acción (en %): 1
```

Mes	SMR	Estado
1-1999	0,8000	---
2-1999	1,0000	---
3-1999	0,9000	---
4-1999	1,1000	---
5-1999	1,4000	---
6-1999	1,6000	Alerta
7-1999	1,5000	---
8-1999	1,3000	---
9-1999	1,7000	Alerta
10-1999	1,4000	---
11-1999	2,3000	Acción
12-1999	1,3000	---

A la vista de los resultados anteriores, se declara el estado de *alerta* en los meses de junio y septiembre, y el estado de *acción* en el mes de noviembre.

1.3. Método Poisson.

El método de Poisson, como su nombre indica, está basado en la distribución de Poisson y permite detectar **agregaciones** de casos en el **tiempo**. Es un procedimiento muy sencillo que ha sido aplicado, por ejemplo, en la vigilancia de malformaciones congénitas en nacidos vivos. El nombre es debido al matemático francés Siméon Denis Poisson (1781-1840), al que se le conoce por sus diferentes trabajos en el campo de la electricidad, también hizo publicaciones sobre la geometría diferencial y la teoría de probabilidades.

La distribución que lleva su nombre y que publicó, junto con su teoría de probabilidad, en 1838 en su trabajo *Recherches sur la probabilité des jugements en matières criminelles et matière civile* ("Investigación sobre la probabilidad de los juicios en materias criminales y civiles"). El trabajo estaba enfocado en ciertas variables aleatorias N que cuentan, entre otras cosas, un número de ocurrencias discretas (muchas veces llamadas "arribos") que tienen lugar durante un intervalo de tiempo de duración determinada. Si el número esperado de ocurrencias en este intervalo es λ , entonces la probabilidad de que haya exactamente k ocurrencias (siendo k un entero no negativo, $k = 0, 1, 2, \dots$) es igual a:

$$f(k; \lambda) = \frac{e^{-\lambda} \lambda^k}{k!},$$

dónde

- e es el base del logaritmo natural ($e = 2.71828\dots$),
- $k!$ es el factorial de k ,
- k es el número de ocurrencias de un evento,
- λ es un número real positivo, equivalente al número esperado de ocurrencias durante un intervalo dado. Por ejemplo, si los eventos ocurren de media cada 4 minutos, y se está interesado en el número de eventos ocurriendo en un intervalo de 10 minutos, se usaría como modelo una distribución de Poisson con $\lambda = 2.5$.

Los **datos** necesarios para la aplicación de esta técnica son los siguientes:

- La serie temporal de casos observados (en semanas, cuatrisesmanas, meses o años): O_i .
- La serie temporal de casos esperados (en semanas, cuatrisesmanas, meses o años): E_i .

Suponiendo que el número de casos observados en el período i -ésimo (O_i) sigue una distribución de Poisson con media igual al número esperado de casos (E_i), este método basa

su regla de decisión en la probabilidad de observar O_i casos o más:

$$P_i = \Pr[\text{Poisson}(E_i) \geq O_i]$$

Si esta probabilidad es pequeña (menor que el nivel de significación de la prueba) se rechaza la hipótesis nula de que no existe agregación de casos en el período considerado.

La mayor **limitación** del método Poisson es que no siempre se puede suponer que las enfermedades siguen una distribución de Poisson. Si, por ejemplo, el área de estudio es pequeña, o la enfermedad es poco frecuente, puede producirse lo que se denomina sobredispersión cuando la varianza es mayor que la media (en una distribución de Poisson son iguales). En esos casos, distribuciones como la binomial negativa pueden ser más adecuadas.

Ejercicio: Durante el año 1987 se diagnostican 105 casos de cierto tipo de cáncer en una región. La serie mensual de casos observados y esperados puede verse en la Tabla 5. En el período de estudio, ¿hay evidencia de agregación de casos en algún mes?

Tabla 5. Casos de cáncer observados y esperados mensualmente durante el año 1987.

AÑO	MES	OBSERVADOS	ESPERADOS
1987	1	11	10
1987	2	8	10
1987	3	9	10
1987	4	14	10
1987	5	19	10
1987	6	11	10
1987	7	9	8
1987	8	8	8
1987	9	7	8
1987	10	4	8
1987	11	0	8
1987	12	5	8

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato

Dbase, Excel o Access (entrada automática). En ambos casos, el usuario debe seleccionar en primer lugar la unidad de medida del tiempo entre semanas, cuatrisesmanas, meses y años. A continuación, si la entrada es manual, se introducen el período inicial (por ejemplo, el mes y año cuando la serie de casos es mensual) y el número total de períodos (en ese caso, número de meses); entonces, el programa genera una tabla con tantas filas como períodos y sólo hay que introducir el número de casos observados y esperados en cada período.

Si la entrada es automática, Epidat 3.1 necesita que la tabla de datos tenga una estructura determinada, con dos variables que contengan, respectivamente, las series de casos observados y esperados, y una o dos variables que identifiquen el período temporal de ocurrencia, según la unidad de medida seleccionada (semanas, cuatrisesmanas, meses o años).

Por ejemplo, si la serie de casos es semanal, mensual o cuatrisesmanal, la tabla debe contener una variable para identificar el año y otra para el número de semana, mes o cuatrisesmana, respectivamente (como ejemplo, véase la Tabla 5); en el caso de series anuales, sólo es necesario el año de ocurrencia.

Resultados con Epidat 3.1

```
Vigilancia: Detección de clusters, agregaciones temporales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \POISSON.xls

Campo que identifica:
Casos observados: OBSERVADOS
Casos esperados: ESPERADOS
Meses: MES
Años: AÑO

Método          : Poisson
Tipo de período: Meses
```

Mes Observados	Esperados	Valor p	
1-1987	11	10,0000	0,4170
2-1987	8	10,0000	0,7798
3-1987	9	10,0000	0,6672
4-1987	14	10,0000	0,1355
5-1987	19	10,0000	0,0072
6-1987	11	10,0000	0,4170
7-1987	9	8,0000	0,4075
8-1987	8	8,0000	0,5470
9-1987	7	8,0000	0,6866
10-1987	4	8,0000	0,9576
11-1987	0	8,0000	1,0000
12-1987	5	8,0000	0,9004

Valor p: probabilidad de que n° de casos esperado sea >= al observado.

Epidat, con la información de los casos observados y de los casos esperados, calcula, para cada mes, la probabilidad de observar esos casos o más bajo una distribución de Poisson. Si esa probabilidad resulta ser muy pequeña, se concluye que hay evidencia de una agregación temporal que no es producto del azar. Nótese como en el mes de mayo se da un número inusitado de casos con una baja probabilidad.

1.4. Método Cusum

Esta técnica fue diseñada para el control de calidad de algunos procesos industriales y se utiliza en epidemiología para detectar incrementos en la incidencia de una enfermedad, por ejemplo para vigilar malformaciones congénitas.

Los **datos** necesarios para la aplicación de esta técnica son los siguientes:

- La serie temporal de casos observados (en días, semanas, cuatrisesemanas o meses): O_i .
- E: el número esperado de casos por unidad de tiempo.
- El CUSUM inicial.

A partir del número esperado de casos, E, se obtienen dos valores K y h, conocidos como valor de referencia y de alarma, respectivamente, de modo que maximicen la capacidad de detección del método y minimicen la probabilidad de producir falsas alarmas. Estos valores están tabulados en función de E cuando $E < 9$.

Si $E > 9$ se toman $K=1$ y $h=2$.

La técnica basa su decisión en la comparación del valor CUSUM, calculado para cada intervalo de tiempo, con el valor h. Si $CUSUM > h$, se detecta un cambio significativo en la frecuencia de la enfermedad y se declara una alerta.

Para obtener los valores CUSUM se procede del modo siguiente:

1. Para cada período k se calcula:

$$S_k = \text{Max}\{0; O_i - K\} \text{ si } E < 9$$

$$S_k = \text{Max}\left\{0; \frac{O_i - E}{\sqrt{E}} - 1\right\} \text{ si } E \geq 9$$

2. El CUSUM del período i es la suma acumulada de valores S_k desde el período inicial, si aún no se ha declarado ninguna alerta o, en caso contrario, desde el período siguiente al de la última alerta, y hasta el período i. Para calcular el primer valor acumulado CUSUM1 se añade a S_1 el valor del cusum inicial CUSUM0.

Las **limitaciones** del método Cusum son:

- La sensibilidad del método depende de la incidencia de la enfermedad en la población.
- Cuando el número esperado de casos es mayor o igual que 9 el método funciona peor.

Ejercicio: En una población se han observado 26 casos de anencefalia en los 7 primeros meses del año 1990 y una revisión de los registros del año anterior indica que el número esperado de casos para esta región es de 2 casos mensuales. Los datos, que se han tomado del manual del CLUSTER13 se presentan en la Tabla 6 y se encuentran, además, en el archivo CUSUM.xls distribuido con Epidat 3.1. Partiendo de un CUSUM inicial igual a 0, ¿hay evidencia de un cluster de anencefalia en algún mes del período estudiado?

Los datos pueden introducirse desde el teclado (entrada manual) o importarse en formato Dbase, Excel o Access (entrada automática). En ambos casos, el usuario debe seleccionar en primer lugar la unidad de medida del tiempo entre días, semanas, cuatrisesmanas o meses.

A continuación, si la entrada es manual, se introducen el período inicial (por ejemplo, el mes y año cuando la serie de casos es mensual) y el número total de períodos (en ese caso, número de meses); entonces, el programa genera una tabla con tantas filas como períodos y sólo hay que introducir el número de casos observados en cada período.

Si la entrada es automática, Epidat 3.1 necesita que la tabla de datos tenga una estructura determinada, con una variable que contenga la serie temporal de casos observados, y una o dos variables que identifiquen el período temporal de ocurrencia, según la unidad de medida seleccionada (días, semanas, cuatrisesmanas o meses). Por ejemplo, si el período elegido es días, sólo se necesita una variable temporal que contenga la fecha de ocurrencia de los casos. Si la serie de casos es semanal, mensual o cuatrisesmanal, la tabla debe contener una variable para identificar el año y otra para el número de semana, mes o cuatrisesmana, respectivamente (como ejemplo, véase la Tabla 6).

Tabla 6. Casos mensuales de anencefalia entre enero y julio de 1990.

AÑO	MES	OBSERVADOS
1990	1	1
1990	2	3
1990	3	2
1990	4	4
1990	5	5
1990	6	5
1990	7	6

En cualquiera de los dos casos deben introducirse, además, el número esperado de casos por período de tiempo y el CUSUM inicial.

Resultados con Epidat 3.1:

```
Vigilancia: Detección de clusters, agregaciones temporales

Archivo de trabajo: C:\Archivos de programa \Epidat 3.1 \Ejemplos
\Vigilancia \CUSUM.xls

Campo que identifica:
  Casos observados: OBSERVADOS
  Días: FECHA
  Meses: MES
  Años: AÑO

Método          : Cusum
Tipo de período: Meses
Cusum inicial   : 0
Nº esperado de casos por Meses: 2
```

	Mes Observados	Cusum
1-1990	1	0,00
2-1990	3	0,00
3-1990	2	0,00
4-1990	4	1,00
5-1990	5	3,00
6-1990	5	5,00
7-1990	6	8,00 Alarma

Los resultados indican que el nivel de alerta se superó en el mes de julio, indicando una posible agrupación de casos de anencefalia en ese mes.

Bibliografía

- Webs:

<http://www.wikipedia.org>
<http://www.google.com>

- Libros y artículos de consulta:

Chen R, Mantel N. "A study of tree techniques for time-space clustering in Hodkin's disease". Stat Med. 1984

Chen R, "Statistical techniques in Viti defects surveillance Systems". Contr Epidemiol Biostat. 1979.

Glaz J, "Approximations for the tail probabilities and moments of the scan statics". Stat Med. 1993.

"Continuous inspection schemes". Biometrika. 1954.