



UNIVERSIDAD DE LA HABANA

## TRABAJO DE DIPLOMA

### **Metodologías para el análisis de la variación geográfica del riesgo relativo para la morbi-mortalidad de enfermedades**

Autora: Vianka Orovio Cobo  
Tutoras: MSc. Angela Damary Tuero Iglesias

Trabajo de Diploma presentado  
para obtener el título de  
Licenciada en Matemáticas

Julio del año 2003

A mis padres.

## AGRADECIMIENTOS

Quiero agradecer a todos los que me han ayudado y apoyado en el desarrollo y culminación de esta tesis de diplomado y en especial:

*A mi tutora MSc Angela Damary Tuero Iglesias*, quien me motivó a la realización de esta tesis y posteriormente me guió prestándome una valiosísima ayuda.

*A mis padres* por haber depositado en mi toda su confianza y cariño, por brindarme cada día su amor y darme su apoyo incondicional en todo momento.

*A mi novio*, de corazón, por el gran apoyo que me ha dado en estos años.

*A mi familia*, por el amor y apoyo que he recibido de ellos siempre.

*A mis profesores*, por enseñarme todo lo que hoy sé sobre mi profesión.

*A mis amigos*, que me han alentado y guiado, hasta en los momentos más difíciles; a los que siempre me han deseado lo mejor.

A todos aquellos que de una forma u otra me han ayudado en la realización de esta tesis.

## ÍNDICE

<b>Introducción</b> .....	<b>1</b>
<b>Capítulo 1: Metodología para el cálculo del riesgo relativo para la morbi-mortalidad de enfermedades. Mapeo de su distribución geográfica</b>	
1.1- Introducción.....	3
1.2- Conceptos de interés.....	4
1.3- Metodología clásica para el cálculo del riesgo relativo para la morbi-mortalidad de enfermedades y el mapeo de su distribución geográfica.....	8
1.4- Metodología bayesiana para el cálculo del riesgo relativo para la morbi-mortalidad de enfermedades y el mapeo de su distribución geográfica.....	11
<b>Capítulo 2: Sistema computacional utilizado en el cálculo del riesgo relativo para la morbi-mortalidad de enfermedades y de su representación geográfica en los mapas</b>	
2.1- Introducción.....	22
2.2- Estudio del sistema WinBUGS 1.4 para el cálculo del riesgo relativo utilizando el enfoque bayesiano.....	23
2.3- Estudio del sistema GeoBUGS para la representación en mapas de la distribución del riesgo relativo.....	31
<b>Capítulo 3: Análisis de los resultados obtenidos al aplicar la metodología clásica y la bayesiana para el cálculo del riesgo relativo para la mortalidad por enfermedades cerebrovasculares en los municipios del país</b>	
3.1- Introducción.....	33
3.2- Resultados obtenidos de aplicar la metodología clásica al cálculo y mapeo del riesgo relativo.....	35
3.3- Resultados obtenidos de aplicar la metodología bayesiana al cálculo y mapeo del riesgo relativo.....	38
3.4- Comparación de los resultados obtenidos por ambas metodologías.....	39
<b>Conclusiones</b> .....	<b>40</b>
<b>Bibliografía</b> .....	<b>42</b>
<b>Anexos:</b>	
<b>Anexo 0:</b> Tabla de datos en Excel.....	<b>45</b>
<b>Anexo 1:</b> Defunciones por grupos de edades para las Enfermedades Cerebrovasculares en Cuba, año 2002.....	<b>47</b>

<b>Anexo 2:</b> Hoja de datos para el cálculo utilizando el WinBUGS para el análisis bayesiano.....	<b>48</b>
<b>Anexo 3:</b> Salidas de Excel obtenidas de aplicar la metodología clásica en el cálculo del riesgo relativo.....	<b>52</b>
<b>Anexo 4:</b> Salidas del WinBUGS 1.4 obtenidas de aplicar la metodología bayesiana en el cálculo del riesgo relativo.....	<b>54</b>

## INTRODUCCIÓN

Son diversas las soluciones estadísticas que se han propuesto para el tratamiento de la morbi-mortalidad de enfermedades cuando estas tienen poca frecuencia de ocurrir y las áreas geográficas son pequeñas.

Inicialmente el método que se empleó fue el ajuste de tasas, que se hace de dos formas, la directa y la indirecta, prefiriéndose la utilización de esta última cuando las áreas geográficas son pequeñas y las enfermedades son raras, mediante el cálculo de las Standardized Mortality Ratios (SMR). Para evaluar la significación estadística de los efectos estimados surge lo que se conoce como aproximación estadística clásica al estudio de las variaciones geográficas de la mortalidad o incidencia de enfermedades, estas estimaciones por máxima verosimilitud del riesgo relativo bajo un modelo de Poisson de los casos observados (cantidad de defunciones) van a coincidir con los SMR calculados.

Este enfoque clásico presenta muchas deficiencias puesto que el modelo no es capaz de explicar toda la variabilidad del riesgo relativo a través del mapa. Por lo que constituyó un nuevo reto para los investigadores darle solución a este problema, lo cual fue posible de lograr gracias al advenimiento del enfoque bayesiano en las técnicas estadísticas.

En los últimos 15 años se ha podido avanzar en la reformulación y tratamiento del cálculo y mapeo del riesgo relativo para la morbi-mortalidad de enfermedades a través del enfoque bayesiano y gracias al desarrollo que se ha obtenido en la ciencia de la computación. Ya hoy día se cuenta con un enfoque que da solución a todas las deficiencias del enfoque clásico sobre este problema y esta siendo muy aplicado en los países desarrollados como Italia, Francia, Canadá y Estados Unidos.

En nuestro país el Ministerio de Salud Pública tiene como objetivo primario la búsqueda de técnicas novedosas para el tratamiento de enfermedades raras y áreas pequeñas como es el caso de los municipios del país, lo cual motivó la realización de este trabajo de diploma, donde nos planteamos como objetivos, realizar una revisión bibliográfica extensa y profunda de ambas metodologías, permitiéndonos hacer una comparación detallada de las mismas para poder conformar un material de estudio y consultas para los especialistas cubanos y el otro objetivo es dar solución a un problema muy importante para la Salud Pública como es la determinación del

riesgo relativo y los patrones de su distribución geográfica en el mapa, para las defunciones por las enfermedades cerebrovasculares ocurridas en todos los municipios del país en el año 2002.

Para lo anterior, el trabajo se estructuró en tres capítulos. En el primer capítulo trataremos las metodologías usadas para el cálculo del riesgo relativo en la morbi-mortalidad de enfermedades y el mapeo de su distribución geográfica, en el segundo capítulo se estudia el sistema computacional WinBUGS 1.4 para el tratamiento de la metodología bayesiana con el GeoBUGS para el mapeo de la distribución geográfica del riesgo relativo, aquí se trata de que se logre una mayor identificación entre el especialista y el sistema de cómputo empleado. En el tercer capítulo se discuten los resultados obtenidos de aplicar ambas metodologías a un estudio sobre la distribución del riesgo relativo de las defunciones por las enfermedades cerebrovasculares en todos los municipios del país con el objetivo de determinar cuáles son los de alto riesgo para esta mortalidad y así poder dar una atención especial a estas áreas a través de los programas de salud de nuestro país.

# CAPÍTULO 1

## Metodología para el cálculo del riesgo relativo para la morbi-mortalidad de enfermedades. Mapeo de su distribución geográfica

### 1.1- Introducción

La estadística esta basada en la Teoría de Probabilidades. Formalmente la probabilidad es una función que cumple con ciertas condiciones, pero en general puede entenderse como una medida o cuantificación de la incertidumbre. Aunque la definición de función de probabilidad es una, existen varias interpretaciones de la probabilidad (Nieto, L. E. 2001):

**Clásica:** Supone que el experimento aleatorio produce resultados igualmente verosímiles (posibles) y propone como medida de probabilidad el cociente entre los casos favorables y los casos totales,

$$P(A) = \frac{N_A}{N}$$

**Frecuentista:** Supone que un experimento aleatorio puede ser repetido un número infinito de veces bajo condiciones similares y propone como medida de probabilidad la proporción de veces que ocurrió el evento de interés,

$$P(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N}$$

**Subjetiva:** La probabilidad de un evento  $A$  es una medida del grado de conocimiento de alguien en la ocurrencia de  $A$ . Es decir, la probabilidad de un evento está condicionada a la información disponible sobre él. Es un juicio personal sobre un evento único. Esta probabilidad debe ser coherente (obedecer las leyes de la probabilidad) y debe tener alguna relación con la realidad.

La metodología Bayesiana está basada en la interpretación subjetiva de la probabilidad y tiene como punto central el Teorema de Bayes.

Los métodos Bayesianos (llamados así debido al Reverendo y matemático inglés Thomas Bayes) proveen alternativas que permiten combinar información *a priori* sobre un parámetro poblacional con la información contenida en una muestra para conducir el proceso de inferencia estadística. Una distribución de probabilidad para el parámetro de interés es especificada primero. Luego es

obtenida la información muestral y son combinadas a través del teorema de Bayes para obtener una distribución de probabilidad *a posteriori*.

## 1.2- Conceptos de interés.

Ahora veremos algunos **conceptos básicos** (Bernardo, J. M. 1981), con que se nutre la estadística bayesiana y en especial su razón de ser: **el teorema de Bayes**.

### - Parámetros de interés $\beta$

Cantidades aleatorias en cuyos valores estamos interesados, las mismas pueden ser clasificadas como cantidades aleatorias discretas o continuas. Se denotarán con la letra griega  $\beta$ .

### - Distribuciones a priori de $\beta$

Medida de probabilidad que expresa la información disponible inicialmente sobre  $\beta$ , describiendo el grado de creencia del investigador en la ocurrencia de los distintos valores posibles de  $\beta$ , antes de tomar en cuenta los datos, la misma será su distribución de probabilidad.

Si  $\beta$  es una cantidad aleatoria discreta, su distribución de probabilidad puede ser descrita mediante la correspondiente función de probabilidad  $p(\beta) = \{p_1, p_2, \dots\}$ .

Si  $\beta$  es una cantidad aleatoria continua su distribución de probabilidad puede ser descrita mediante la correspondiente función de densidad de probabilidad  $\rho(\beta)$ .

Las distribuciones a priori pueden ser caracterizadas de la siguiente manera (Spiegelhalter, D. 2001):

- Son en principio subjetivas.
- Deben ser sacadas por expertos.
- Pueden ser más convincentes si están basadas en datos existentes. Pero la relación asumida es siempre un juicio subjetivo.
- Los arquetipos de a priori que expresan escepticismo y entusiasmo son útiles.
- Estas pueden ser asumidas desconocidas y clasificadas como a priori no informativas.

La distribución a priori es la distribución probabilística que refleja y sintetiza la idea acerca de lo que dicha persona razonablemente piensa antes de observar los datos. Está claro que la realidad objetiva es de determinada manera y también es obvio que no se comporta aleatoriamente, pero

nuestra visión sobre ella sí, y en gran medida depende de nuestra comprensión del problema y sobre todo, del conocimiento general prevaleciente en el momento del estudio.

Este conocimiento a priori acerca de  $\beta$  puede ser usado para especificar una densidad a priori con una forma funcional particular. Una familia paramétrica de densidades puede ser definida. Aunque muy a menudo estas familias pueden hacerse fácilmente del análisis, debemos ser cuidadosos y estar seguros que la densidad escogida realmente representa la información disponible. Los parámetros de estas formas funcionales, se conocen como hiperparámetros, y deben ser especificados en correspondencia con la información subjetiva disponible (Benavides, A. et al, 2002).

### **- Función de verosimilitud**

Con objeto de mejorar nuestra información sobre los parámetros de interés  $\beta$ , puede realizarse un experimento cuyo resultado  $Y$  es una cantidad aleatoria con una distribución  $p(Y/\beta)$ , tal observación proporcionará, indirectamente, información sobre el valor de  $\beta$ .

Cuando es considerada como una función de  $\beta$  para un  $Y = Y_0$  fijo  $p(Y_0/\beta)$  es una función que describe la verosimilitud de los distintos valores de  $\beta$  a la luz del resultado experimental observado  $Y_0$ , por lo que es llamada función de verosimilitud. Los valores de  $\beta$  que hacen grande  $p(Y_0/\beta)$  son aquellos que hacían más plausibles a priori la observación del resultado  $Y_0$ , estos valores resultan más verosímiles que los demás. Esta es una función positiva de  $\beta$ , no una función de probabilidad, en consecuencia, ni la suma de sus valores si  $\beta$  es discreta, ni su integral si es continua, tienen por qué ser la unidad (Bernardo, J. M. 1981).

En general la verosimilitud intenta reducir la incertidumbre obteniendo más información sobre su valor a la vez que relaciona todas las variables en un modelo de probabilidad completo.

### **- Distribución predictiva**

Sea  $p(Y/\beta)$  la distribución del resultado  $Y$  de un determinado experimento. Generalmente el valor de  $\beta$  es desconocido y en consecuencia,  $p(Y/\beta)$  puede no utilizarse para determinar los valores de  $Y$  que resultan más probables. Sin embargo aunque el valor exacto de  $\beta$  es desconocido, siempre se dispone de cierta información sobre  $\beta$ , que es la descrita por su distribución a priori  $p(\beta)$ . Esta distribución puede combinarse con la distribución  $p(Y/\beta)$  para

poder describir la información que se posee sobre los posibles valores de  $Y$ . En efecto: la función de probabilidad:

$$p(Y) = \sum_i p(Y/\beta_i)p(\beta_i) \text{ si } \beta \text{ es discreta} \quad (1.2.1)$$

y la función de densidad de probabilidad

$$p(Y) = \int p(Y/\beta)p(\beta)d\beta \text{ si } \beta \text{ es continua} \quad (1.2.2)$$

proporcionan una distribución de  $Y$  totalmente conocida que podemos utilizar entre otras cosas, para hacer predicciones sobre los valores de  $Y$  a que dará lugar el experimento, por lo que recibe el nombre de distribución predictiva a priori. A priori porque no es condicional sobre una observación previa del proceso y predictiva porque es la distribución para una cantidad que es observable.

Las expresiones (1.2.1) y (1.2.2) son una forma del Teorema de la Probabilidad Total.

#### - Teorema de Bayes y distribución a posteriori

Con objeto de mejorar la información de que se dispone sobre el valor de  $\beta$ , se realiza un experimento que da lugar a unos resultados  $Y$  con probabilidad (o densidad de probabilidad si  $Y$  es una cantidad aleatoria continua)  $p(Y/\beta)$ . Sea  $p(\beta)$  la distribución a priori de  $\beta$ . Después de observar el resultado  $Y$  del experimento, la información de que disponemos sobre el valor de  $\beta$  estará descrita por su distribución a posteriori  $p(\beta/Y)$ .

El teorema de Bayes permite obtener la distribución a posteriori  $p(\beta/Y)$  a partir de la distribución a priori  $p(\beta)$  y de la función de verosimilitud del resultado obtenido  $p(Y/\beta)$ .

**Teorema de Bayes:** Sea  $Y$  los resultados del experimento definido mediante el modelo  $p(Y/\beta)$  y sea  $p(\beta)$  la distribución a priori de  $\beta$ . La distribución a posteriori es entonces:

$$p(\beta/Y) = \frac{p(Y/\beta)p(\beta)}{p(Y)} \quad (1.2.3)$$

donde  $p(Y)$  es la distribución predictiva de  $Y$ .

El teorema anterior es una generalización natural a funciones de probabilidad o de densidad de probabilidad del teorema de Bayes para sucesos.

Sin pérdida de generalidad el teorema de Bayes puede expresarse en la forma:

$$p(\beta/Y) \propto p(Y/\beta)p(\beta) \quad (1.2.4)$$

de manera que si dos modelos probabilísticos son proporcionales para todo  $\beta$ , dan necesariamente lugar a la misma distribución a posteriori.

La constante de proporcionalidad puede determinarse en cualquier momento utilizando la propiedad de que, por ser  $p(\beta/Y)$  una distribución de probabilidad,  $\int p(\beta/Y) d\beta = 1$  en el caso continuo, y  $\sum p(\beta_i/Y) = 1$  en el discreto.

Específicamente, si  $p(\beta/Y) = C p(Y/\beta) p(\beta)$  entonces:

$$\int p(\beta/Y) d\beta = C \int p(Y/\beta) p(\beta) d\beta = 1$$

$$C = 1/\int p(Y/\beta) p(\beta) d\beta = 1/p(Y) \text{ y análogamente:}$$

$$\sum p(\beta_i/Y) = C \sum p(Y/\beta_i) p(\beta_i) = 1$$

$$C = 1/\sum p(Y/\beta_i) p(\beta_i) = 1/p(Y)$$

La expresión (1.2.4) es generalmente más cómoda de manejar que la expresión (1.2.3).

Con los conceptos vistos anteriormente ya estamos en condiciones de describir mediante una distribución de probabilidad, la información a priori que se posee sobre la verosimilitud de su ocurrencia y determinar mediante el teorema de Bayes, la distribución a posteriori que describe la información que se tiene sobre los sucesos inciertos tras incorporar a la información a priori la que proporcionan los resultados experimentales.

De modo que se tiene que las expresiones vistas anteriormente encierran la técnica central de la inferencia Bayesiana.

### 1.3- Metodología clásica para el cálculo del riesgo relativo para la morbi-mortalidad de enfermedades y el mapeo de su distribución geográfica

El análisis de las variaciones geográficas en la proporción de la mortalidad o incidencia de enfermedades es muy útil en la formulación de las hipótesis etiológicas. El objetivo del mapeo de enfermedades es explicar la distribución geográfica que esta detrás de la proporción de enfermedades, y para identificar áreas con bajas o altas proporciones.

Como es sabido, las tablas de tasas por grupos de edad y sexo no son manejables para el investigador debido a la cantidad de datos que contienen, de ahí la necesidad de buscar recursos expresivos más sencillos, fácilmente interpretables, que ayuden en la toma de decisiones. La representación en mapas proporciona una imagen más intuitiva de la distribución geográfica de las enfermedades.

El uso de las tasas brutas puede llevar a conclusiones erróneas, puesto que las diferencias entre las poblaciones son imputables, no sólo a la intensidad del fenómeno, sino a confusores, de los cuales el más clásico es la estructura por edad, que pueden diferir sensiblemente de una población a otra.

De modo que, con el objetivo de neutralizar el efecto de las distintas composiciones por edad en el cálculo de las tasas brutas, se realiza el ajuste de tasas por edad y para ello se pueden manejar dos métodos muy utilizados por los especialistas de la Salud (Benavides, A. et al, 2002)

- **Método Directo:**

Consiste en aplicar las tasas específicas de mortalidad en la población estudiada a una población hipotética denominada población estándar, es decir, denotando por  $Y_j$  el número de defunciones en el j-ésimo grupo de edad, por  $n_j$  las personas por años, entonces  $t_j = Y_j/n_j$  es la tasa de mortalidad en cada grupo de edad j y denotando por  $w_j$  la proporción de individuos en el grupo de edad j en la población estándar; entonces la tasa ajustada se define como la media ponderada de las tasas específicas por edad:

$$T = \sum_j w_j t_j \quad (1.3.1)$$

- **Método indirecto:**

Compara el número observado de muertes con el número esperado que se obtiene al aplicar las tasa estándar a la estructura de edad de la población objeto de estudio. Esto es lo que se conoce por razón de mortalidad estándar (SMR).

La SMR se define como una media ponderada de las razones de las tasas específicas, donde  $Y_j$  denota el número de muertes observadas en el j-ésimo grupo de edad,  $n_j$  es el total de personas por años,  $t_j^*$  representa la tasa de mortalidad en cada grupo de edad y los pesos  $w_j = n_j t_j^*$  son el número esperado de muertes en el j-ésimo grupo de edad

$$SMR = \frac{\sum_j Y_j}{\sum_j n_j t_j^*} = \frac{Y}{E} \quad (1.3.2)$$

donde  $Y$  denota el número total de muertes observadas y  $E$  el número de muertes esperadas.

La SMR es más apropiado cuando la muestra es pequeña, lo cual ocurre cuando la enfermedad es rara o se trabaja con unidades geográficas pequeñas. En estos casos, los SMR gozan de más estabilidad que las tasas calculadas por el método directo.

Debido a la poca consistencia desde el punto de vista estadístico de estos métodos, directo e indirecto, se decidió realizar una aproximación estadística clásica al estudio de las variaciones geográficas en la proporción de la mortalidad o incidencia de enfermedades, utilizando estimaciones de máxima verosimilitud del riesgo relativo bajo un modelo de Poisson de los casos observados (cantidad de defunciones), las cuales van a coincidir con las SMRs calculadas para cada área. Esta aproximación es lo que se conoce como metodología clásica para la determinación del riesgo relativo para la morbi-mortalidad de enfermedades.

El mapa es dividido en  $N$  áreas contiguas  $i=1, \dots, N$ . Sea  $Y = \{Y_1, \dots, Y_N\}$  que denota el número de muertes por la enfermedad de interés durante el período estudiado. Los valores esperados de las muertes  $E = \{E_1, \dots, E_N\}$  van a ser conocidos, constantes durante el periodo de estudio y calculados aplicando la tasa específica por edad y sexo a la población de riesgo en un área. Denotando  $\xi_i$  como el riesgo relativo desconocido para la i-ésima área genérica. Suponiendo que los  $\xi_i$  son mutuamente independientes y que los  $E_i$  son conocidos. Además que cada  $Y_i$  se distribuye de acuerdo a una distribución Poisson con media,  $E_i \xi_i$ , la cual se escribirá como:

$$[Y_i / E_i, \xi_i] \sim Po(Y_i; E_i \xi_i) \quad (1.3.3)$$

La verosimilitud del riesgo relativo  $\xi$  es:

$$f(Y_i/E_i, \xi) = \frac{(E_i \xi)^{Y_i}}{Y_i!} \lambda^{-E_i \xi} \quad (1.3.4)$$

La estimación máximo verosímil es como sigue:

$$\begin{aligned} V(\xi_i, Y_i) &= C(Y_i) \prod_{i=1}^N \frac{(E_i \xi_i)^{Y_i} \lambda^{-E_i \xi_i}}{Y_i!} \\ &= C(Y_i) \frac{\prod_{i=1}^N (E_i \xi_i)^{Y_i} \lambda^{-\sum_{i=1}^N E_i \xi_i}}{\prod_{i=1}^N Y_i!} \end{aligned} \quad (1.3.5)$$

tomando  $C(Y_i) = \prod_{i=1}^N Y_i!$  tenemos entonces

$$V(\xi_i, Y_i) = \left[ \prod_{i=1}^N (E_i \xi_i)^{Y_i} \right] \lambda^{-\sum_{i=1}^N E_i \xi_i} \quad (1.3.6)$$

aplicando el logaritmo natural nos queda:

$$\begin{aligned} lv(\xi_i, Y_i) &= \ln \left[ \prod_{i=1}^N (E_i \xi_i)^{Y_i} \right] - \sum_{i=1}^N E_i \xi_i \\ &= \sum_{i=1}^N Y_i \ln(E_i \xi_i) - \sum_{i=1}^N E_i \xi_i \\ &= \sum_{i=1}^N Y_i \ln E_i + \sum_{i=1}^N Y_i \ln \xi_i - \sum_{i=1}^N E_i \xi_i \end{aligned} \quad (1.3.7)$$

$$\text{derivando nos queda: } lv'(\xi_i) = \frac{Y_i}{\xi_i} - E_i = 0 \text{ para cada } i = 1, \dots, N \quad (1.3.8)$$

por lo que el estimador máximo verosímil de  $\xi_i$  es  $\hat{\xi}_i = Y_i/E_i$  coincide con los SMR (razón de mortalidad estándar) para la i-ésima área (Bernardinelli, L. et al, 1995)

Sin embargo, el método de Poisson tiene desventajas debido a que estos pueden resultar notablemente dependientes de los tamaños poblacionales, con el consecuente impacto sobre la variabilidad de las estimaciones, fenómeno que ocurre con más frecuencia cuando las áreas geográficas sobre las que se quieren representar son pequeñas. Dicho de otro modo: cuando las enfermedades son raras y las áreas pequeñas, la estimación de máxima verosimilitud del modelo dado puede producir estimaciones inestables y poco informativas de las tendencias específicas de

cada área: típicamente, las estimaciones extremas (las cuales dominarían el mapa) tienden a ubicarse en aquellas áreas cuyos tamaños poblacionales son más pequeños (Bernardinelli, L. et al, 1992)

El acercamiento basado en los SMRs ha sido criticado por varios autores, notablemente Clayton y Kaldor, quienes señalan que los mapas de los SMRs están altamente afectados por el “ruido”, que es debido a la variabilidad de las estimaciones. Esto es particularmente evidente cuando enfermedades raras son investigadas en áreas pequeñas. Por lo tanto las características más sobresalientes en el mapa son con frecuencia contribuidas por los datos más pequeños y por lo tanto no daría una interpretación epidemiológica muy útil.

La variabilidad del riesgo relativo dentro de cada área excede a la que se podía esperar mediante una distribución Poisson. Dadas estas debilidades de la modelación de Poisson se plantea la necesidad de buscar un acercamiento alternativo que nos permita superar los problemas de sobredispersión. Esto nos llevará hacia la modelación Bayesiana, la cual da estimaciones más suavizadas de los riesgos relativos (Lawson, A. B. 1995).

#### **1.4- Enfoque Bayesiano al mapeo de la variación geográfica de enfermedades. Análisis Bayesiano Puro**

Una vez expuestas las debilidades de la modelación Poisson se plantea la necesidad de buscar un acercamiento alternativo que nos permita superar los problemas de la sobredispersión.

Este nuevo reto nos lleva hacia la modelación Bayesiana, la cual da estimaciones suavizadas de los riesgos relativos. Incluso si el SMR es la mejor estimación del riesgo relativo los métodos Bayesianos producen conjuntos de estimaciones con menor error estándar que los SMR crudos. Un acercamiento Bayesiano a la modelización de los riesgos relativos, el cual fue propuesto por Clayton y Kaldor, evita la inestabilidad del mapeo de los SMR crudos. La idea de Clayton y Kaldor fue imponer una estructura espacial a los riesgos relativos modelándolos colectivamente como un proceso estocástico espacial (Bernardinelli, L. et al, 1995)..

La modelación Bayesiana tiene la ventaja de que al estimar el riesgo de un área pequeña tienen como idea central utilizar la información que le brindan otras áreas, sobre todo las más vecinas, las unidades espaciales vecinas definidas por criterios de adyacencias. Al basarse en esta información disminuyen el efecto de las fluctuaciones aleatorias no asociadas al riesgo y

permiten ofrecer mapas suavizados. También contribuye a esto el hecho de que se toma en cuenta la correlación espacial entre áreas vecinas y los mapas resultantes son más informativos (Lawson, A. B. et al,1997).

Estos modelos parten de la consideración que la variación del comportamiento del evento en el mapa tiene dos fuentes: una es la variación Poisson (representada por el uso de la Razón estandarizada de Mortalidad (SMR) y otra por la variación extrapoisson, la cual explica el resto de la variación inexplicada en el modelo.

El tratamiento bayesiano permite representar estas variaciones en un modelo jerárquico en el cual aparecen representados varios términos lineales: uno para la *media global del mapa* y otro para los *efectos del área*.

El modelo que describiremos es muy semejante al que propuso Clayton y Kaldor. Como ya habíamos visto en el enfoque clásico, los  $Y_i$  son mutuamente independientes y se distribuyen de acuerdo a una distribución de Poisson con media,  $E_i \xi_i$ .

$$[Y_i / E_i, \xi_i] \sim Po(Y_i; E_i \xi_i) \quad (1.4.1)$$

Ahora comenzaremos denotando  $\eta_i$  como el logaritmo natural del riesgo relativo desconocido para el área i:

$$\eta_i = \ln(\xi_i) \quad (1.4.2)$$

$\eta_i$  puede ser modelado como la suma de una media global, denotada por  $\mu$ , que expresa el nivel global de el log-riesgo relativo por todo el mapa, y un efecto específico-área, denotado por  $\phi_i$ , que representa la diferencia entre el log-riesgo relativo para el área i y la media global. Por lo tanto para la i-ésima área genérica podemos escribir:

$$\log(\xi_i) = \eta_i = \mu + \phi_i \quad (1.4.3)$$

y rescribimos (1.4.1) como sigue:

$$[Y_i / E_i, \mu, \phi_i] \sim Po(Y_i; E_i \exp(\mu + \phi_i)) \quad (1.4.4)$$

La cantidad  $\exp(\mu)$  representa realmente la media geométrica del riesgo relativo sobre el mapa. Los efectos  $\phi_i$  pueden ser vistos como sustitutos para las variables desconocidas o no observadas que afectan al riesgo.

Si fuera posible observar las variables  $\phi_i$  algunas manifestarían estructura no geográfica mientras otras manifestarían altos patrones de estructura geográfica de variación. Si las primeras dominan,

los  $\phi_i$ 's verdaderos manifestarán una heterogeneidad no estructurada, mientras que si en cambio dominan las últimas, los  $\phi_i$ 's verdaderos manifestarán heterogeneidad estructurada geográficamente, en que los valores de riesgos respectivos para un par de áreas vecinas serán generalmente más semejantes que para un par de áreas arbitrarias (Bernardinelli, L. et al, 1992).

El acercamiento Bayesiano requiere que sea especificada una distribución priori multivariante para los  $\phi_i$ 's. Esta reflejará la creencia priori acerca del tipo de heterogeneidad geográfica de los  $\phi_i$ 's. El compromiso para una hipótesis de heterogeneidad no estructurada producirá una priori suponiendo que los  $\phi_i$ 's son independientes e idénticamente distribuidos. Esta suposición en sí misma implica un modelo priori intercambiable (EX). Por otra parte, un compromiso para una hipótesis de heterogeneidad estructurada producirá la incorporación en el modelo priori de una correlación espacial para los  $\phi_i$ 's, lo cual implica un modelo priori autorregresivo condicional (CAR). Llamaremos a lo planteado anteriormente modelos priori simples, puesto que ellos modelan cualquiera de las dos, heterogeneidades estructurada y no estructurada.

Cuando carecemos de una idea priori como para el tipo de heterogeneidad de los efectos de áreas, uno "mixto", más bien que un modelo priori simple de su variación puede ser adoptado. En el modelo mixto, descrito en Clayton y Bernardinelli, un parámetro extra que es estimado de los datos, elige entre los tipos de variación estructurados y no estructurados. Para simplificar, en este trabajo consideramos una aproximación en la cual el investigador solamente usa modelos priori simples. El conocimiento priori epidemiológico proporciona formas para elegir el modelo priori simple apropiado, por ejemplo si creemos que los determinantes del riesgo varían en una escala que es pequeña en comparación con el tamaño promedio de las áreas, podemos asumir que los  $\phi_i$ 's están incorrelacionados, y un modelo de intercambiabilidad será apropiado. Si no obstante creemos que estos factores varían en una escala grande, un modelo de correlaciones espaciales será apropiado. Entonces trataremos con modelos priori simples en que un parámetro desconocido, denotado por  $\lambda$  (oficialmente un parámetro escalar de la densidad conjunta de los efectos de áreas) representan la variabilidad geográfica. Estos parámetros controlan la cantidad de variación en la distribución del riesgo por todo el mapa (Bernardinelli, L. et al, 1995, Lawson, A. B. et al, 1997).

Ahora, el caso de muerte observado genérico  $Y_i$  se acepta como una variable Poisson con valor esperado (esperanza)  $E_i \exp(\mu + \phi_i)$ , como se propuso en (1.4.4).

Condicionales en  $\phi$  y  $E$ , las  $Y$ 's son mutuamente independientes.

Por lo tanto la distribución condicional  $[Y/E, \mu, \phi]$ , que representa la verosimilitud de los datos que está dado por el producto:

$$[Y/E, \mu, \phi] = \prod_{i=1}^N [Y_i/E_i, \mu, \phi_i] = \prod_{i=1}^N Po(Y_i; E_i \exp(\mu + \phi)) \quad (1.4.5)$$

Si llamamos  $mu_i = E_i \xi_i$  entonces teniendo en cuenta que  $\xi_i = \exp(\mu + \phi_i)$  se tiene:

$$\ln(mu_i) = \ln E_i + \mu + \phi_i \quad (1.4.6)$$

El enfoque bayesiano requiere que se especifique una distribución a priori para los  $\phi_i$  dependiente de un hiperparámetro  $\lambda$ .

La especificación de la densidad condicional  $[\phi/\lambda]$  que está asociada con el parámetro  $\phi$  expresa nuestra creencia priori concerniente a la colección de los efectos de áreas desconocidos, vistos como un proceso espacial. Como una especificación conveniente para  $[\phi/\lambda]$  elegimos una densidad para la familia:

$$[\phi/\lambda] \propto \exp \left[ -\frac{\lambda}{2} \sum_{i=1}^N w_{i+} \phi_i \left( \phi_i - \frac{1}{w_{i+}} \sum_{j=1}^N w_{ij} \phi_j \right) \right] \quad (1.4.7 a)$$

$$\text{con: } w_{ii} = 0 \quad i = 1, \dots, N \quad (1.4.7 b)$$

donde  $\lambda$  actúa como un hiperparámetro, los  $w_{ij}$  son asignados pesos no-negativos, y el índice “i+” denota la suma sobre j.

Notemos que (1.4.7 a) depende de las diferencias respectivas entre cada  $\phi_i$  y una media de los pesos de todas las otras  $\phi$ 's. Una posible selección fuera de la diagonal de los pesos  $w_{ij}$  es especificada:

$$w_{ij} = 1 \quad i = 1, \dots, N \quad j = 1, \dots, N \quad i \neq j \quad (1.4.7 c)$$

Esto posibilita inducir intercambiabilidad en el riesgo relativo de las diferentes áreas geográficas, en el sentido que da (1.4.7 c) en el lado derecho de (1.4.7 a) queda invariable bajo una permutación de  $\phi_1, \phi_2, \dots, \phi_N$ , así bajo (1.4.7 c), (1.4.7 a), (1.4.7 b) define una priori de intercambiabilidad (EX priori), y se simplifica a:

$$[\phi/\lambda]_{EX} \propto \exp\left[-\frac{\lambda}{2} \sum_{i=1}^N (N-1)\phi_i(\phi_i - \bar{\phi}_{-i})\right] \quad (1.4.8)$$

(donde  $\bar{\phi}_{-i} = \frac{1}{N-1} \sum_{j \neq i} \phi_j$  representa la media aritmética de todos los efectos de área excluyendo  $\phi_i$ )

El modelo priori EX (1.4.8) supone que la distribución priori condicional de los  $\phi_i$  individuales genéricos, dado todos los  $\phi$ 's restantes, es una Normal con media  $\bar{\phi}_{-i}$  y varianza  $((N-1)\lambda)^{-1}$ , que es:

$$[\phi_i/\phi_j, j \neq i, \lambda]_{EX} \sim N\left(\bar{\phi}_{-i}, \frac{1}{(N-1)\lambda}\right) \quad (1.4.9)$$

De (1.4.9) es evidente que el modelo priori EX tiende a desplazar las estimaciones de los efectos de áreas individuales hacia una media global. Además queda claro de (1.4.9) que la probabilidad condicional de cada log-efecto de área dado todos los otros es invariante con respecto a adicionar una constante a todos log-efectos de áreas. Significando que la densidad priori EX (1.4.8) es uniforme con respecto a la suma de las  $\phi$ 's. En terminología Bayesiana (1.4.8) es llamada distribución impropia.

La intercambiabilidad estricta debe ser descartada cuando hay creencias priori de correlaciones espaciales en los riesgos relativos. La estructura espacial puede ser incorporada en (1.4.7) por elección de valores altos para  $w_{ij}$  cuando las áreas i y las áreas j son gráficamente cercanas (donde “cercanas” puede ser definido de varias maneras). Un extremo escogido en esta dirección tiene  $w_{ij}=1$ , si i y j son áreas contiguas, y  $w_{ij}=0$ , en otro caso.

En este caso (1.4.7) produce la priori Autorregresiva Condicional (CAR priori)

$$[\phi/\lambda]_{CAR} \propto \exp\left[-\frac{\lambda}{2} \sum_{i=1}^N n_i \phi_i (\phi_i - \bar{\phi}_i)\right] \quad (1.4.10)$$

donde  $n_i$  representa el número de áreas contiguas a la i-ésima área y  $\bar{\phi}_i = \frac{1}{n_i} \sum_{j=1}^N w_{ij} \phi_j$  es la media aritmética de los log efectos de áreas que son contiguos a la i-ésima área. El modelo priori CAR

(1.4.10) supone que la distribución condicional del efecto individual genérico  $\phi_i$  dado todos los restantes  $\phi_j$ 's, es Normal con media  $\bar{\phi}_i$  y varianza  $\frac{1}{\lambda n_i}$ , que es:

$$[\phi_i / \phi_j, j \neq i, \lambda]_{CAR} \sim N\left(\bar{\phi}_i, \frac{1}{\lambda n_i}\right) \quad (1.4.11)$$

De (1.4.11) nos parece que el modelo priori CAR tiende a desplazar cada estimación de los efectos de áreas individual hacia un “efecto medio” local  $\bar{\phi}_i$ .

Por las mismas razones dadas en relación al modelo priori EX, también la priori CAR es impropia.

De (1.4.9) y (1.4.11) parece que  $\lambda$  es inversamente proporcional a la varianza priori del log-efecto de área genérico  $\phi_i$ . Nosotros elegimos una distribución Chi-Cuadrado, puesto que garantiza la conjugación con respecto a la distribución condicional Normal  $[\phi/\lambda]$ . Más precisamente, para un número positivo dado  $\lambda^*$ , asumamos que la cantidad  $\nu\lambda/\lambda^*$  está distribuida como una Chi-Cuadrado con  $\nu$  grados de libertad, tal que:

$$[\lambda] \sim \frac{X_\nu^2 \lambda^*}{\nu} \quad (1.4.12)$$

Puesto que el valor esperado para la distribución anterior es  $\nu$ , de acuerdo a (1.4.12)  $[\lambda]$  tiene valor esperado  $\lambda^*$ . De este modo,  $\lambda^*$  representa una conjetura, suposición priori para  $\lambda$ . Además de (1.4.12) se tiene que  $\text{var}(\lambda) = 2\lambda^{*2}/\nu$ , de lo cual parece que aumenta  $\nu$  para una priori fuerte, robusta alrededor de  $\lambda^*$ . En otras palabras, el parámetro  $\nu$  representa el grado de confianza en  $\lambda^*$ .

La conjetura  $\lambda^*$  depende de si elegimos modelos priori EX o CAR. Primero consideremos el modelo EX donde se tiene de (1.4.9) que  $((N-1)\lambda)^{-1}$  es la varianza priori de la distribución (normal) de cada  $\phi_i$ , y por lo tanto de cada uno de los log-riesgo relativo  $\eta_i$ . En ausencia de creencias priori esos riesgos son más altos y bajos que la referencia, es razonable suponer que la densidad priori de el i-ésimo log-riesgo relativo,  $[\eta_i]_{EX}$ , es simétrica alrededor de 0.

La expresión (1.4.11) sugiere que bajo el modelo CAR una suposición razonable  $\lambda_{CAR}^*$  para  $\lambda$  es dada por  $\lambda_{EX}^*$  dividida por un estimador de el valor promedio para  $n_i$ . Por ejemplo si asumimos que, en promedio, un área tiene 5 áreas adyacentes, la regla anterior produce,  $\lambda_{CAR}^* = \lambda_{EX}^*/5$ . A causa de que es razonable asumir que nuestra ignorancia de  $\mu$  es total, le asignemos a  $\mu$  una priori impropia vaga, a saber una densidad uniforme no informativa en la recta real:

$$[\mu] \sim U(-\infty, +\infty) \quad (1.4.13)$$

Las condicionales (1.4.6), (1.4.7), (1.4.12), (1.4.13) especifican la posterior conjunta para  $[\phi, \lambda, \mu/Y, E]$  de las cantidades desconocidas en el modelo. Más precisamente, sobre la base de la estructura de independencia condicional, esta distribución puede ser factorizada en el producto:

$$[\phi, \lambda, \mu/Y, E] = [Y/E, \phi, \mu] \times [\phi/\lambda] \times [\lambda] \times [\mu] \quad (1.4.14)$$

En el análisis Bayesiano, el cálculo de la posterior deseada  $[\phi, \lambda, \mu/Y, E]$  implica la integración de (1.4.14) sobre  $\lambda$ . Esta integración es analíticamente tratable debido a los recientes desarrollos en la ciencia de la computación. La idea es la de muestrear la distribución posterior conjunta del riesgo relativo. Ideas innovativas recientes, basadas en estrategias de muestreo y remuestreo que propone el uso de una técnica Monte Carlo, llamada muestreador de Gibbs, para generar muestras de la posterior  $[\phi, \lambda, \mu/Y, E]$  (Bernardinelli, L. et al, 1992).

Un sistema computacional creado para el análisis complejo bayesiano es el WinBUGS, el cual tiene implementadas estas técnicas de simulación y toda la teoría que veremos a continuación para la estimación de los parámetros del modelo y el análisis de la convergencia.

#### **- Estimación de los parámetros del modelo.**

Las distribuciones a posteriori necesarias no siempre pueden ser obtenidas analíticamente y las aproximaciones asintóticas son solo apropiadas bajo ciertas circunstancias. Por ello se utilizan métodos Cadena de Markov-Integración Monte Carlo (MCMC), específicamente el algoritmo Gibbs Sampler (muestreo Gibbs) para simular las realizaciones de la distribución posterior conjunta. Ha sido demostrado que el método MCMC tiene un potencial enorme para los análisis Bayesianos de modelos estadísticos complejos, los cuales no pueden ser tratados con aproximaciones numéricas o analíticas (Natarajan, R. et al, 2000).

A continuación describiremos como se realiza todo este proceso.

- **Simulación MCMC.**

Como mencionamos anteriormente la inferencia posteriori Bayesiana puede ser realizada por vía de la integración Monte Carlo, usando valores simulados de todas las cantidades desconocidas  $\beta$  en el modelo generado de una Cadena de Markov con  $p(\beta/Y)$  como distribución estacionaria.

Este proceso parte de la distribución posteriori conjunta:

$$p(\beta/Y) \propto p(Y/\beta)p(\beta) \quad (1.4.15)$$

necesitamos inferencias sobre los parámetros simples  $\beta_h$  lo cual es realizado por medio de la posteriori conjunta (integrando) sobre todas las otras  $\beta_1, \beta_2, \dots, \beta_{h-1}, \beta_{h+1}, \dots, \beta_k$  desconocidas, es decir:

$$p(\beta_h/Y) = \int p(\beta/Y) d\beta_1 d\beta_2 \dots d\beta_{h-1} d\beta_{h+1} \dots d\beta_k \quad (1.4.16)$$

Puesto que la integración analítica o numérica es prácticamente insoluble para problemas de la vida real, la integración Monte Carlo será usada puesto que la misma es igual a simular la posteriori conjunta (Dey, D. K et al 1997).

- Integración Monte Carlo.

Supongamos que podemos sacar muestras de la distribución a posteriori conjunta para  $\beta$ , es decir:

$\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(n)}$   $p(\beta/Y)$ , entonces:

$$E(g(\beta)) = \int g(\beta) p(\beta/Y) d\beta \approx \sum_{i=1}^n g(\beta^{(i)}) \quad (1.4.17)$$

- ¿Cómo sacar muestras de la a posteriori?

Queremos muestras independientes de la distribución a posteriori conjunta  $p(\beta/Y)$ , lo cual resulta muy difícil, pero muestras independientes de una Cadena de Markov con  $p(\beta/Y)$  como distribución estacionaria (equilibrio) es fácil.

**Definición:** Una sucesión de variables aleatorias  $\beta^{(0)}, \beta^{(1)}, \dots$  en un espacio E de  $\mathfrak{R}^d$  forman una Cadena de Markov si la distribución de  $\beta^{(i+1)}$  condicional al pasado  $\beta^{(i)}, \beta^{(i-1)}, \dots$  es la distribución de  $\beta^{(i+1)}$  condicional a  $\beta^{(i)}$ .

Los Teoremas de existencias muestran que:  $\frac{1}{n} \sum_{i=1}^n f(\beta^{(i)}) \rightarrow E(f(\beta))$  cuando  $n \rightarrow \infty$  y cuando  $\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(n)}$  son muestras de una cadena de Markov conveniente.

El muestreo Gibbs (ver Geman and Geman (1984), Gelfand and Smith (1990), Casella and George (1992)) es un caso especial del algoritmo Metrópolis-Hasting el cual genera una cadena de Markov por muestreo de distribuciones condicionales completas (Spiegelhalter, D. 2001).

- Muestreo de Gibbs.

En la simulación de la distribución conjunta el algoritmo Gibbs Sampler comienza especificando valores iniciales para todas las cantidades no observadas en el modelo.

Sea nuestro vector de  $\beta$  desconocidos que consiste en k-subcomponentes  $\beta = (\beta_0, \dots, \beta_k)^T$ .

1- Empecemos seleccionando valores  $\beta_1^{(0)}, \beta_2^{(0)}, \dots, \beta_k^{(0)}$ .

2- Muestra  $\beta_1^{(1)}$  de  $p(\beta_1 / \beta_2^{(0)}, \beta_3^{(0)}, \dots, \beta_k^{(0)}, Y)$

Muestra  $\beta_2^{(1)}$  de  $p(\beta_2 / \beta_1^{(1)}, \beta_3^{(0)}, \dots, \beta_k^{(0)}, Y)$

...

Muestra  $\beta_k^{(1)}$  de  $p(\beta_k / \beta_1^{(1)}, \beta_2^{(1)}, \dots, \beta_{k-1}^{(1)}, y)$

3- Repetir paso 2 muchos miles de veces, a la larga finalmente se obtiene muestra de  $p(\beta/Y)$ .

Las  $\beta^{(n)}$  forman una Cadena de Markov con una distribución estacionaria  $p(\beta/Y)$ .

En resumen el método Gibbs Sampler parte de los valores iniciales asignados a los parámetros y va generando valores de los mismos, en el primer paso obtiene el del primer parámetro, teniendo en cuenta los valores iniciales de los restantes, en el segundo paso obtiene la estimación del segundo parámetro teniendo en cuenta la estimación actual del primer parámetro y los valores iniciales de los restantes, y así sucesivamente hasta formar una cadena de Markov.

Obtener los valores generados de una distribución de probabilidad es a menudo sencillo con modernas técnicas de computación que contamos en la actualidad para trabajar los métodos Bayesianos y muestra de ello es el sistema computacional utilizado en este trabajo (WinBUGS).

### - Diagnósticos de convergencia.

Debido a que los métodos de análisis son iterativos, se necesita evaluar su convergencia; para esto sólo necesitamos una Cadena de Markov para poder enfocar la convergencia para la integración Monte Carlo y como consecuencia producir una estimación consistente de la esperanza verdadera.

Existen varios diagnósticos de convergencia, como son:

- Geweke (1992)
- Gelman y Rubin (1992)
- Raftery y Lewis (1992)
- Heidelberger y Welch (1983)
- Autocorrelaciones
- Correlaciones cruzadas

Pero el de Gelman y Rubin ha demostrado ser en múltiples trabajos el más seguro de todos los diagnósticos discutidos en la literatura hasta el momento para monitorear la convergencia del Gibbs Sampler, por lo que será el que tendremos en cuenta en nuestro trabajo (Spiegelhalter, D. 2001).

Este test se basa en una comparación de las varianzas de cada variable y proporciona los cuantiles al 50% y al 97,5% de la distribución marginal a posteriori; si ambos cuantiles son aproximadamente 1 se diagnostica la convergencia. Dicho test está implementado en el programa BOA.

Nosotros monitoreamos la convergencia de la simulación iterativa para estimar el factor por el cual el escalar de la distribución corriente para  $\psi$  puede ser reducido si las simulaciones fueron continuadas en el  $\lim n \rightarrow \infty$  (Gelman, A. et al 1995).

Esta reducción escalar potencial es estimada por:

$$\sqrt{\hat{R}} = \sqrt{\frac{\hat{\text{var}}^+(\psi/Y)}{W}} \quad (1.4.18)$$

(que no es más que el cociente entre la varianza posterior marginal estimada y la varianza de las sucesiones “dentro”)

la cual declina a 1 cuando  $n \rightarrow \infty$ .

En la práctica si el valor de la reducción potencial escalar es alta (mayor que 1.2, esto indica convergencia pobre), entonces tenemos razón para creer que procediendo con simulaciones

adicionales podemos mejorar nuestra inferencia sobre la distribución utilizada (Lawson, A. B. et al 1997).

Después de la convergencia, se necesitan más iteraciones para obtener muestras de las inferencias a posteriori. Puesto que muchas iteraciones equivalen a muchas estimaciones puntuales a posteriori, la exactitud de las estimaciones a posteriori puede ser calculada por el error estándar Monte Carlo para cada parámetro, el cual es reportado por el sistema computacional empleado a través de la siguiente ecuación:

$$\text{MC error} = \frac{\sigma}{\sqrt{N}} \quad (1.4.19)$$

quedando:  $\sqrt{N}(\overline{f(\beta)}_N - E_x[f(\beta)]) \rightarrow \text{Normal}(0, \sigma^2)$ , es decir el MC error es proporcional a la desviación estándar de la diferencia entre la media de los valores muestreados de  $\beta$  y la media a posteriori verdadera.

Por lo que se debe correr la simulación hasta que el MC error sea pequeño en relación a la desviación estándar a posteriori

La mala selección de los valores iniciales puede causar problemas de convergencia (Spiegelhalter, D. 2001)

## CAPÍTULO 2

### **Sistema computacional utilizado en el cálculo del riesgo relativo para la morbi-mortalidad de enfermedades y de su representación geográfica en los mapas**

#### **2.1- Introducción**

Para el cálculo de riesgo relativo para la morbi-mortalidad de enfermedades utilizamos el sistema WinBUGS 1.4, sistema que se oferta libremente en Internet, desarrollado por Imperial College & Medical Research Council, United Kingdom.

El BUGS es un programa para la inferencia Bayesiana de los problemas estadísticos complejos para los que no hay ninguna solución analítica exacta, y para incluso las técnicas de aproximación normales que tienen dificultades. Un método Cadena de Markov Monte Carlo (MCMC) del acercamiento a la integración numérica es usado: Los BUGS Versión 0.6 generalmente usan el univariado muestreador de Gibbs, aunque puede usar un Metrópoli-Hastings dentro de la rutina de Gibbs cuando es necesario: WinBUGS tiene un catador de Metrópoli univariado más sofisticado.

El BUGS mantiene un idioma declaratorio para la especificación real de modelos estadísticos basada en la estructura gráfica supuesta, aunque hay algunas restricciones en las clases de modelos que han sido analizados actualmente. Un compilador entonces procesa el modelo, los datos y las corridas de las distribuciones muestrales requeridas por el Muestreador de Gibbs. Finalmente, muestreando, los algoritmos apropiados se llevan a cabo para simular valores de las cantidades desconocidas en el modelo. El trabajo en los BUG empezó en 1989, y llevaron a hacer los BUGS para varias plataformas. El reciente desarrollo tiene el enfoque en el WinBUGS.

La versión más avanzada de BUGS es ahora WinBUGS que corre bajo Windows. Esto comprende un programa para la especificación gráfica de modelos llamado DoodleBUGS, y una versión educativa de WinBUGS. Esto permite el análisis de todos los ejemplos disponibles de los BUGS clásicos, y para los usuarios el análisis de pequeños ejemplos de su propia invención. Una versión sin restricción de WinBUGS 1.4 está actualmente libremente disponible en Internet para los usuarios.

## 2.2- Estudio del sistema WinBUGS 1.4 para el cálculo del riesgo relativo utilizando el enfoque bayesiano

El sistema WinBUGS es actualmente el programa más utilizado para análisis estadísticos complejos, el cual asume como uso familiar el lenguaje BUGS y los modelos MCMC (ver Manual WinBUGS 1.4).

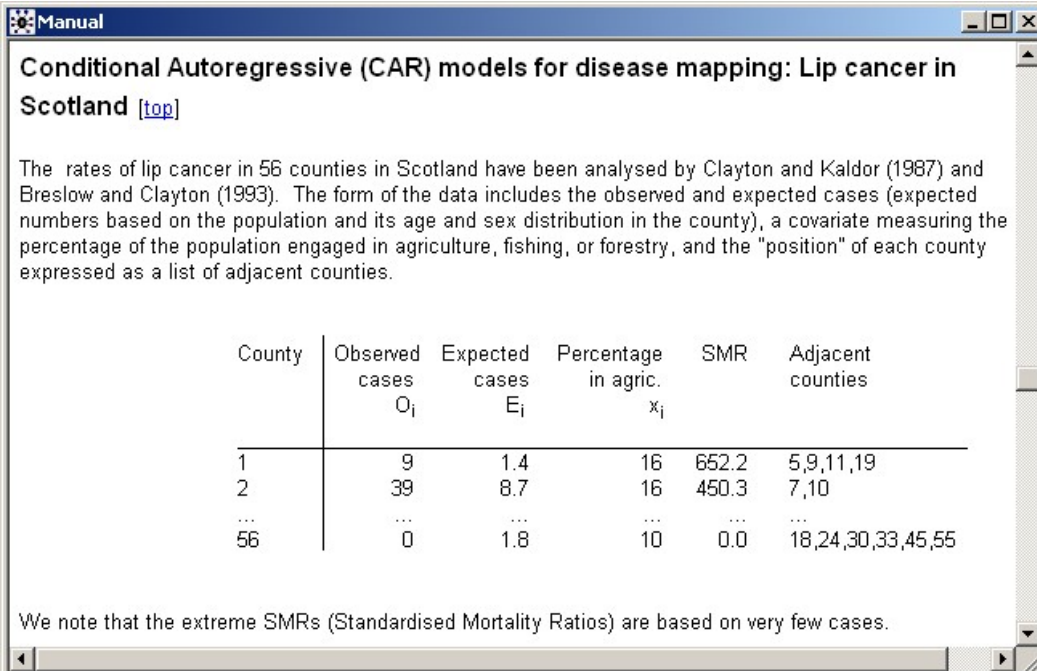
Para su uso, primeramente debemos usar un documento compuesto para la especificación del modelo, los datos, y los valores iniciales.

- Ahora explicamos que es un **Documento Compuesto**.

Es un documento que contiene varios tipos de información (textos, tablas, fórmulas, gráficos, etc.), presentados en una única ventana y almacenados en un único fichero \*.odc. Las herramientas necesarias para crear y manipular estos tipos de información variada están siempre disponibles, además no hay necesidad de estar continuamente moviéndose entre programas diferentes. El software ha sido diseñado para que produzca los resultados directamente a un documento compuesto.

Los elementos de un documento compuesto pueden ser enfocados, seleccionados, movidos, copiados, borrados, etc.

Un ejemplo de un documento compuesto es:



**Manual**

### Conditional Autoregressive (CAR) models for disease mapping: Lip cancer in Scotland [\[top\]](#)

The rates of lip cancer in 56 counties in Scotland have been analysed by Clayton and Kaldor (1987) and Breslow and Clayton (1993). The form of the data includes the observed and expected cases (expected numbers based on the population and its age and sex distribution in the county), a covariate measuring the percentage of the population engaged in agriculture, fishing, or forestry, and the "position" of each county expressed as a list of adjacent counties.

County	Observed cases $O_i$	Expected cases $E_i$	Percentage in agric. $x_i$	SMR	Adjacent counties
1	9	1.4	16	652.2	5,9,11,19
2	39	8.7	16	450.3	7,10
...	...	...	...	...	...
56	0	1.8	10	0.0	18,24,30,33,45,55

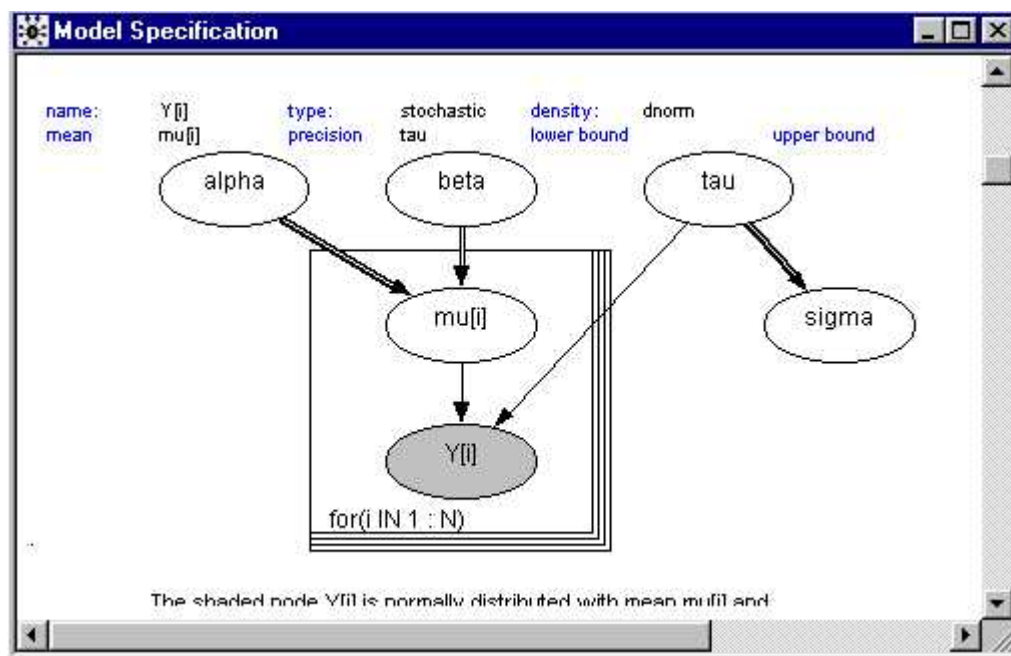
We note that the extreme SMRs (Standardised Mortality Ratios) are based on very few cases.

El software trabaja con muchos tipos diferentes de elementos de los cuales los más interesantes son los Doodles, los cuales permiten modelos estadísticos para ser descritos en términos gráficos.

- Ahora explicamos de forma general **como se trabaja** en este nuevo sistema, los detalles particulares para su uso y manejo lo encontrarán en el menú HELP que da acceso a la documentación del programa con una serie de ejemplos. Este programa se encuentra disponible en Internet.

- Para **especificar los modelos estadísticos** con los cuales vamos a trabajar veremos que esto puede ser hecho de dos formas:

1) Como modelos gráficos cuya construcción se realiza abriendo una nueva ventana, seleccionando la opción NEW en el menú Doodle. Ejemplo:



2) Usando el texto basado en el lenguaje BUGS; los cuales pueden ser construidos como parte de un documento en formato rectangular (hoja de trabajo del WingBUBS), encabezado por la palabra 'model'.

```
model {  
    texto basado de la descripción del gráfico en lenguaje BUGS  
}
```

Ejemplo:

```
model
{
  for (i in 1 : N) {
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha0 + alpha1 * X[i]/10 + b[i]
    RR[i] <- exp(alpha0 + alpha1 * X[i]/10 + b[i]) # Area-specific relative risk (for maps)
  }

  # CAR prior distribution for random effects:
  b[1:N] ~ car.normal(adj[], weights[], num[], tau)
  for(k in 1:sumNumNeigh) {
    weights[k] <- 1
  }

  # Other priors:
  alpha0 ~ dflat()
  alpha1 ~ dnorm(0.0, 1.0E-5)
  tau ~ dgamma(0.5, 0.0005) # prior on precision
  sigma <- sqrt(1 / tau) # standard deviation
}
```

- En la **Especificación de los Datos y de los Valores Iniciales** debemos tener en cuenta cualquiera de estas dos maneras:

- 1) En parte de un documento como datos en un arreglo en formato rectangular.
- 2) Como parte de un documento en formato S\_PLUS comenzando siempre con la palabra clave **List**.

Ejemplo:

```
list(N = 44,
O = c(4, 8, 3, 6, 2, 7, 7, 2, 10, 13, 5, 7, 9, 3, 3, 1, 2, 3, 19, 7, 5, 4,
      9, 2, 4, 6, 14, 13, 6, 7, 9, 3, 4, 10, 10, 8, 23, 24, 18, 10, 6, 4, 17, 7),
E = c(7.20902956, 7.81436792, 3.46693788, 7.04393728, 7.37412184,
      4.45749156, 6.3285374, 4.45749156, 6.3285374, 6.76878348, 9.30019844,
      .....
      .....
      15.7486080432, 17.6933951016, 8.642580858, 4.969277628, 6.3736626232,
      11.2890101064, 6.3736626232),
depriv = c(1.233, 8.162, 0.919, -0.78, -1.182, 3.647, 6.47, 0.948, 4.479,
           .....
           .....
           2.754, 7.55, 3.961),
```

```

num = c(5, 4, 6, 7, 4, 5, 6, 6, 3, 6,
        .....
        5, 3, 8, 6
        ),
adj = c(
17, 12, 9, 8, 4,
14, 13, 10, 7,
.....
.....
38, 34, 31, 30, 23, 16, 10, 7,
41, 32, 31, 26, 25, 22
),
sumNumNeigh = 212
)

```

- En el **menú Model** la aplicación de cualquier comando es para todo el modelo estadístico, abriendo cuadros de diálogos. Este menú es solamente presentado si se enfoca en una ventana un texto o un Doodle.

- En la **opción Specification** del menú Model

1) En la Verificación del modelo:

a) Si la vista enfocada contiene un Doodle, WinBUGS asume que el modelo ha sido especificado gráficamente. El Doodle es analizado. Para todos los errores de sintaxis que sean detectados, una descripción del error es dada en la caja de diálogo apropiada.

b) Si la vista enfocada contiene un texto, WinBUGS asume que el modelo está especificado en lenguaje BUGS y es analizado.

2) En la Verificación de los Datos (**Data**):

a) Si los datos están especificados como parte de un documento en formato rectangular la ventana que contiene el documento necesita ser enfocada cuando el comando 'load data' es usado.

b) Si los datos están especificados en un documento en formato S\_PLUS, la palabra list es enfocada cuando el comando 'load data' es usado.

3) Para la opción **Compile**:

Este comando construye la estructura de los datos necesarios para llevar a cabo el muestreo Gibbs (GS). En el modelo es chequeada su consistencia con los datos. Este comando vuelve a activarse enseguida que el modelo ha sido sucesivamente cambiado.

- 4) En la Verificación de los Valores Iniciales ([Inits](#)):
- Si los datos están especificados como parte de un documento en formato rectangular la ventana que contiene el documento necesita ser enfocada cuando el comando 'load data' es usado.
  - Si los datos están especificados en un documento en formato S\_PLUS, la palabra list es enfocada cuando el comando 'load data' es usado.

- Línea de trabajo hasta el momento

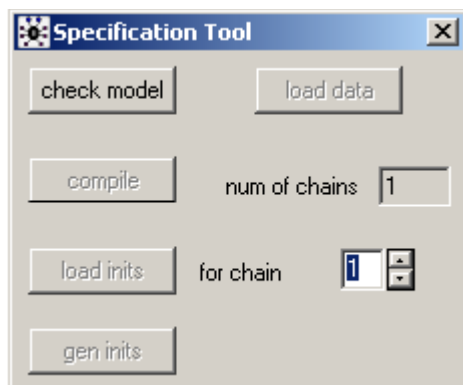
Para chequear que el modelo está correcto destacamos la palabra **Model** y entonces ejecutemos el comando Check Model de la opción Specification del menú Model, deberá aparecer en la barra de estado "Model is syntactically correct".

Para chequear y cargar un conjunto de datos, destaquemos la palabra clave **List** y entonces ejecutamos el comando Load Data de la opción Specification del menú Model, aparecerá en la barra de estado "data loaded".

Seguidamente ejecutamos el comando Compilar de la opción Specification del menú Model para establecer la estructura de los datos usados por el muestreador de Gibbs, aparecerá en la barra de estado "Model Compiled".

Para cargar y chequear los Valores Iniciales destacamos la palabra clave **List** y entonces ejecutamos el comando Load Inits de la opción Specification del menú Model, deberá aparecer en la barra de estado "initial values loaded: model initialized".

Ejemplo de la opción Specification del menú Model:



WinBUGS está ahora disponible para generar muestras.

Note que primero necesitamos chequear el modelo y luego cargar los datos, compilar el modelo y cargar los valores iniciales.

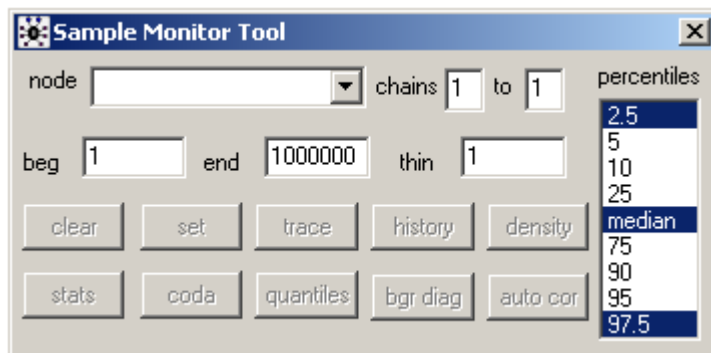
- La **opción Update** del menú Model

Este comando se activa una vez que el modelo ha sido compilado e inicializado. Este comando abre una caja de diálogo (se puede tener abierta, no hace falta ejecutarla al momento) con campos de entrada numérica. Updates para el número de iteraciones MCMC que será llevado a cabo. Estos generan muestras múltiples en cada iteración.



- En el **menú Inference** se abre una caja de diálogo para hacer inferencias sobre los parámetros del modelo, los comandos están divididos en tres secciones: la primera sección con tres comandos se refiere a un monitoreo entero del conjunto de valores para una variable, en la próxima sección sus dos comandos son space-saving short-cuts que controlan la corrida estadística y la última sección de un solo comando se refiere a la evaluación del Criterio de la información Deviance. Es importante asegurarse de que la simulación ha convergido antes de usar Summary, Rank o DIC; notemos que si la simulación MCMC tiene una fase adaptativa, no será posible usar para hacer inferencias valores muestreados antes de la terminación de esta fase.

- Un ejemplo del menú Inerence es la **opción Samples** donde se abre un diálogo para analizar las muestras almacenadas de las variables del modelo producidas por la simulación MCMC. La variable de interés debe ser escrita en el campo de texto Node y luego se debe usar el botón Set.



1. El campo numérico Node es utilizado para escribir la variable de interés.
2. El campo Chains... to puede ser usado para seleccionar las cadenas que contribuirán al cálculo del proceso estocástico.
3. Los campos numéricos Beg y el End son usados para seleccionar un subconjunto de las muestras almacenadas para el análisis.

4. El campo numérico Thin es usado para seleccionar todas las k-ésimas iteraciones de cada cadena que contribuirán a que el proceso estocástico sea calculado, donde k es el valor del campo.
5. El botón Clear quita los valores almacenados de la variable en memoria de la computadora.
6. El botón Set debe ser usado para comenzar registrando una cadena de valores de la variable.
7. El botón Trace plotea el valor de la variable contra el número de iteraciones, este trazado es dinámico, en cada tiempo la pantalla es redibujada. El botón Trace además comienza registrando una cadena de valores de la variable.
8. El botón History plotea un trace completo de la variable.
9. El botón Density plotea una estimación del núcleo de la densidad suavizada para la variable si esta es continua y un histograma si es discreta.
10. El botón Stats produce resúmenes estadísticos para las variables, aunados sobre las cadenas seleccionadas. Los percentiles requeridos pueden ser seleccionados usando la caja de selección de percentiles. La cantidad reportada en las columnas del error MC da una estimación de  $\sigma/N^{1/2}$ , el error estándar Monte Carlo de la media.
11. El botón Coda saca una representación ascii de los valores monitoreados apropiados para usar en el paquete diagnóstico CODA.
12. El botón Quantile plotea la media corriente con un intervalo de confianza corriente contra el número de iteraciones.
13. El botón Bgr diag calcula la convergencia estadística de Gelman-Rubin, como fue modificada por Brooks y Gelman (1998). La amplitud del intervalo central al 80% de las corridas juntas es verde, la amplitud promedio de los intervalos al 80% dentro de las corridas individuales es azul, y su razón R (juntas/dentro) es roja –por el propósito del ploteo la amplitud de los intervalos de las juntas y las de dentro fueron normalizadas para hacer un máximo total de una. Las estadísticas fueron calculadas en papeleras de longitud de 50: generalmente podríamos esperar que R sea mayor que 1 si los valores iniciales están debidamente sobredispersados. Brooks y Gelman enfatizaron que uno debería preocuparse a la vez de la convergencia de R hacia 1 y de la convergencia de ambas la amplitud de los intervalos de las juntas y de las de dentro para la estabilidad.
14. El botón Auto Cor plotea la función de autocorrelación de la variable de salida para un retraso de 50.

Para analizar la convergencia de los resultados muestrales se utiliza el programa BOA ya que es más eficiente para esto. El Programa para el Análisis de los Resultados Bayesianos (BOA) es un programa de S-PLUS o de R, el cual nos da los diagnósticos de convergencia, resúmenes estadísticos y análisis de gráficos de las muestras producidas por el método Monte Carlo. Este puede ser usado como un procesador de los resultados para el software BUGS o para cualquier otro programa que produzca resultados muestrales.

Dicho programa esta basado en el software de los Diagnósticos de Convergencia y Análisis de Resultados para las salidas del muestreador de Gibbs (CODA). Este incluye todas las opciones de análisis encontradas en el CODA y otras más. Está diseñado para que sea rápido, más eficiente, y ofrece más flexibilidad en el manejo de los datos que el CODA

El BOA puede ser usado por líneas de comando. De cualquier manera, el programa permite al usuario enfocarse en el análisis que tiene a mano más que en la manipulación de los datos.

Podemos obtenerlo en internet: <http://www.public-health.uiowa.edu/BOA>

Para su uso utilizando el interfaz escribimos : `> boa.menu()`

Se comienza importando los resultados archivados del BUGS, o S-PLUS matrices.

Primero debemos entrar el Path que es la dirección de la memoria de la máquina en donde vamos a trabajar, por ejemplo:

```
> boa.par(path="D:\\BUGS\\voll\\line\\", title="Line problem from the BUGS manual").
```

Seguidamente para importar los resultados usamos:

```
>boa.chain.import("_____"). (Dentro de las comillas se escribe el nombre de la cadena que tiene los resultados muestrales del BUGS)
```

Para ver la información que acabamos de importar escribimos:

```
>boa.print.info().
```

Luego para ver los resultados del diagnóstico de convergencia de Gelman y Rubin que es lo que nos interesa, se escribe:

```
>boa.print.gandr().
```

Si lo deseamos podemos graficar este diagnóstico de convergencia con el comando:

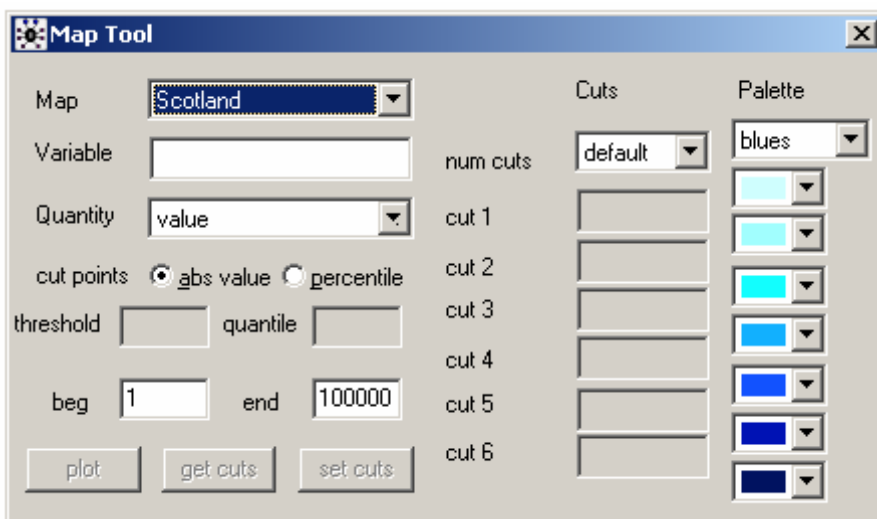
```
>boa.plot("gandr")
```

Y con esto ya tenemos el análisis de la convergencia de los resultados obtenidos en el WinBUGS.

### 2.3- Estudio del sistema GeoBUGS para la representación en mapas de la distribución del riesgo relativo

El sistema computacional GeoBUGS aparece como el **menú Map** dentro del WinBUGS, es el que se utiliza para el mapeo geográfico de la morbi-mortalidad de enfermedades, o mejor dicho para graficar en un mapa dado un parámetro de interés.

- La **opción Mapping Tool** del menú Map es seleccionada para producir un mapa del parámetro estadístico de interés, veamos como se utiliza:



- 1) Se selecciona el mapa con que se va a trabajar en la caja de texto Map (el mapa con que se trabajará debió haber sido importado con anterioridad).
- 2) Se escribe la variable con la cual se va a realizar el mapeo en la caja de texto Variable, dicha variable debe tener ya puesta una muestra o resumen para este parámetro y debe de tenerse mucho cuidado con algunos update.
- 3) Si la variable son datos (por ejemplo el SMR crudo, los E esperados, o una covariada) escoja la opción value en el menú Quantity y entonces pulse el botón Plot: un mapa obscurecido según los valores de la variable aparecerá.
- 4) Si la variable es una cantidad estocástica (por ejemplo los riesgos relativos) hay varias opciones que puede seleccionar del menú Quantity:
  - Si se ha monitoreado la variable poniendo un resumen, entonces se debe seleccionar la opción mean(summary) de este menú, como solo las medias posteriores son almacenadas por el resumen.

- Si se ha monitoreado la variable poniendo una muestra (que almacena la muestra posterior total), se puede seleccionar cualquiera de las opciones restantes del menú Quantity:

El Mean(sample) mapeará las medias posteriores de la variable.

El Percentil trazará los cuantiles posteriores de la variable - si se selecciona esta opción, se debe teclear el percentil requerido.

El Prob Greater trazará la probabilidad posterior de que el valor de la variable sea mayor o igual al umbral especificado que debemos teclear en la caja de texto threshold.

El Prob less trazará la probabilidad posterior de que el valor de la variable sea menor o igual al umbral especificado que debemos teclear en la caja de texto threshold.

En Cuts se delimitan los intervalos en que deseamos dividir los riesgos.

En Palette se seleccionan las diferentes tonalidades de colores en que serán dibujadas cada una de las regiones del mapa de acuerdo a los intervalos indicados.

Cuando se ha seleccionado la cantidad que se quiere trazar, se pulsa el botón Plot para desplegar el mapa.

- Los números en los anaqueles mostrados en la leyenda del mapa dan el número de áreas clasificado en cada categoría en el mapa.

Este sistema al ser muy interactivo con el usuario permite obtener una serie de resultados complementarios, por los cuales es necesario tener un conocimiento del modelo con que se trabaja, para al final obtener un modelo ajustado de los datos.

## CAPÍTULO 3

### **Análisis de los resultados obtenidos al aplicar la metodología clásica y la bayesiana para el cálculo del riesgo relativo para la mortalidad por enfermedades cerebrovasculares en los municipios del país**

#### **3.1- Introducción**

La importancia de analizar la variación geográfica del riesgo relativo de las defunciones por enfermedades cerebrovasculares es, que nos posibilita identificar aquellas regiones del país de mayor riesgo, pudiendo actuar en la planificación y distribución de los recursos destinados a mejorar la calidad de la atención médica disminuyendo la mortalidad por esta causa en los lugares de mayor riesgo, aplicar un tratamiento eficaz para lo cual es necesario la atención especializada por equipos multidisciplinarios, impartir una adecuada educación sanitaria al paciente, la cual garantice su posterior rehabilitación.

Planteamiento del problema:

Las enfermedades cerebrovasculares (ECV), constituyen uno de los problemas de salud más importantes en todos los países desarrollados y en nuestro país. Múltiples recursos se emplean con el fin de atenuar las pérdidas de vidas humanas que alcanzan niveles considerables cada año, y el grado de invalidez que presentan estos pacientes (Buergo, Dr. M. A. 2000).

En Cuba constituyen la tercera causa de muerte, observándose en las últimas tres décadas un ascenso de las mismas en tasas brutas, afecta fundamentalmente, a los mayores de 65 años detectándose una tendencia a aumentar en los grupos de edades más jóvenes y a presentar una sobremortalidad femenina, en diferentes provincias, aunque globalmente se observa una tendencia a disminuir.

Las enfermedades cerebrovasculares, constituyen un grupo de enfermedades, las cuales involucran invariablemente a los vasos sanguíneos encargados de la irrigación del sistema nervioso (SN). Sus consecuencias son la isquemia y el infarto del 85 al 90% del total de los casos, las hemorragias intracraneales del 10 al 15 %. Dentro de las patologías consideradas se encuentran las congénitas y las adquiridas, formando así un amplio espectro de posibilidades en ambos grupos. El accidente cerebrovascular (ACV) o ictus, que incluye al infarto cerebral, la

hemorragia intraparenquimatosa y la hemorragia subaracnoidea, es una emergencia médica, ya que sólo el tratamiento médico-quirúrgico agresivo puede minimizar las consecuencias, a veces catastróficas, de la enfermedad.

La incidencia del accidente cerebrovascular (ACV) oscila alrededor de 200 casos entre 100000 habitantes por años y la frecuencia relativa de cada tipo es, suprimiendo valores medios, infartos cerebrales el 80% del total (20 % de origen embólico y 69 % aterotrombótico), hemorragias el 20% restante (aproximadamente del 12 al 15 % son hemorragias intraparenquimatosa y de un 5 a un 8% subaracnoidea).

Para ilustrar los métodos detallados en el Capítulo I se utilizarán los datos de mortalidad por enfermedades cerebrovasculares para cada municipio del país en el año 2002.

En calidad de áreas geográficas se considerarán los 169 municipios del país determinados por la división político-administrativa vigente desde 1975.

La información relacionada con las defunciones se obtuvo de la base de datos de la Dirección Nacional de Estadística del Ministerio de Salud Pública y los datos relacionados con la distribución de la población por edad del año de estudio se obtuvieron de las bases de datos de la Oficina Nacional de Estadística.

Los casos observados de defunciones son denotados por ( $Y_i$ ) y los casos esperados de defunciones para cada área son denotados por ( $E_i$ ) y para su cálculo las edades fueron divididas en 7 grupos:

0	a	14
15	a	24
25	a	34
35	a	44
45	a	54
55	a	64
65	años	y más

### 3.2- Resultados obtenidos de aplicar la metodología clásica al cálculo y mapeo del riesgo relativo

Como ya habíamos dicho anteriormente, el análisis de los datos a través de los métodos tradicionales exige el cálculo de las SMRs, usando la expresión (1.3.2) del Capítulo I para cada municipio.

Los datos para cada uno de los municipios se muestran en el Anexo 0.

Para mostrar dichos cálculos, consideremos el municipio de Maisí de la provincia de Guantánamo. Toda la información que necesitamos para este cálculo se resume en la siguiente tabla:

Grupo de edades	Defunciones en Maisí	Población de Maisí	Tasa de mortalidad por edades
0 - 14	0	8 025	0.0000037
15 - 24	0	4 541	0.0000059
25 - 34	0	5 828	0.0000201
35 - 44	0	4 021	0.000109
45 - 54	0	2 611	0.000347
55 - 64	1	2 101	0.0009951
65 - más	20	2 216	0.0053489

Las defunciones observadas para el municipio de Maisí son:

$$\sum_{j=1}^7 d_j = 0 + 0 + 0 + 0 + 0 + 1 + 20 = 21$$

donde  $d_j$  es la defunción observada en el grupo de edad  $j$  (dichos datos se encuentran en la segunda columna de la tabla) y esta suma es el termino del numerador de (1.3.2).

Los casos esperados se calculan dividiendo el total de defunciones de cada grupo de edad en Cuba por la población total de ese grupo de edad en Cuba, esta división es la tasa de mortalidad por edad (columna 4) (ver Anexo 1), esta se multiplica por la población total de Maisí del grupo de edad correspondiente.

Hecho el cálculo para cada grupo de edad se suman para obtener el valor esperado:

$$\begin{aligned} \sum_{j=1}^7 n_j t_j^* &= 0.0000037 * 8025 + 0.0000059 * 4541 + 0.0000201 * 5828 + 0.000109 * 4021 \\ &+ 0.000347 * 2611 + 0.0009951 * 2101 + 0.0053489 * 2216 \end{aligned}$$

Finalmente el SMR es igual a 15,4618.

La Figura 1 muestra las SMRs de las enfermedades cerebrovasculares del 2002 calculadas como explicamos anteriormente para los 169 municipios del país (ver Anexo 3). Las SMRs son valores que se ubican de manera natural alrededor del 1; para que sus valores resulten más fácilmente visualizables en este ejemplo, se representan categorizadas de acuerdo a su distribución en quintiles. Dichos valores se clasifican según los siguientes intervalos:

0	a	0.7368
0.7369	a	0.8588
0.8589	a	0.9977
0.9978	a	1.1928
1.1929	a	1.0818

Tal agrupación conforma grupos con tamaños bastantes homogéneos (4 con 34 municipios y 1 con 33). Gráficamente son representados en 5 tonalidades de azul.

Como se puede observar en este mapa no es posible discernir patrón alguno de mortalidad, puesto que existe mucha dispersión por todo el mapa de un mismo grupo de riesgo, regiones con muy pocos casos de la enfermedad y poblaciones pequeñas aparecen representadas como de alto riesgo, hecho que se debe, como era de esperar, al predominio de los valores extremos. Obsérvese además la dependencia de los tamaños poblacionales, con el consecuente impacto sobre la variabilidad de las estimaciones, fenómeno que ocurre con más frecuencia, como ya hemos visto, con enfermedades raras y poblaciones pequeñas y debido a esta variabilidad de las estimaciones el mapa de los SMRs está altamente afectado por el “ruido”, por lo que no se daría una interpretación epidemiológica útil, en este caso la variabilidad del riesgo relativo dentro de cada área excede a la que se podía esperar mediante una distribución Poisson; un ejemplo claro en el mapa donde se observan estas deficiencias del método de Poisson es el municipio utilizado anteriormente para ilustrar los cálculos de los SMRs, Maisí, el cual tiene una población pequeña de tan solo 29 343 habitantes, este municipio exhibe un tono azul oscuro, podríamos pensar que es una zona de elevado riesgo, muy diferente al de sus municipios vecinos. Sin embargo esta cifra de alto riesgo alarmante se debe a sólo 21 defunciones.

Figura 1

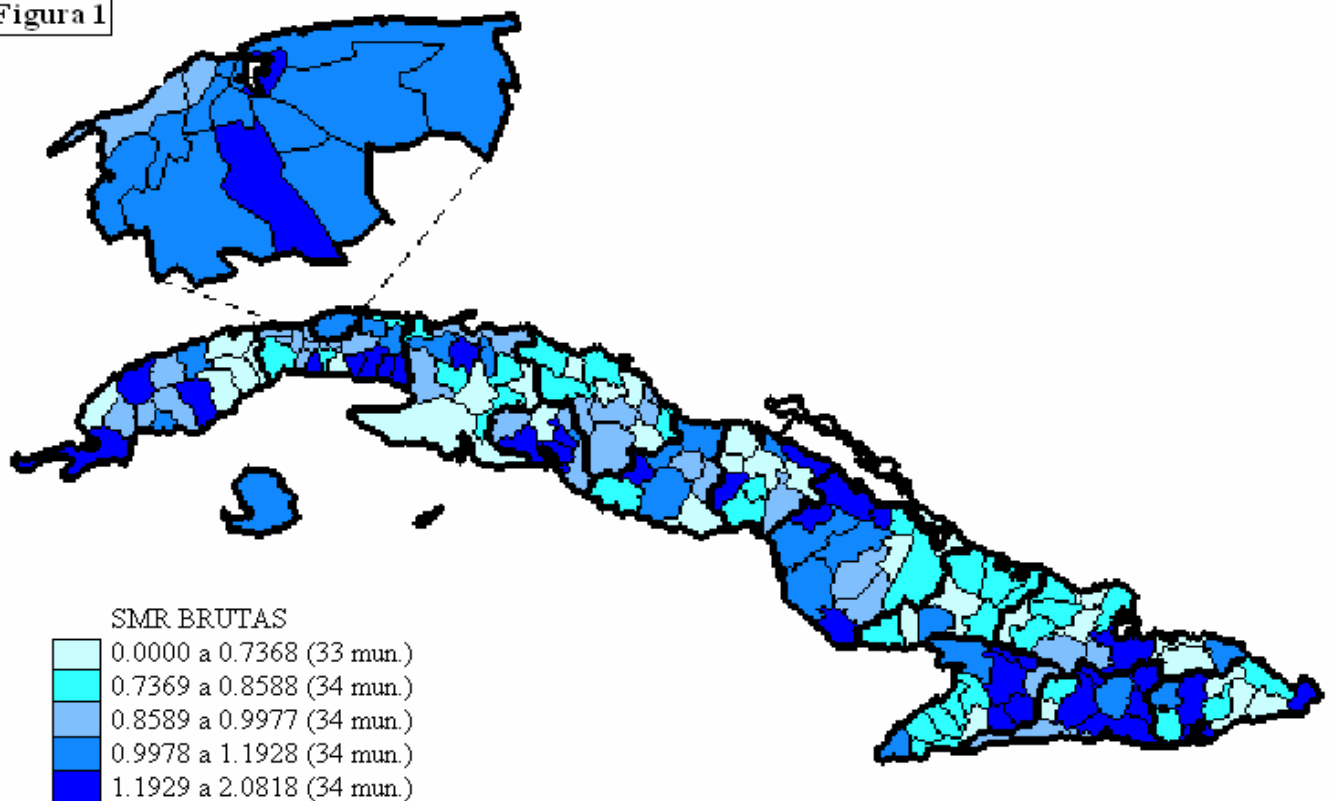
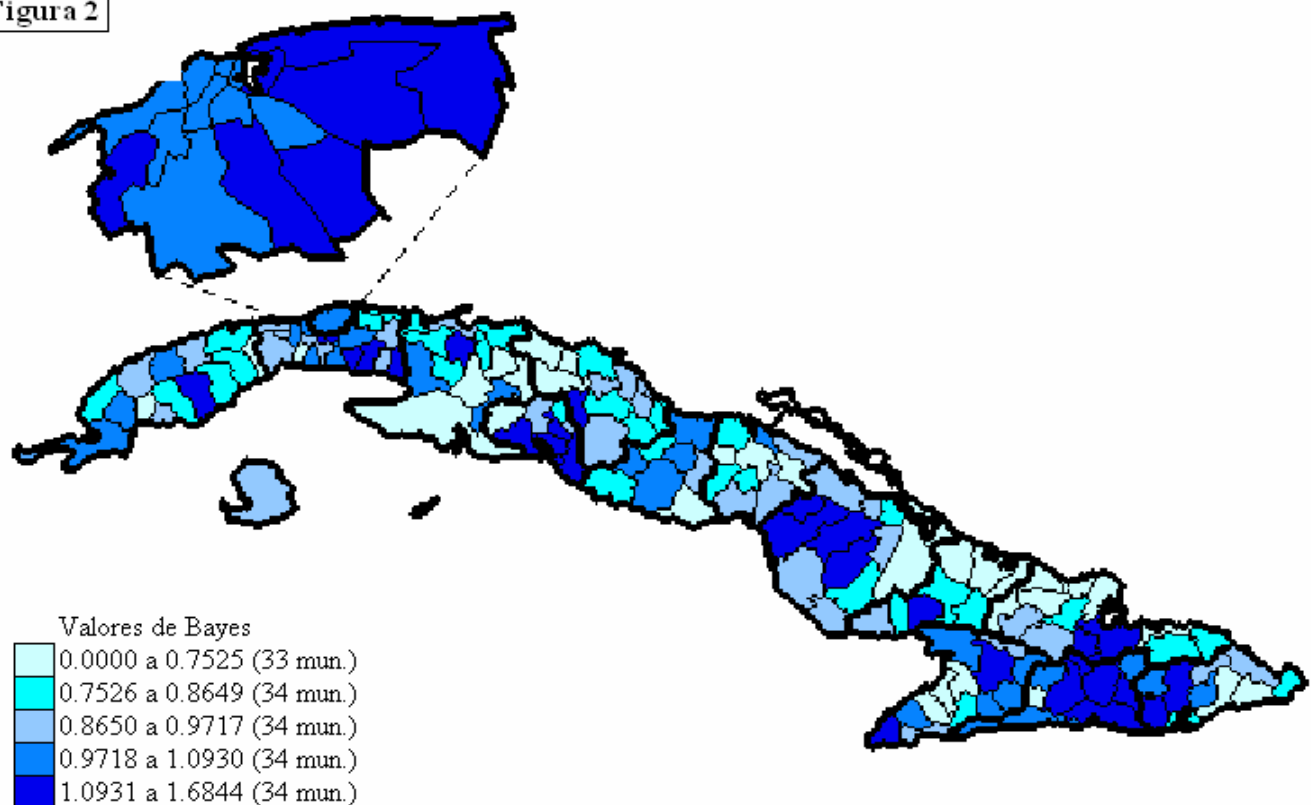


Figura 2



### **3.3- Resultados obtenidos de aplicar la metodología bayesiana al cálculo y mapeo del riesgo relativo**

Para obtener las estimaciones del riesgo relativo, se utiliza el programa WinBUGS 1.4 que simula valores de parámetros e hiperparámetros utilizando el método de Monte Carlo cadena de Markov y el muestreador de Gibbs, de modo que se obtiene así la distribución a posteriori de interés. Con este modelo jerárquico se hicieron 11 000 iteraciones, fueron despreciadas las primeras 1 000 iteraciones para eliminar el efecto de los valores iniciales y la inferencia a posteriori se basó en las cadenas de las otras 10 000 muestras.

El modelo jerarquico utilizado en el análisis para este estudio y los datos se pueden encontrar en el Anexo 2.

La Figura 2 muestra la distribución geográfica de los riesgo relativo calculados según el método bayesiano explicado (ver Anexo 4). Como en el mapa anterior, cada una de las categorías representa los quintiles de la distribución de las estimaciones de los riesgo relativo en cinco tonos de azul.

En este mapa se observa claramente el efecto de la suavización, en contraposición con el mapa anterior confeccionado con las técnicas tradicionales, y revela con mayor claridad como se distribuyen los patrones de mortalidad por enfermedades cerebrovasculares en los diferentes municipios del país. Dicho mapa elude la inestabilidad del mapeo de los SMRs brutos y la no informatividad por lo que el mapa se suaviza consiguiendo así una mejor interpretación epidemiológica. Como se puede observar aquí, la estimación del riesgo relativo en un área está fuertemente influenciada por la estimación de áreas adyacentes geográficamente y sólo influenciada indirectamente por la estimación de las otras áreas del mapa, hecho que hace que las estimaciones individuales estén mas desplazadas hacia un valor medio local que hacia un valor medio global. El acercamiento bayesiano corrige la sobredisperción del método de Poisson; para cada área este acercamiento combina la información expresada en los datos de las áreas con la información del resto del mapa, expresada en el modelo priori. Este mecanismo tiende a suavizar las partes del mapa donde las informaciones son poco fiables.

Si observamos el mapa de la Figura 2 podemos apreciar que la mayor concentración de municipios de alto riesgo para la mortalidad por ECV se encuentran en las provincias de: Ciudad de la Habana, Cienfuegos, Camagüey y Santiago de Cuba, observándose algunos municipios

aislados de alto riesgo en las provincias de: Pinar del Río, La Habana, Matanzas, Villa Clara, Las Tunas, Holguín, Granma y Guantánamo.

A modo de ilustración final veamos que ocurre en el municipio de Maisí (Figura 2). Como se observa aparece con un tono más claro, de marcada homogeneidad con los municipios que lo rodean. A diferencia de la Figura 1, Maisí se identifica como un área de bajo riesgo.

### **3.4- Comparación de los resultados obtenidos por ambas metodologías**

- El enfoque bayesiano permite incorporar evidencias de experiencias (información a priori) y los resultados de experimentos previos (función de verosimilitud) en el proceso de análisis formal de los datos y el enfoque clásico no.
- El enfoque clásico estima los parámetros del modelo por máxima verosimilitud y el enfoque bayesiano lo hace a través de la simulación MCMC
- Cuando se calculan los riesgos relativos a través de métodos tradicionales, no se puede apreciar con claridad en el mapa los patrones de mortalidad, sobre todo cuando se trata de una enfermedad rara o de áreas pequeñas, debido al predominio de valores extremos. Con los métodos bayesianos, sin embargo el análisis de los patrones de mortalidad tiene innegables ventajas, dadas por la estabilidad de los estimadores que se utilizan, además, el hecho de que con este procedimiento se puedan tener en cuenta en cada área los datos de las áreas vecinas, permite la obtención de unos resultados más sensibles y respetuosos de los verdaderos patrones espaciales.

## CONCLUSIONES

- Se realizó una revisión bibliográfica extensa sobre las dos metodologías empleadas en la determinación del riesgo relativo para la morbi-mortalidad de enfermedades, la metodología clásica y la bayesiana.
- En la metodología clásica debemos determinar los casos observados (defunciones) y los casos esperados para cada una de las regiones en el estudio y luego calcular los SMRs, para cada una de las regiones.
- En la metodología bayesiana debemos determinar los casos observados y los casos esperados para cada una de las regiones a analizar y luego se recomienda asignar distribuciones a priori que recojan información sobre la interrelación que exista entre las regiones en dependencia de que sean vecinas o no, posteriormente se deben estimar los parámetros de modelo, valorar la convergencia del método y representar la distribución de los riesgos en el mapa para poder determinar patrones de riesgo para las diferentes regiones.
- Por todo lo analizado en los capítulos anteriores, hemos llegado a la conclusión de que: mediante la metodología clásica los SMRs altos ocurren en algunas ocasiones en áreas con la población pequeña y poca ocurrencia de casos, produciendo un mapa dominado por estos ruidos en el que es imposible determinar patrones de riesgo para la morbi-mortalidad de enfermedades. Por otro lado la metodología Bayesiana corrige la sobredispersión para cada área, este acercamiento combina la información expresada en los datos de las áreas con la información del resto del mapa, expresada en el modelo priori, obteniéndose mapas más suavizados, donde si se pueden determinar claros patrones de riesgo para la mortalidad. Esta metodología brinda resultados con mucha más precisión en cuanto a que se acerca más a la realidad del fenómeno que la metodología clásica, a la vez que permite al especialista poder utilizar toda la información a priori que tiene acerca del estudio, lo cual hace más realista la solución del problema.
- En el caso del estudio investigativo sobre la determinación del riesgo relativo para la mortalidad por enfermedades cerebrovasculares y su posterior representación en el mapa para determinar los municipios del país mas afectados por esta causa de muerte, se llegó a la conclusión de que los patrones de alto riesgo se encuentran en municipios pertenecientes a las provincias de: Ciudad de la Habana, Cienfuegos, Camagüey y

Santiago de Cuba, observándose algunos municipios aislados de alto riesgo en las provincias de: Pinar del Río, La Habana, Matanzas, Villa Clara, Las Tunas, Holguín, Granma y Guantánamo.

## BIBLIOGRAFÍA

- Assuncao, R. M., Barreto, S. M., Guerra, H. L., and Sakurai, E., (1998), Mapas de taxas epidemiológicas: uma abordagem Bayesiana.  
Cad. Saúde Pública, vol.14 n.4 Río de Janeiro
- Benavides, A. C., and Silva, L. C., (2002), Insuficiencias del paradigma frecuentista y el enfoque Bayesiano como alternativa.  
Tesis doctoral, Escuela Nacional de Salud Pública de Cuba
- Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C. , Ghislandi, M., and Songini, M., (1995), Bayesian analysis of space-time variation in disease risk.  
Statistic in Medicine, vol 14, 2433-2443
- Bernardinelli, L., and Montomoli, C., (1992), Empirical Bayes versus Fully Bayesian analysis of geographical variation in disease risk.  
Statistic in Medicine, vol 11, 983-1007
- Bernardo, J. M., (1981), Bioestadística: una perspectiva Bayesiana.  
Editorial, Vicens-Vives
- Bernardo, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M., (1985), Generalized Linear Models. Scale Parameters, Outlier Accommodation and Prior Distributions.  
Elsevier Science Publisher B. V. (North-Holland)
- Breslow, N. E., and Clayton, D. G., (1993), Approximate Inference in Generalized Linear Mixed Models  
Royal Statistical Society
- Best, N. G., Spiegelhalter, D., Thomas, A., and Brayne, C. E. G., (1995), Bayesian Analysis of Realistically Complex Models.

Royal Statistical Society

- Buergo, Dr. M. A., (2000), Programa Nacional de Prevención y Control de la Enfermedades Cerebrovasculares (ECV)  
Ministerio de Salud Pública
- Dey, D. K., Gelfand, A. E. And Peng, F., (1997), Overdispersed Generalized Linear Models.  
Journal of Statistical
- Gelman, A., Carlin, J. B., Starn, H. S., and Rubin, D. B., (1995), Bayesian Data Analysis.  
Chapman & Hall: 331-332
- Lawson, A. B., (1995), Markov Chain Monte Carlo Methods for putative pollution source problems in environmental epidemiology.  
Statistic in Medicine, vol 14, 2473-2486
- Lawson, A. B., Bigger, A., Bohning, D., Lesaffre, E., Viel, J. F., and Bertollini, R., (1997), Disease Mapping and Risk Assessment for Public Health.  
John Wiley & sons, LTD
- Lawson, A. B., and Clark, A., (2002), Spatial mixture relative risk models applied to disease mapping.  
Statistic in Medicine, vol 21, 359-370
- Lawson, B. A., and Cressie, N., (2000), Spatial Statistical Methods for Environmental Epidemiology.  
P. K. Sen and C. R. Rao, eds., Handbook of Statistics, vol 18, Elsevier Science B. V.
- Lawson, B. A., and Waller, L., (1996), A review of point pattern methods for spatial modelling and events around sources of pollution.  
Environmetrics 7, 471-488

- Lee, P. M., (1997), Bayesian Statistics an Introduction.  
Second Edition. Arnold, a member of the Hodder. HeadLine Group, 338 Euston Road,  
London NW1 3BH
- Lindley, D. V., and Smith, A. F., (1971), Bayes Estimates for the Linear Model  
Royal Statistical Society
- Natarajaran, R., and Kass, R. E., (2000), Reference Bayesian Methods for Generalized Linear  
Mixed Models.  
American Statistical Association. Journal of the American Statistical Association, vol 95, No.  
449, Theory and Methods
- Nieto, L. E., (2001), Estadística Bayesiana, Módulo 3, Diplomado de Estadística Aplicada,  
PDF, Internet.
- Silva, L. C., (2000), Que es la inferencia Bayesiana?  
JANO, Medicina y Humanidades 58 (1338): 65-66
- Spiegelhalter, D., (2001), Bayesian Methods in Health Research.  
III Jornada Nacional de Bioestadística, I Encuentro de Estadística Médica. Varadero, Cuba
- Spiegelhalter, D., Thomas, A., and Best, N., (2002), WinBUGS 1.4 Software. Gratis en  
internet; <http://www.mrc-bsu.com.ac.uk/bugs>

## ANEXO 0

### Tabla de datos de Excel

#	Municipios	Valor Observado	Valor Esperado
1	sandino	30	22,3
2	mantua	8	15,2
3	m. matahambre	26	21,3
4	viñales	16	16,8
5	la palma	24	21,2
6	bahia honda	15	27,2
7	candelaria	3	11,7
8	san cristobal	26	38,6
9	los palacios	17	25,3
10	c. del sur	76	59
11	pinar del rio	110	124,4
12	san luis	25	24,2
13	san. j martinez	30	34,1
14	guane	20	20,9
15	mariel	28	24,6
16	guanajay	19	22,1
17	caimito	26	23,6
18	bauta	28	31,7
19	san a. baños	28	31,6
20	bejucal	17	18,3
21	san j. de lajas	44	46,2
22	jaruco	25	23,8
23	santa c.del norte	19	23,3
24	madruga	27	23,2
25	nueva paz	26	20,3
26	san nicolas	21	16,3
27	guines	74	57,9
28	melena del sur	22	14,7
29	batabano	13	18
30	quivicán	18	21
31	guira de melena	34	27,8
32	alquizar	15	17,3
33	artemisa	45	52,4
34	playa	153	155,1
35	p. revolucion	172	172,4
36	centro habana	171	147,3
37	la habana vieja	87	79
38	regla	41	33,6
39	la habana del este	117	112,4
40	guanabacoa	89	80,6
41	san m. del padron	118	113,4
42	10 de octubre	237	229,2
43	cerro	137	123
44	marianao	126	111,2
45	la lisa	86	83,4
46	boyeros	126	124,3

#	Municipios	Valor Observado	Valor Esperado
47	arroyo naranjo	173	83,1
48	cotorro	59	49,7
49	matanzas	82	106,7
50	cardenas	66	68,3
51	varadero	13	13,5
52	marti	15	16,9
53	colon	38	49,5
54	perico	23	21,3
55	jovellanos	45	36,3
56	pedro betancourt	19	23,8
57	limonar	19	17,4
58	union de reyes	33	34,9
59	cienaga de zapata	2	4,2
60	jaguey grande	21	38
61	calimete	16	21,1
62	los arabos	10	18
63	corralillo	18	21
64	q. de guines	12	19,4
65	sagua la grande	37	50,2
66	encrucijada	21	30
67	camajuani	50	57,8
68	caibarien	30	31,6
69	remedios	33	44,2
70	placetas	49	67,2
71	santa clara	159	177,7
72	cifuentes	22	32
73	santo domingo	32	42,3
74	ranchuelo	51	54,8
75	manicaragua	49	52,2
76	aguada de p	21	21,8
77	rodas	17	25,7
78	palmira	40	25
79	lajas	18	18,1
80	cruces	32	27,4
81	cumanayagua	35	35,2
82	cienfuegos	127	105,8
83	abreus	28	19,2
84	yaguajay	55	50,4
85	jatibonico	27	30,6
86	taguasco	27	30,1
87	cabaiguan	70	60,7
88	fomento	38	29,6
89	trinidad	41	48,7
90	sancti spiritus	101	91,7
91	la sierpe	7	10,7
92	chambas	19	29

#	Municipios	Valor Observado	Valor Esperado
93	moron	45	43,2
94	bolivia	14	8,2
95	primero de enero	12	17,4
96	ciro redondo	12	17,8
97	florencia	8	18,2
98	majagua	23	19,1
99	ciego de avila	76	90
100	venezuela	14	17
101	baragua	17	19,2
102	c. m. de cespedes	20	14,7
103	esmeralda	27	20,4
104	sierra de cubitas	12	9
105	minas	20	24,9
106	nuevitas	16	25,6
107	guaimaro	32	39,9
108	sibanicu	14	19
109	camaguey	277	239,1
110	florida	61	52,9
111	vertientes	32	30,8
112	jimaguayu	9	9,3
113	najasa	10	10,1
114	santa cruz del sur	42	29,5
115	manati	14	18,1
116	puerto padre	48	59,8
117	jesus menendez	27	35,2
118	majibacoa	17	22,1
119	las tunas	73	103,1
120	jobabo	29	27,6
121	colombia	14	19,8
122	a. rodriguez	20	24,1
123	gibara	43	58,2
124	rafael freyre	18	35,6
125	banes	46	60
126	antilla	7	8
127	baguanos	17	40,3
128	holguin	145	195,9
129	calixto garcia	23	39,7
130	cacocum	19	25

#	Municipios	Valor Observado	Valor Esperado
131	urbano noris	24	26,2
132	cueto	20	22,4
133	mayari	79	63,1
134	frank pais	8	16,3
135	sagua de tanamo	24	33
136	moa	27	22,8
137	rio cauto	26	22,3
138	cauto cristo	16	9,7
139	jiguani	38	39
140	bayamo	151	121,7
141	yara	28	34,9
142	manzanillo	65	87,4
143	campechuela	23	27,6
144	media luna	17	20,1
145	niquero	27	22,9
146	pilon	11	14,9
147	bartolome maso	20	28,2
148	buey arriba	21	17,2
149	guisa	37	30,2
150	contramaestre	50	61,2
151	mella	22	16,6
152	san luis	60	50,3
153	segundo frente	32	19,7
154	songo la maya	67	53,9
155	s. de cuba	438	292,4
156	palma soriano	97	77,6
157	tercer frente	10	15,5
158	guama	17	18
159	el salvador	24	21,6
160	guantanamo	168	137,4
161	yateras	6	8,8
162	baracoa	38	47,5
163	maisi	21	15,5
164	imias	11	15,3
165	san a. del sur	8	13,2
166	manuel tames	6	7,3
167	caimanera	7	5,3
168	niceto perez	6	7,5
169	i de la juventud	38	33,6

## ANEXO 1

### Defunciones por grupos de edades para las Enfermedades Cerebrovasculares en Cuba, año 2002

<b>Grupo de Edad</b>	<b>Total de defunciones en Cuba</b>	<b>Población total de Cuba</b>	<b>Tasa de mortalidad por edades</b>
0-14	9	2 432 355	0.0000037
15-24	9	1 507 039	0.0000059
25-34	44	2 181 446	0.0000201
35-44	186	1 705 959	0.000109
45-54	459	1 322 686	0.000347
55-64	985	989 756	0.0009951
65-más	5 964	1 114 986	0.0053489

## ANEXO 2

### Hoja de datos para el cálculo utilizando el WinBUGS para el análisis bayesiano

```
model
{
  for (i in 1 : N) {
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha0 + b[i]
    RR[i] <- exp(alpha0 + b[i]) # Area-specific relative risk (for maps)
  }

  # CAR prior distribution for random effects:
  b[1:N] ~ car.normal(adj[], weights[], num[], tau)
  for(k in 1:sumNumNeigh) {
    weights[k] <- 1
  }

  # Other priors:
  alpha0 ~ dflat()
  tau ~ dgamma(0.5, 0.0005) # prior on precision
  sigma <- sqrt(1 / tau) # standard deviation
}
```

```
Data list(N = 169,
  O = c(30, 8, 26, 16, 24, 15, 3, 26, 17, 76, 110, 25, 30, 20, 28, 19, 26, 28, 28, 17,
44, 25, 19, 27, 26, 21, 74, 22, 13, 18, 34, 15, 45, 153, 172, 171, 87, 41, 117, 89,
118, 237, 137, 126, 86, 126, 173, 59, 82, 66, 13, 15, 38, 23, 45, 19, 19, 33, 2, 21,
16, 10, 18, 12, 37, 21, 50, 30, 33, 49, 159, 22, 32, 51, 49, 21, 17, 40, 18, 32,
35, 127, 28, 55, 27, 27, 70, 38, 41, 101, 7, 19, 45, 14, 12, 12, 8, 23, 76, 14,
17, 20, 27, 12, 20, 16, 32, 14, 277, 61, 32, 9, 10, 42, 14, 48, 27, 17, 73, 29,
14, 20, 43, 18, 46, 7, 17, 145, 23, 19, 24, 20, 79, 8, 24, 27, 26, 16, 38, 151,
28, 65, 23, 17, 27, 11, 20, 21, 37, 50, 22, 60, 32, 67, 438, 97, 10, 17, 24, 168,
6, 38, 21, 11, 8, 6, 7, 6, 38),
  E = c(22.3, 15.2, 21.3, 16.8, 21.2, 27.2, 11.7, 38.6, 25.3, 59.0, 124.4, 24.2, 34.1,
20.9, 24.6, 22.1, 23.6, 31.7, 31.6, 18.3, 46.2, 23.8, 23.3, 23.2, 20.3, 16.3, 57.9, 14.7, 18.0, 21.0,
27.8, 17.3, 52.4, 155.1, 172.4, 147.3, 79.0, 33.6, 112.4, 80.6, 113.4, 229.2, 123.0, 111.2, 83.4,
124.3, 83.1, 49.7, 106.7, 68.3, 13.5, 16.9, 49.5, 21.3, 36.3, 23.8, 17.4, 34.9, 4.2, 38.0, 21.1, 18.0,
21.0, 19.4, 50.2, 30.0, 57.8, 31.6, 44.2, 67.2, 177.7, 32.0, 42.3, 54.8, 52.2, 21.8, 25.7, 25.0, 18.1, 27.4,
35.2, 105.8, 19.2, 50.4, 30.6, 30.1, 60.7, 29.6, 48.7, 91.7, 10.7, 29.0, 43.2, 8.2, 17.4, 17.8, 18.2, 19.1, 90.0, 17.0,
19.2, 14.7, 20.4, 9.0, 24.9, 25.6, 39.9, 19.0, 239.1, 52.9, 30.8, 9.3, 10.1, 29.5, 18.1, 59.8, 35.2, 22.1, 103.1, 27.6,
19.8, 24.1, 58.2, 35.6, 60.0, 8.0, 40.3, 195.9, 39.7, 25.0, 26.2, 22.4, 63.1, 16.3, 33.0, 22.8, 22.3, 9.7, 39.0, 121.7,
34.9, 87.4, 27.6, 20.1, 22.9, 14.9, 28.2, 17.2, 30.2, 61.2, 16.6, 50.3, 19.7, 53.9, 292.4, 77.6, 15.5, 18.0, 21.6, 137.4,
8.8, 47.5, 15.5, 15.3, 13.2, 7.3, 5.3, 7.5, 33.6),
  num = c(2, 3, 5, 4, 5, 5, 3, 4, 3, 4, 5, 2, 4, 4, 4, 3, 6, 5, 6, 5,
11, 4, 4, 9, 3, 3, 4, 3, 3, 5, 3, 4, 6, 4, 3, 3, 4, 4, 5, 5,
5, 4, 6, 4, 4, 6, 6, 4, 4, 6, 1, 5, 5, 5, 5, 4, 6, 6, 5, 7,
5, 7, 4, 3, 4, 3, 6, 3, 5, 6, 5, 6, 8, 6, 6, 5, 7, 5, 5, 4,
5, 4, 4, 7, 6, 4, 6, 5, 4, 6, 4, 4, 4, 3, 6, 7, 5, 6, 3, 4,
5, 4, 6, 3, 5, 3, 9, 5, 7, 5, 5, 4, 6, 3, 4, 4, 4, 6, 6, 4,
3, 4, 4, 4, 4, 1, 7, 5, 6, 5, 8, 4, 7, 4, 5, 4, 6, 7, 5, 6,
5, 3, 4, 3, 2, 5, 6, 5, 6, 5, 5, 6, 6, 5, 5, 7, 4, 7, 5, 5,
5, 5, 2, 3, 4, 3, 1, 5, 0),
  adj = c(2, 14,
1, 3, 14,
2, 4, 11, 13, 14,
3, 5, 10, 11,
4, 6, 8, 9, 10,
```

5, 7, 8, 15, 33,  
6, 8, 33,  
5, 6, 7, 9,  
5, 8, 10,  
4, 5, 9, 11,  
3, 4, 10, 12, 13,  
11, 13,  
3, 11, 12, 14,  
1, 2, 3, 13,  
6, 33, 16, 17,  
15, 17, 33,  
15, 16, 18, 19, 32, 33,  
17, 19, 34, 45, 46,  
17, 18, 20, 30, 31, 32,  
19, 21, 30, 46, 47,  
20, 22, 24, 27, 28, 29, 30, 39, 40, 47, 48,  
21, 23, 24, 39,  
22, 24, 39, 49,  
21, 22, 23, 25, 26, 27, 49, 57, 58,  
24, 26, 58,  
24, 25, 27,  
21, 24, 26, 28,  
21, 27, 29,  
21, 28, 30,  
19, 20, 21, 29, 31,  
19, 30, 32,  
17, 19, 31, 33,  
6, 7, 15, 16, 17, 32,  
18, 35, 44, 45,  
34, 36, 43,  
35, 37, 43,  
36, 38, 42, 43,  
37, 39, 40, 41,  
21, 22, 23, 38, 40,  
21, 38, 39, 41, 48,  
38, 40, 42, 47, 48,  
37, 41, 43, 47,  
35, 36, 37, 42, 44, 46,  
34, 43, 45, 46,  
18, 34, 44, 46,  
18, 20, 43, 44, 45, 47,  
20, 21, 41, 42, 46, 48,  
21, 40, 41, 47,  
23, 24, 50, 57,  
49, 51, 52, 54, 55, 57,  
50,  
50, 53, 54, 62, 63,  
52, 54, 60, 61, 62,  
50, 52, 53, 55, 60,  
50, 54, 56, 57, 60,  
55, 57, 58, 60,  
24, 49, 50, 55, 56, 58,  
24, 25, 56, 57, 59, 60,  
58, 60, 61, 76, 83,  
53, 54, 55, 56, 58, 59, 61,  
53, 59, 60, 62, 76,  
52, 53, 61, 63, 73, 76, 77,  
52, 62, 64, 73,  
63, 65, 73,  
64, 66, 72, 73,  
65, 67, 72,  
66, 68, 69, 70, 71, 72,  
67, 69, 84,  
67, 68, 70, 84, 87,  
67, 69, 71, 75, 87, 88,  
67, 70, 72, 74, 75,  
65, 66, 67, 71, 73, 74,  
62, 63, 64, 65, 72, 74, 77, 79,  
71, 72, 73, 75, 79, 80,  
70, 71, 74, 81, 88, 89,

59, 61, 62, 77, 83,  
62, 73, 76, 78, 79, 82, 83,  
77, 79, 80, 81, 82,  
73, 74, 77, 78, 80,  
74, 78, 79, 81,  
75, 78, 80, 82, 89,  
77, 78, 81, 83,  
59, 76, 77, 82,  
68, 69, 85, 86, 87, 92, 97,  
84, 86, 90, 91, 97, 98,  
84, 85, 87, 90,  
69, 70, 84, 86, 88, 90,  
70, 75, 87, 89, 90,  
75, 81, 88, 90,  
85, 86, 87, 88, 89, 91,  
85, 90, 98, 100,  
84, 93, 96, 97,  
92, 94, 95, 96,  
93, 95, 103,  
93, 94, 96, 101, 102, 103,  
92, 93, 95, 97, 98, 99, 101,  
84, 85, 92, 96, 98,  
85, 91, 96, 97, 99, 100,  
96, 98, 100,  
91, 98, 99, 101,  
95, 96, 100, 102, 110,  
95, 101, 103, 110,  
94, 95, 102, 104, 109, 110,  
103, 105, 109,  
104, 106, 107, 108, 109,  
105, 107, 115,  
105, 106, 108, 113, 115, 119, 120, 121, 122,  
105, 107, 109, 112, 113,  
103, 104, 105, 108, 110, 111, 112,  
101, 102, 103, 109, 111,  
109, 110, 112, 113, 114,  
108, 109, 111, 113,  
107, 108, 111, 112, 114, 122,  
111, 113, 122,  
106, 107, 116, 119,  
115, 117, 118, 119,  
116, 118, 123, 129,  
116, 117, 119, 129, 137, 138,  
107, 115, 116, 118, 120, 137,  
107, 119, 121, 137,  
107, 120, 122,  
107, 113, 114, 121,  
117, 124, 128, 129,  
123, 125, 127, 128,  
124, 126, 127, 133,  
125,  
124, 125, 128, 130, 131, 132, 133,  
123, 124, 127, 129, 130,  
117, 118, 123, 128, 130, 138,  
127, 128, 129, 131, 138,  
127, 130, 132, 138, 139, 150, 151, 156,  
127, 131, 133, 151,  
125, 127, 132, 134, 151, 152, 153,  
133, 135, 136, 153,  
134, 136, 153, 159, 160,  
134, 135, 161, 162,  
118, 119, 120, 138, 140, 141,  
118, 129, 130, 131, 137, 139, 140,  
131, 138, 140, 149, 150,  
137, 138, 139, 141, 148, 149,  
137, 140, 142, 147, 148,  
141, 143, 147,  
142, 144, 146, 147,  
143, 145, 146,  
144, 146,



### ANEXO 3

#### Salidas de Excel obtenidas de aplicar la metodología clásica en el cálculo del riesgo relativo

#	Municipios	SMRs Brutas
1	sandino	1,34529148
2	mantua	0,526315789
3	m. matahambre	1,220657277
4	viñales	0,952380952
5	la palma	1,132075472
6	bahia honda	0,551470588
7	candelaria	0,256410256
8	san cristobal	0,67357513
9	los palacios	0,671936759
10	c. del sur	1,288135593
11	pinar del rio	0,884244373
12	san luis	1,033057851
13	san. j martinez	0,879765396
14	guane	0,956937799
15	mariel	1,138211382
16	guanajay	0,859728507
17	caimito	1,101694915
18	bauta	0,883280757
19	san a. baños	0,886075949
20	bejucal	0,928961749
21	san j. de lajas	0,952380952
22	jaruco	1,050420168
23	santa c.del norte	0,815450644
24	madrugá	1,163793103
25	nueva paz	1,280788177
26	san nicolas	1,288343558
27	guines	1,27806563
28	melena del sur	1,496598639
29	batabano	0,722222222
30	quivicán	0,857142857
31	guira de melena	1,223021583
32	alquizar	0,867052023
33	artemisa	0,858778626
34	playa	0,986460348
35	p. revolucion	0,997679814
36	centro habana	1,16089613
37	la habana vieja	1,101265823
38	regla	1,220238095
39	la habana del este	1,040925267
40	guanabacoa	1,104218362
41	san m. del padron	1,040564374

#	Municipios	SMRs Brutas
42	10 de octubre	1,034031414
43	cerro	1,113821138
44	marianao	1,133093525
45	la lisa	1,03117506
46	boyeros	1,013676589
47	arroyo naranjo	2,081829122
48	cotorro	1,187122736
49	matanzas	0,768509841
50	cardenas	0,966325037
51	varadero	0,962962963
52	marti	0,887573964
53	colon	0,767676768
54	perico	1,079812207
55	jovellanos	1,239669421
56	pedro betancourt	0,798319328
57	limonar	1,091954023
58	union de reyes	0,945558739
59	cienaga de zapata	0,476190476
60	jaguey grande	0,552631579
61	calimete	0,758293839
62	los arabos	0,555555556
63	corralillo	0,857142857
64	q. de guines	0,618556701
65	sagua la grande	0,737051793
66	encrucijada	0,7
67	camajuani	0,865051903
68	caibarien	0,949367089
69	remedios	0,746606335
70	placetas	0,729166667
71	santa clara	0,89476646
72	cifuentes	0,6875
73	santo domingo	0,756501182
74	ranchuelo	0,930656934
75	manicaragua	0,938697318
76	aguada de p	0,963302752
77	rodas	0,661478599
78	palmira	1,6
79	lajas	0,994475138
80	cruces	1,167883212
81	cumanayagua	0,994318182
82	cienfuegos	1,200378072

#	Municipios	SMRs Brutas
83	abreus	1,458333333
84	yaguajay	1,091269841
85	jatibonico	0,882352941
86	taguasco	0,897009967
87	cabaiguan	1,153212521
88	fomento	1,283783784
89	trinidad	0,841889117
90	sancti spiritus	1,101417666
91	la sierpe	0,654205607
92	chambas	0,655172414
93	moron	1,041666667
94	bolivia	1,707317073
95	primero de enero	0,689655172
96	ciro redondo	0,674157303
97	florencia	0,43956044
98	majagua	1,204188482
99	ciego de avila	0,844444444
100	venezuela	0,823529412
101	baragua	0,885416667
102	c. m. de cespedes	1,360544218
103	esmeralda	1,323529412
104	sierra de cubitas	1,333333333
105	minas	0,803212851
106	nuevitas	0,625
107	guaimaro	0,802005013
108	sibanicu	0,736842105
109	camaguey	1,158511083
110	florida	1,153119093
111	vertientes	1,038961039
112	jimaguayu	0,967741935
113	najasa	0,99009901
114	santa cruz del sur	1,423728814
115	manati	0,773480663
116	puerto padre	0,802675585
117	jesus menendez	0,767045455
118	majibacoa	0,769230769
119	las tunas	0,708050436
120	jobabo	1,050724638
121	colombia	0,707070707
122	a. rodriguez	0,829875519
123	gibara	0,738831615
124	rafael freyre	0,505617978
125	banes	0,766666667
126	antilla	0,875

#	Municipios	SMRs Brutas
127	baguanos	0,421836228
128	holguin	0,740173558
129	calixto garcia	0,579345088
130	cacocum	0,76
131	urbano noris	0,916030534
132	cueto	0,892857143
133	mayari	1,251980983
134	frank pais	0,490797546
135	sagua de tanamo	0,727272727
136	moa	1,184210526
137	rio cauto	1,165919283
138	cauto cristo	1,649484536
139	jiguani	0,974358974
140	bayamo	1,240755957
141	yara	0,802292264
142	manzanillo	0,743707094
143	campechuela	0,833333333
144	media luna	0,845771144
145	niquero	1,179039301
146	pilon	0,738255034
147	bartolome maso	0,709219858
148	buey arriba	1,220930233
149	guisa	1,225165563
150	contramaestre	0,816993464
151	mella	1,325301205
152	san luis	1,192842942
153	segundo frente	1,624365482
154	songo la maya	1,243042672
155	s. de cuba	1,497948016
156	palma soriano	1,25
157	tercer frente	0,64516129
158	guama	0,944444444
159	el salvador	1,111111111
160	guantanamo	1,222707424
161	yateras	0,681818182
162	baracoa	0,8
163	maisi	1,35483871
164	imias	0,718954248
165	san a. del sur	0,606060606
166	manuel tames	0,821917808
167	caimanera	1,320754717
168	niceto perez	0,8
169	i de la juventud	1,130952381

## ANEXO 4

### Salidas del WinBUGS 1.4 obtenidas de aplicar la metodología bayesiana en el cálculo del riesgo relativo.

model is syntactically correct	RR[37]	1.032775034939811
data loaded	RR[38]	1.535199439094386
model compiled	RR[39]	1.189991225776745
model is initialized	RR[40]	1.127870267852249
RR[1] 1.057655339069627	RR[41]	1.074770340849583
RR[2] 0.8043656549511383	RR[42]	1.019194914913713
RR[3] 0.9560850051260041	RR[43]	1.082876410020525
RR[4] 1.056664854425874	RR[44]	1.054277809993766
RR[5] 0.9525133859891475	RR[45]	1.24289782456483
RR[6] 0.8351941557231463	RR[46]	1.089603735935869
RR[7] 0.6636280536838112	RR[47]	1.684395803044239
RR[8] 0.7769831605796481	RR[48]	1.267829657263943
RR[9] 0.7806488729971196	RR[49]	0.7783220336774144
RR[10] 1.453457781033609	RR[50]	0.9368837541508381
RR[11] 0.8649120487383039	RR[51]	0.7098514865527216
RR[12] 0.8894290351895872	RR[52]	0.8134009420767793
RR[13] 0.6886634897875318	RR[53]	0.7213501369527363
RR[14] 1.045883428657047	RR[54]	0.7608539579791845
RR[15] 0.9570803493410092	RR[55]	1.101561232531768
RR[16] 0.8947892172622842	RR[56]	0.8321565950022575
RR[17] 1.010228391773138	RR[57]	0.8318410054672045
RR[18] 1.088923480644541	RR[58]	0.9797224670044242
RR[19] 0.885470106720285	RR[59]	0.670398703546827
RR[20] 1.183384253493658	RR[60]	0.7159009049577475
RR[21] 1.063455694422203	RR[61]	1.002449769593764
RR[22] 0.7866303671627974	RR[62]	0.677125643409674
RR[23] 0.773439407103946	RR[63]	0.6128030393865621
RR[24] 0.9199617723924368	RR[64]	0.5965374176517388
RR[25] 1.362101795650043	RR[65]	0.7905653471048137
RR[26] 0.9014640074615133	RR[66]	0.9419370576397982
RR[27] 1.197236713502839	RR[67]	0.9717493683779919
RR[28] 1.270023668454425	RR[68]	0.839424690650003
RR[29] 1.053918937143587	RR[69]	0.8337155707350996
RR[30] 0.9564055880480858	RR[70]	0.7669072847074183
RR[31] 1.03390406664641	RR[71]	0.8601878853818821
RR[32] 0.746469165270339	RR[72]	0.9636776870915551
RR[33] 0.888058324022536	RR[73]	0.6978668741813584
RR[34] 1.069626173961675	RR[74]	1.125309567450534
RR[35] 1.093038271620124	RR[75]	0.873607204259681
RR[36] 1.046887620293197	RR[76]	0.7269103729557775

RR[77]	0.9222177241079162	RR[124]	0.4619896709180821
RR[78]	1.246963510963581	RR[125]	0.7393845585746207
RR[79]	0.8245997045580507	RR[126]	1.110373040942709
RR[80]	0.9043433525587899	RR[127]	0.8171428915056846
RR[81]	1.094883356299626	RR[128]	0.6915819540728491
RR[82]	1.160692854248416	RR[129]	0.7310496871078973
RR[83]	1.261824770060987	RR[130]	0.9126647782734464
RR[84]	1.033265505000974	RR[131]	0.9493902204340947
RR[85]	0.9023098155887588	RR[132]	1.155268822473846
RR[86]	0.9728672917027008	RR[133]	1.149553224247732
RR[87]	0.8351196697552956	RR[134]	0.68633604812022
RR[88]	1.088377499473056	RR[135]	0.8186371926594036
RR[89]	0.848683650727331	RR[136]	0.8644835362197757
RR[90]	0.9940024897479327	RR[137]	1.023570498098362
RR[91]	0.6046692450748894	RR[138]	0.9608250075822528
RR[92]	0.8084886708154351	RR[139]	1.088447913904729
RR[93]	0.9744755174659505	RR[140]	1.103833877119771
RR[94]	0.8833322393287252	RR[141]	0.6860552703656005
RR[95]	0.7524755670329656	RR[142]	0.7147687650275059
RR[96]	0.7298079895253996	RR[143]	0.6279763743573183
RR[97]	0.6596872258731815	RR[144]	1.006571682722194
RR[98]	0.7979407163485808	RR[145]	1.293914263310702
RR[99]	0.8240476699086228	RR[146]	0.9095297648422668
RR[100]	0.9368826289817045	RR[147]	0.7926209983318749
RR[101]	0.9234450425568816	RR[148]	0.9869985062446444
RR[102]	1.210208492702288	RR[149]	1.053072233343896
RR[103]	0.8858224084958094	RR[150]	1.056984879342537
RR[104]	0.9514070151192783	RR[151]	1.453266809874995
RR[105]	0.8068259582001873	RR[152]	1.257378053470903
RR[106]	0.7331662544664694	RR[153]	1.01119221393609
RR[107]	0.6958244103064657	RR[154]	1.301082564743386
RR[108]	0.9044180752117844	RR[155]	1.481648005694098
RR[109]	1.189964977706175	RR[156]	1.299295342601746
RR[110]	1.168959963311472	RR[157]	0.7320665744754604
RR[111]	0.9425383643672086	RR[158]	1.013390700805806
RR[112]	1.155451968265028	RR[159]	0.9871481928569099
RR[113]	0.7657794257855025	RR[160]	1.126228885105228
RR[114]	0.9427706557083235	RR[161]	0.8890563122004537
RR[115]	0.6702127621710949	RR[162]	0.9569684521385015
RR[116]	0.7137723601671492	RR[163]	0.7731936764348339
RR[117]	0.7074854863681881	RR[164]	0.648203099386811
RR[118]	0.8317637744691119	RR[165]	0.6921484515930971
RR[119]	0.8209753118013839	RR[166]	1.058411018562678
RR[120]	1.286079655257339	RR[167]	0.8278729269405186
RR[121]	0.767549479503422	RR[168]	1.142111193033644
RR[122]	0.9367366840792086	RR[169]	0.921037709464926
RR[123]	0.5618166242019045		