



UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO

---

---

FACULTAD DE CIENCIAS

Análisis espacio-temporal de incidencias  
del SIDA en hombres. México, nivel  
estatal. 1998-2007

T E S I S

QUE PARA OBTENER EL TÍTULO DE:  
ACTUARIO

PRESENTA:  
FRANCISCO JAVIER MÉNDEZ CORREA

DIRECTOR DE TESIS:  
DR. CARLOS DÍAZ ÁVALOS



2011



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



## Hoja de datos del jurado

1. Datos del alumno  
Méndez  
Correa  
Francisco Javier  
Universidad Nacional Autónoma de México  
Facultad de Ciencias
2. Datos del tutor  
Dr  
Carlos  
Díaz  
Ávalos
3. Datos del sinodal 1  
Dr  
Ramsés Humberto  
Mena  
Chávez
4. Datos del sinodal 2  
Dra  
Silvia  
Ruiz-Velazco  
Acosta
5. Datos del sinodal 3  
Act  
Harim  
García  
Lamont
6. Datos del sinodal 4  
Dra  
Ruth Selene  
Fuentes  
García
7. Datos del trabajo escrito  
Análisis espacio-temporal de incidencias del SIDA en hombres. México, nivel estatal. 1998-2007  
84 p  
2011



## Agradecimientos

A María Fernanda Agoitia, mi tutora escolar y gran amiga, responsable de la mitad de mis créditos, servicio social y mi formación académica dentro de esta magna casa de estudios.

A Timón, Jiman, Noob, Tulemon, Animal, Mevendez y Enrique por trazar conmigo en esta etapa de mi vida.

A Lalo por la gran ayuda que me brindo en la parte final de la carrera, y por el Torres que deberá pagar.

A mis papas y mi hermana por el apoyo incondicional que me han brindado a lo largo de toda mi vida.

Al Dr. Carlos Díaz Ávalos por su gran ayuda, asesoría, apoyo y en especial por su paciencia en la elaboración de esta tesis.

A la Dra. Ruth Fuentes, Dr.a Silvia Ruiz-Velazco, Dr. Ramsés Mena y Act. Harim García por sus valiosas aportaciones para la conclusión de esta tesis.

Y a ti, por leer esta tesis.

# Índice general

<b>1</b>	<b>VIH y SIDA</b>	<b>11</b>
1.1	Descripción de la enfermedad . . . . .	11
1.1.1	Historia natural de la enfermedad . . . . .	11
1.1.2	Mecanismos de transmisión del VIH . . . . .	13
1.2	El VIH desde una perspectiva histórica . . . . .	13
1.3	Distribución de la Epidemia en el Mundo . . . . .	14
1.3.1	Mortalidad por SIDA . . . . .	16
1.4	La epidemia en México . . . . .	17
1.4.1	Grupos de riesgo . . . . .	18
1.4.2	Atención a pacientes . . . . .	19
1.4.3	Mortalidad . . . . .	20
1.4.4	Retos epidemiológicos . . . . .	20
<b>2</b>	<b>Estadística espacial</b>	<b>23</b>
2.1	Definición y motivación . . . . .	23
2.2	Modelos espaciales . . . . .	24
2.3	Sistemas de lattice . . . . .	25
2.4	Vecino más cercano (Nearest Neighbor) . . . . .	26
2.5	Teorema de Hammersley-Clifford . . . . .	26
2.6	Campos aleatorios de Markov . . . . .	29
2.7	Lema de Brook . . . . .	30
2.8	Automodelos o modelos condicionales autorregresivos . . . . .	31
2.8.1	Esquema auto-poisson . . . . .	32
2.8.2	Auto-normal . . . . .	33
2.9	Procesos espacio-temporales . . . . .	34
<b>3</b>	<b>Herramientas bayesianas</b>	<b>35</b>
3.1	Estadística Bayesiana . . . . .	35

3.2	Modelos jerárquicos . . . . .	35
3.2.1	Modelos jerárquicos espaciales . . . . .	37
3.3	MCMC . . . . .	38
3.3.1	Método de Monte Carlo . . . . .	38
3.3.2	Cadenas de Markov . . . . .	38
3.3.3	Metropolis-Hasting . . . . .	40
3.3.4	Gibbs Sampler . . . . .	40
<b>4</b>	<b>Análisis exploratorio de los datos</b>	<b>43</b>
4.1	Descripción de los datos . . . . .	43
4.2	Análisis global de las variables . . . . .	44
4.2.1	Análisis de la matriz de correlación . . . . .	44
4.2.2	Dendograma . . . . .	47
4.3	Defunciones por SIDA . . . . .	49
4.3.1	Serie de tiempo . . . . .	49
4.3.2	Proporción hombre / mujer . . . . .	49
4.3.3	Análisis por estado . . . . .	49
<b>5</b>	<b>Análisis espacio temporal</b>	<b>59</b>
5.1	Descripción del modelo . . . . .	59
5.1.1	Variables . . . . .	59
5.2	Modelo jerárquico . . . . .	60
5.3	Estimación de parámetros . . . . .	63
5.4	Resultados . . . . .	65
5.5	Mapas de riesgo . . . . .	67
5.6	Series de tiempo . . . . .	68
5.7	Observaciones y comparaciones . . . . .	69
5.8	Análisis por estado . . . . .	71
5.8.1	Veracruz . . . . .	72
5.8.2	D.F. . . . .	74
5.8.3	Quintana Roo . . . . .	75
5.8.4	Jalisco . . . . .	76
5.8.5	Baja California . . . . .	76
<b>6</b>	<b>Conclusiones</b>	<b>79</b>
<b>A</b>	<b>Apéndice</b>	<b>81</b>

# Índice de cuadros

4.1	Matriz de correlación de las variables involucradas en el estudio realizado en esta tesis.(La descripción de las variables se encuentran al inicio de este capítulo.)	44
5.1	Valores de los cuantiles 2.5 % y 97.5 % de las distribuciones de las $\beta's$	66
5.2	Tabla de cuantiles y moda de las distribuciones finales correspondientes al riesgo relativo multiplicado por 1,000,000 de Veracruz y la moda de sus estados vecinos mas representativos desde 1998 hasta 2007.	76
5.3	Tabla de cuantiles y moda de las distribuciones finales correspondientes al riesgo relativo del D.F. multiplicado por 1,000,000 y la moda de sus estados vecinos desde 1998 hasta 2007.	77
5.4	Tabla de cuantiles y moda de las distribuciones finales correspondientes al riesgo relativo multiplicado por 1,000,000 de Quintana Roo y la moda de sus estados vecinos desde 1998 hasta 2007.	77
5.5	Tabla de cuantiles y moda de las distribuciones finales correspondientes al riesgo relativo multiplicado por 1,000,000 de Jalisco y la moda de sus estados vecinos mas importantes desde 1998 hasta 2007.	78
5.6	Tabla de cuantiles y moda de las distribuciones finales correspondientes al riesgo relativo multiplicado por 1,000,000 de Baja California y la moda de Baja California Sur, su estado vecino mas importante, desde 1998 hasta 2007.	78

# Introducción

La epidemia del SIDA ha causado en las últimas décadas mas de 20 millones de muertes, lo que es uno de los grandes problemas a los que se enfrentan los sistemas de salud de todo el mundo.[13]

El reto que representa controlar la epidemia en México es muy grande, esto causado principalmente por la falta de tratamientos efectivos que controlen la enfermedad, los mecanismos de transmisión y a la ignorancia de muchas personas infectadas de VIH respecto a su condición. A pesar que la epidemia en nuestro país afecta mayormente a ciertos grupos muy específicos, el resto de la población no es ajena al riesgo.

Es por eso que entender el comportamiento espacio-temporal de la enfermedad proporciona una herramienta para modelar y predecir, generando información que permite acciones de prevención concretas y funcionales para controlar la epidemia.

El objetivo de esta tesis es modelar espacial y temporalmente el impacto de mortalidad ocasionado por esta epidemia, analizando el número de muertes de SIDA por sexo y estado, durante el periodo de 1998 a 2007 registrados por el IMSS, con ayuda de covariables demográficas, mediante el uso de modelos jerárquicos espaciales.

Para esto, el trabajo fue estructurado en 6 capítulos. En el primero se centra la atención en el VIH y la enfermedad que ocasiona; el SIDA, exponiendo de forma general la descripción del virus, la historia natural de la enfermedad, el desarrollo histórico de la epidemia y la situación actual que el mundo esta enfrentando. En los siguientes capítulos se exponen las herramientas matemáticas necesarias para la construcción de modelo.

El capítulo 2 consiste en un resumen que abarca de forma global la estadística espacial, haciendo especial énfasis en los conceptos de mayor trascendencia para este trabajo. En el tercer capítulo estan plasmados tanto la noción de modelos jerárquicos como el algoritmo

del método de Monte Carlo con Cadenas de Markov (MCMC), que es la vía mediante la cuál se buscó estimar los parámetros del modelo sugerido.

El cuarto capítulo se hace un análisis exploratorio de todas aquellas variables demográficas y sociales que pudieran tener un impacto con el desarrollo de la epidemia y por tanto de las muertes causadas. De la misma forma fueron analizadas las muertes ocasionadas por el SIDA de forma temporal, según sexo y estado.

En el capítulo 5 se explica la construcción del modelo jerárquico espacio-temporal, usado para poder encontrar el riesgo relativo por estado en cada periodo de tiempo analizado.

Las conclusiones estan expuestas en el último capítulo.

Las referencias usadas para la elaboración de este trabajo estan enlistadas en la última sección. Debido a que se hacen referencias a lo largo de este trabajo, dicho material fue enumerado para facilitar su localización y hacer más práctica la forma de mencionar que cierta información fue elaborada con el apoyo de un documento esfecifico mencionado en la bibliografía.

# Capítulo 1

## VIH y SIDA

En este primer capítulo se exponen las principales características del VIH, así como el panorama general de la epidemia del SIDA. Para ello se usaron como referencias principales documentos publicados por ONUSIDA [13] CENSIDA [24] y artículos del IMSS. [14]

### 1.1. Descripción de la enfermedad

El Síndrome de la Inmunodeficiencia Humana (SIDA) es una enfermedad que consiste en la incapacidad del sistema inmunológico para combatir a las infecciones y otros agentes patológicos. Es provocado por el Virus de la Inmunodeficiencia Humana (VIH), el cual ataca a una célula del sistema inmunológico conocida como linfocito CD4.

Actualmente se conocen dos tipos de virus que causan el SIDA en el ser humano el VIH-1 y VIH-2, estos virus comparten solo el 40% de homología, por su parte el VIH-2 comparte alrededor del 75% de homología con dos razas conocidas de Virus de Inmunodeficiencia Simiae (VIS).[7]

#### 1.1.1. Historia natural de la enfermedad

Se llama historia natural de la enfermedad al curso que sigue un cierto proceso patológico a lo largo del tiempo cuando no se aplica ningún elemento externo para intentar corregirlo o detenerlo.

El VIH ocasiona múltiples manifestaciones clínicas, estas van desde la ausencia de síntomas, hasta tumores. Se dice que una persona tiene SIDA en el momento que a causa del VIH, su sistema inmunológico ha sufrido un daño, cuya consecuencia es la incapacidad para defenderse contra agentes patógenos, que bajo condiciones normales no producen enfermedades graves.

El desarrollo de cualquier enfermedad incluyendo el SIDA, es el resultado de tres elementos: el agente causal, el huésped y el medio ambiente.

El *agente* que ocasiona el SIDA es el VIH, el cual pertenece a la familia de los retrovirus, los cuales tienen como característica principal convertir a su material genético de ARN a ADN, este proceso es conocido como transcripción reversiva. Dentro de esta familia el VIH se encuentra dentro del género de lentivirus, ya que como su nombre lo indica, el periodo de incubación del virus<sup>1</sup> es muy largo.[7]

Una vez que el VIH se encuentra dentro de un *huesped*, es decir el sujeto infectado, localiza y penetra las membranas plasmáticas de varios tipos de células del sistema inmunológico, pero en especial a los linfocitos T4, en los que se alojará iniciando el proceso de destrucción celular. A partir de este momento, el proceso que sigue el virus para atacar al huésped se divide en cuatro etapas.

La primera etapa es conocida como *infección aguda* e inicia desde el primer momento que el VIH penetra al organismo. En esta etapa pocas veces se presentan algunos síntomas expresados en malestares generales, debido a la replicación del virus. Posteriormente se presenta la segunda etapa conocida como *infección asintomática*, en el cual el virus se adhiere a las células y duplica su código genético. La siguiente etapa es llamada *linfadenopatía generalizada persistente*, conocida así porque el término de linfadenopatía se refiere a la enfermedad dentro de los ganglios linfáticos, miembros del sistema inmunológico. La cuarta etapa es el SIDA, donde el equilibrio de las subpoblaciones de linfocitos se pierde y el sistema inmunológico es incapaz de contener la aparición de infecciones por gérmenes oportunistas, ni capaz de impedir el desarrollo de tumores malignos, principalmente linfomas de células B y sarcoma de Kaposi. Es bajo el porcentaje de pacientes que desarrollan cuadros graves que los llevan rápidamente a la muerte. El resto de los pacientes tiene un deterioro en el cuerpo mas lento, que va desde los 8 meses hasta los dos años.

---

<sup>1</sup>EL periodo de incubación se refiere al intervalo de tiempo entre la exposición a un agente infeccioso y la aparición del primer signo o síntoma de la enfermedad.

Por su parte el *medio ambiente* en el SIDA es en su mayoría social, porque se considera una enfermedad de comportamiento. También existen ciertos factores como los hábitos, las costumbres sexuales, creencias, condiciones sociales de la población como la migración y la infraestructura sanitaria entre otros. [13]

### 1.1.2. Mecanismos de transmisión del VIH

Al ser el SIDA una epidemia a nivel mundial es importante conocer los mecanismos de transmisión del VIH que ONUSIDA en el 2009 ha identificado.

- transmisión sexual
- vía sanguínea (comunmente en los drogadictos por vía intravenosa y los receptores de transfusiones)
- transmisión perinatal
- trasplantes de tejidos u órganos

## 1.2. El VIH desde una perspectiva histórica

Desde la detección del VIH y del SIDA se han creado múltiples teorías sobre su aparición, aunque a pesar de que son muy variadas, todas apuntan a que el SIDA empezó en África.

La explicación más aceptada del origen del SIDA, se basa en la transferencia de cruce de especies, e inicia cuando Noireau en 1987, quien trataba con hábitos sexuales de la gente en África mencionó que "Para estimular a un hombre o a una mujer e inducirlos a una intensa actividad sexual, era directamente inoculada sangre de mono macho para hombre o sangre de mono hembra para mujer en el área púvica y también en los muslos y la espalda." Tales prácticas pudieron haber generado un medio eficaz de transmisión trans-especies y podrían ser responsables de la aparición de la infección del hombre por VIS y de esa manera de SIDA. Con esto se llega a pensar que el VIS pudo haber ocasionado la infección del VIH-2 del hombre del África del Oeste y que un virus distinto a los que son propios del mono, pudo ser aislado, el cual es similar a VIH-1.[7]

El mono africano es el huésped natural de los lentivirus, dentro de los cuales se encuentra el VIH, esto significa que el organismo del mono vive en equilibrio con el virus. Solamente

llega a ser mortal, después de transmitirse a otras especies, entre ellas los hombres.

Aunque no es posible determinar cuando se infectó la primera persona de VIH, el hecho de buscar un inicio histórico a la epidemia es mucho más fácil de conocer. La evidencia epidemiológica más grande apunta a la propagación de la infección por VIH en África desde la segunda guerra mundial. Esta propagación parece coincidir con la introducción masiva de jeringas y agujas junto a los programas de vacunación y antibióticos. Este periodo coincide con el mejoramiento de la comunicación y la migración masiva en África. Con esto se puede suponer que el virus se llevo a otras areas aisladas, mas tarde a poblaciones urbanas y finalmente mediante contactos sexuales a otros países del otro lado del oceano.

De los primeros casos conocidos de personas infectadas de VIH-1 fuera de África, esta el de un marino en Noruega, que probablemente llegó a infectarse a mediados de los 60's, y de quien su esposa y tercera hija murieron de lo que se pudiera considerar actualmente como SIDA.

En América el primer signo se dio en forma de un cáncer raro conocido como sarcoma de Kaposi, entre hombres homosexuales de Nueva York. Este tipo de cáncer solo se habia presentado en personas mayores y en África. En 1982 en EUA se reportó un síndrome que incluía infecciones oportunistas y la reducción de células T4, entre otros. Al mismo tiempo se descubrió que el SIDA se estaba propagando entre usuarios de drogas intravenosas, hemofílicos, algunos transfusores de sangre, y también en países del Caribe. Hasta 1983 se detectaron los primeros casos entre heterosexuales.

Fue hasta mediados de 1983 cuando en el Instituto Pasteur de Francia se pudo identificar al virus causante del SIDA, y hasta 1984 se acordó llamarlo VIH para resolver los problemas de los múltiples nombres.[7]

### 1.3. Distribución de la Epidemia en el Mundo

Todo parece indicar que la propagación del SIDA comenzó a finales de la década de los setentas, en África Oriental y Central entre hombres y mujeres sexualmente promiscuos, en América, Europa Occidental y Oceania entre hombre homosexuales y bisexuales, siendo el VIH-1 el tipo de virus predominante en el mundo causante de SIDA. El número de personas que se han infectado de VIH a lo largo del tiempo ha ido en aumento, en la figura 1.1 se muestra la estimación del comportamiento que ha tenido la epidemia desde 1990 realizada

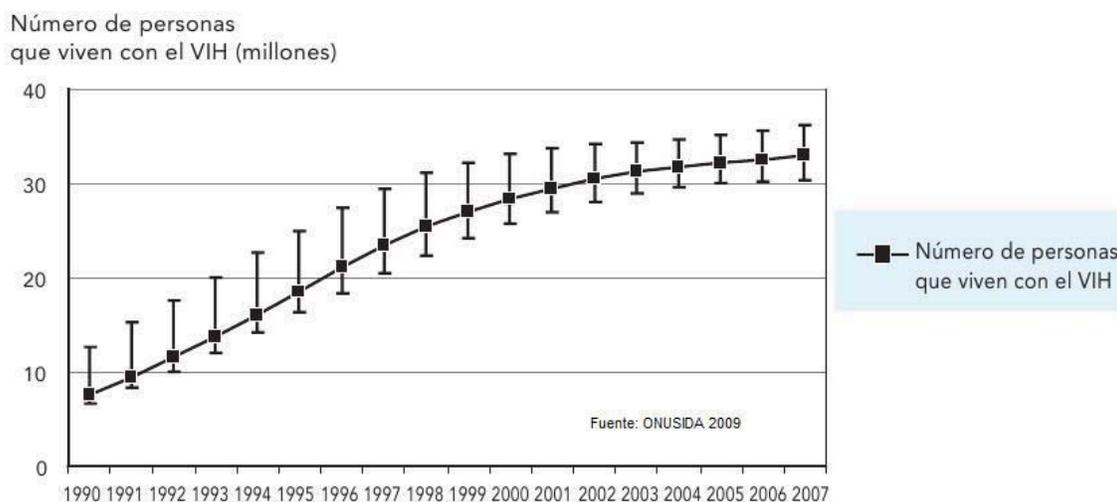


Figura 1.1: Estimación del número de personas en todo el mundo con que viven con el VIH en el periodo de 1990 a 2007, con un intervalo del 90 % de confianza

por ONUSIDA, donde las barras denotan un intervalo del 90 % de confianza de la estimación.

Como la epidemia se ha propagado de diversas formas dependiendo la región geográfica, la Organización Mundial de la Salud (OMS) a partir de 1988, ha analizado a cada región geográfica de forma particular para detectar estos patrones regionales, y obtener una mejor comprensión de la epidemia. A continuación se presenta un breve resumen de la situación regional de todo el mundo expuesto en el *Informe sobre la epidemia mundial de SIDA 2008* publicado por ONUSIDA [13].

- **África Subsahariana:** El coito heterosexual es la principal fuerza impulsora de las epidemias en en esta región, y esto ha generado una mayor población de niños que viven con VIH. Esta zona geográfica ha recibido el impacto demográfico más significativo causado por este virus. Además de contener a los países con mayor prevalencia del VIH, Swazilandia y Botswana. Actualmente las epidemias parecen haberse estabilizado en la mayoría de los países según los reportes de la OMS en 2008.
- **Asia:** Al existir diversas vías de transmisión del VIH (uso de drogas inyectables, relaciones sexuales sin protección) en Asia, hacen que la epidemia en este continente sea una de las más heterogeneas.
- **Europa Oriental y Asia Central :** Casi el 90 % de personas infectadas viven en

Rusia y Ucrania. Aquí las epidemias continúan creciendo. Las principales causas de transmisión se deben a los usuarios de droga inyectables, profesionales del sexo y diversas parejas sexuales.

- **Caribe:** Se estima que el 75 % de personas infectadas viven en República Dominicana y Haití. La principal vía de transmisión del VIH es el coito heterosexual, aunque también las relaciones sexuales entre hombres, en general negadas por la sociedad, constituyen un factor significativo.
- **América Latina :** En esta zona la transmisión se da principalmente entre hombres que tienen relaciones sexuales con hombres, profesionales del sexo y en menor medida, entre usuarios de drogas inyectables.
- **América del Norte y Europa Occidental y Central:** Aquí el acceso al tratamiento antiretroviral ha evitado muertes aceleradas, generado un incremento en el número de personas que viven con el VIH. La principal fuente de transmisión al igual que en América Latina son las relaciones sexuales sin protección entre hombres.
- **África del Norte y Oriente Medio:** Existe una limitada información de las epidemias en esta región, a excepción de Sudán, las epidemias son comparativamente pequeñas. Son diversas la combinación de factores que se asocian con la epidemia, pero destacan las relaciones sexuales remuneradas sin protección y el uso de equipos contaminados para la inyección de drogas.
- **Oceania:** La mayoría de la región tiene epidemias pequeñas excepto Papua Nueva Guinea. Aquí el coito heterosexual sin protección es la principal vía de transmisión de VIH.

### 1.3.1. Mortalidad por SIDA

A lo largo de la historia el VIH ha provocado aproximadamente 25 millones de fallecimientos en todo el mundo, siendo África el continente más afectado.[13]

Es por esto que un factor determinante en la epidemia es el tratamiento a los pacientes infectados con VIH, que además de aumentar la esperanza de vida a las personas con VIH, les proporcionan una mejor calidad de vida, en especial en los países más afectados ligados a una pobreza extrema.

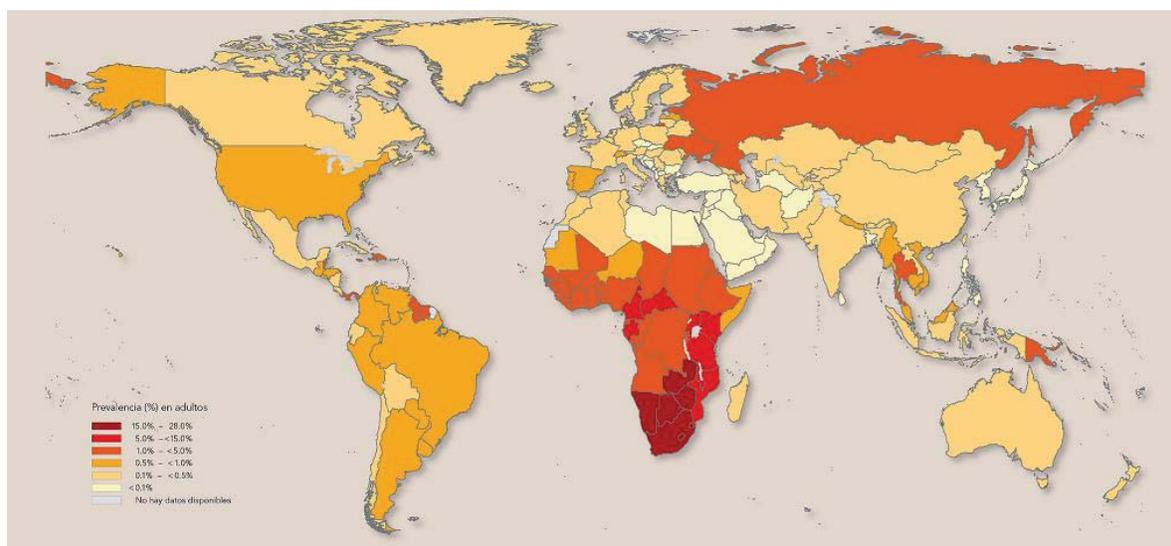


Figura 1.2: Mapa mundial publicado por ONUSIDA en el 2009 donde expone la prevalencia del VIH (%)

La atención a pacientes con SIDA ha pasado por diferentes etapas, de acuerdo a la evolución de la epidemia, el avance científico y tecnológico así como en las alternativas disponibles. A pesar de esto, muchos de los países afectados no han tenido el acceso requerido a este tipo de tratamientos debido a falta de recursos económicos o falta de personal suficiente para proporcionar dichos tratamientos.

## 1.4. La epidemia en México

En México el primer caso de SIDA registrado por el sistema de salud pública fue diagnosticado en 1983, aunque de acuerdo con diferentes técnicas de investigación, el inicio de la epidemia puede ubicarse en 1981.[14]

Desde los inicios de la epidemia hasta el 2009 se han contabilizado 135,003 casos acumulados de SIDA, de las cuales el 82.3 % son hombres y el 17.7 % son mujeres. La evolución de la razón hombre/mujer se ha reducido a lo largo de la epidemia. Disminuyó de 10.8 en 1985 hasta 3.6 en el 2008.[24]

La epidemia en México ha tenido una transmisión predominantemente sexual, ya que

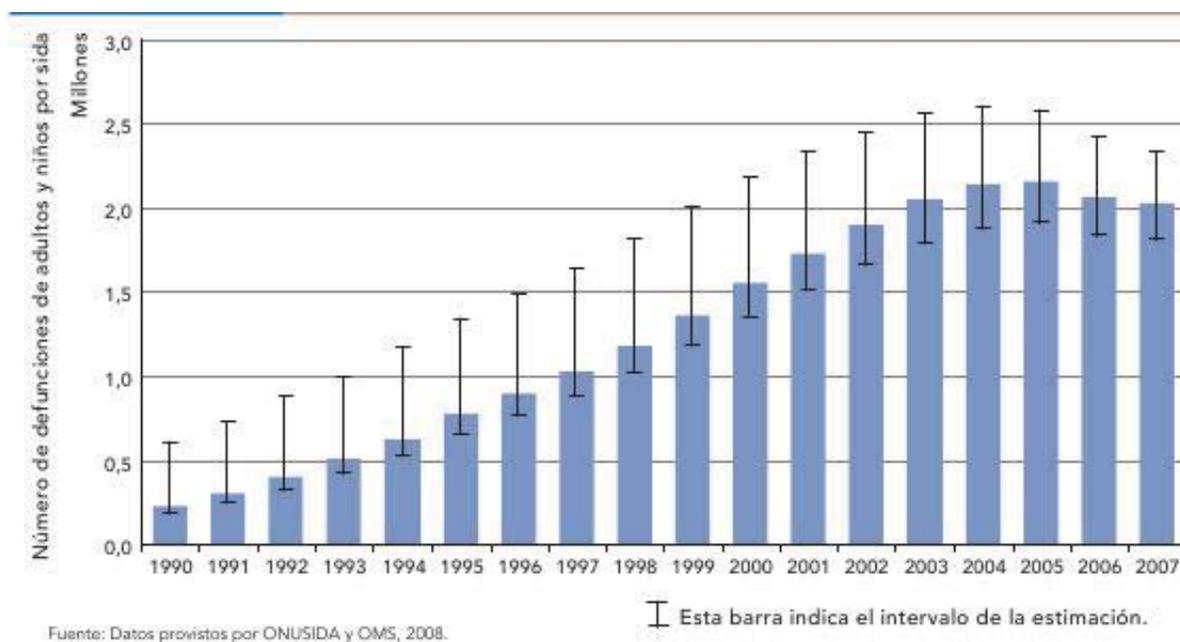


Figura 1.3: Estimación del número de defunciones por SIDA a nivel mundial desde 1990 a 2007, con un intervalo de confianza del 90 %.

este ha sido el causante de mas del 90 % del total de casos acumulados.

#### 1.4.1. Grupos de riesgo

En el 2009 en México el IMSS registró 220,000 personas adultas infectadas por el VIH, es decir la prevalencia de VIH en nuestro país en personas de 15 a 49 años en ese año fue del 0.37 %. La mayoría de estas personas son parte de poblaciones clave que a continuación se presentan:

- **Trabajadores sexuales(varones):** Este grupo tiene una prevalencia media del 15 %, y es el grupo de riesgo más afectado en México.
- **Hombres que tienen sexo con hombres:** Tienen una prevalencia del 11 %.
- **Usuarios de drogas inyectables:** Dentro de esta población la prevalencia es del 5 %.
- **Trabajadoras sexuales:** Tienen tan solo el 2 % de prevalencia.

A pesar de tener estos grupos de riesgo bien identificados, se estima que el 59% de personas infectadas de SIDA en México desconocen que son portadores de dicho virus, generando un gran reto para el control de la epidemia.[13]

### 1.4.2. Atención a pacientes

Debido a la falta de detección de pacientes con VIH, tan solo el 27% de los infectados están recibiendo algún tipo de tratamiento. Por ello los mecanismos de detección juegan un papel sumamente importante en la lucha por controlar esta epidemia.

#### Diagnóstico

En México existen dos tipos de métodos diagnósticos usados por los organismos del sistema de salud pública.

- **Directos:** Estos métodos buscan detectar el virus dentro del organismo. Las pruebas que pertenecen a este grupo son: *detección de ácidos nucleicos, cultivo viral y antígenoemia*
- **Indirectos:** Detectan la presencia de anticuerpos. La prueba de *ELISA* y *las pruebas rápidas* son auxiliadas por la prueba confirmatoria de *Western Blot* para determinar un diagnóstico correcto.

Estas pruebas conforman una de las principales herramientas en la lucha por controlar la epidemia. Desafortunadamente existen varios factores adversos como la discriminación de personas con VIH, el retraso en la entrega de resultados y la poca accesibilidad a estas pruebas en muchos lugares del país, que generan una detección en fase terminal en más del 50% de los enfermos. Al no existir una detección oportuna, el desconocimiento en dichos pacientes puede provocar que pongan en riesgo a más personas a infectarse con el VIH, así como una poca efectividad en los tratamientos que se le administren en su etapa posterior a su detección.[14]

#### Tratamientos

Hasta 1996 la atención a los enfermos de SIDA se concentró en hospitales de especialidades, y hasta 1992 consistía el tratamiento en brindar apoyo psicológico y tratamiento de las infecciones oportunistas. Hasta 1991 se introdujo el tratamiento con antirretrovirales y en 1995 empezó el uso de unos inhibidores de proteasa llamados TARA (Tratamiento

Retroviral Altamente Supresivo), con esto el pronóstico de la enfermedad cambió radicalmente, convirtiendo el SIDA en una enfermedad crónica. Actualmente en México se han adoptado medidas para proporcionar tratamiento a personas infectadas sin seguridad social. Mediante estas medidas se ha podido ampliar la esperanza de vida a las personas infectadas y permitirles una vida similar a la que tenían antes de ser contagiados.[14]

### 1.4.3. Mortalidad

Antes de la llegada de los tratamientos antirretrovirales la tasa de mortalidad general por SIDA era de 1.8 muertes por cada 100,000 habitantes en 1990 y en 1996 alcanzó su máximo con 4.8 muertes. [13]

Desde 1988 se han contabilizado alrededor de 72 mil defunciones ocasionadas por el VIH/SIDA. En un análisis por sexo se ha observado que a lo largo del tiempo la muerte en hombres ha disminuido, mientras que la mortalidad en mujeres ha tenido un pequeño aumento, y siendo en ambos casos el grupo de 25-44 años el grupo más afectado.

### 1.4.4. Retos epidemiológicos

El principal reto epidemiológico en nuestro país consiste en evitar que se generalice a la población en general, también exige estrategias más focalizadas para la prevención.

También se necesita mejorar el acceso y la calidad de los servicios a personas que viven con VIH, en especial en los tratamientos ofrecidos a las mujeres, así como proponer un sistema que permita una detección oportuna del virus.

Al no existir sistemas de información que aproximen las variables de mortalidad por SIDA, se exige una mejoría en la atención integral de los pacientes para poder detectar dichas variables.

Es importante señalar que México se comprometió a detener la epidemia de VIH/SIDA para el año 2015, al firmar la Declaración del Milenio de las Naciones Unidas.

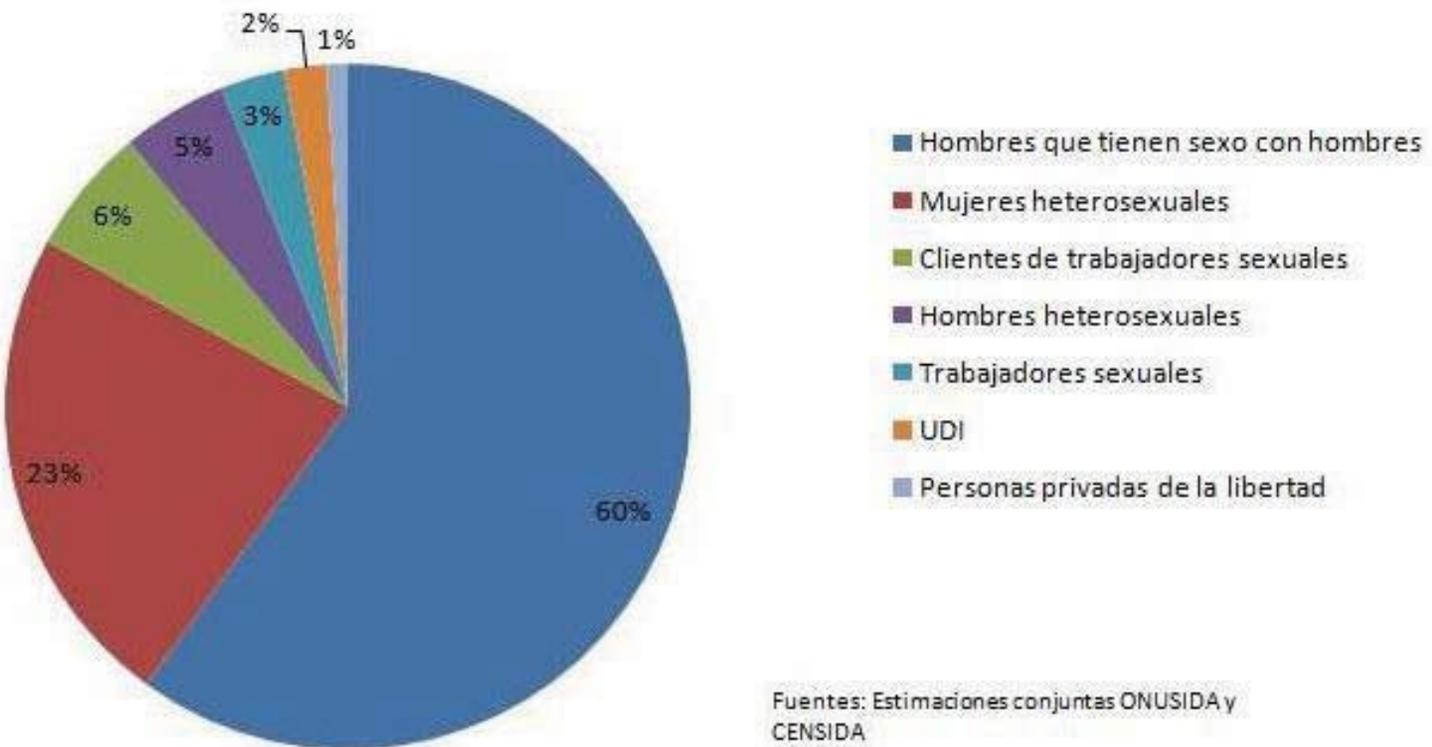


Figura 1.4: Gráfica de pastel que indica el porcentaje de infectados por grupo de riesgo en México



# Capítulo 2

## Estadística espacial

En este capítulo se concentra la definición y características, herramientas y aplicaciones principales de la estadística espacial. Como bibliografía principal para esta sección fueron usados los libros de Cressie *Spatial Statistics* publicado en 1993 [2] y el artículo que en mayor parte dió origen a la estadística espacial, publicado por Besag en 1972 llamado *Spatial interactions and the statistical analysis of lattice systems*[3].

### 2.1. Definición y motivación

La estadística, rama de las matemáticas que, mediante el análisis de los datos, intenta encontrar el orden dentro del desorden, busca a través de la estadística espacial encontrar patrones que permitan modelar información proveniente de diferentes regiones geográficas que interactúan entre sí.

La primera manifestación de estadística para datos espaciales, apareció en forma de mapa de datos. Se tienen registros de que Halley(1686) colocó en un mapa las direcciones de los vientos y monzones entre y cerca de los trópicos buscando asignarle una causa física.

Los modelos espaciales, que son una herramienta básica para el análisis espacial de los datos, aparecieron mucho después. Uno de los primeros que aplicó estos modelos fue Student (1907), que buscaba conocer la distribución de las partículas en un líquido. [2]

Conforme al paso del tiempo, los modelos espaciales se han implementado y desarrollado en problemas donde se busca encontrar la dependencia espacial de los datos. Existen una gran cantidad de fenómenos que proporcionan problemas con esta incógnita y que buscan

encontrar su respuesta en la estadística espacial. Algunos ejemplos son:

- *Fenómenos Físicos*: Predicción del clima, análisis geológicos para la extracción de recursos.
- *Fenómenos Epidemiológicos*: Conocer el comportamiento y transmisión de una enfermedad para su prevención, análisis de mapas de mortalidad.
- *Fenómenos Ecológicos*: Análisis de suelo para alimentación de ganado y mejoramiento de cosecha en agricultura, análisis de contaminación ambiental.
- *Análisis de imágenes*: Su mayor contribución está en auxiliar los avances de la investigación médica para la mejor comprensión del cuerpo humano, así como en el uso para la inteligencia militar y civil.

Este tipo de fenómenos en su mayoría pueden ser modelados y analizados por el modelo espacial general que es presentado en la siguiente sección.

## 2.2. Modelos espaciales

Una herramienta fundamental de la estadística espacial es el *modelo espacial general*, el cual busca relacionar de forma ordenada la región espacial con la variable o fenómeno aleatorio. El modelo espacial general en términos matemáticos se define como:

$$\{\mathbf{X}_s : s \in D\} \tag{2.1}$$

donde:

- $\mathbf{s} = (x, y, z)$  denota las coordenadas (o valor asignado) al sitio observado
- $X_s$  variable aleatoria o propiedad indexada en  $D$ . Esta es la variable de interés en el sitio  $\mathbf{s}$ , y representa de forma numérica el comportamiento del fenómeno.
- $D$  representa el conjunto de los sitios en los cuales los datos se obtuvieron y donde está bien definido el proceso  $\{\mathbf{X}_s : s \in D\}$

Los datos espaciales se pueden clasificar principalmente en 3 grupos, según el contexto de observación del que proviene y dependiendo la naturaleza de  $D$  y  $\mathbf{X}$ . Por tanto en estadística espacial se distinguen principalmente tres tipos de datos:

- **Geostadísticos:** Los datos se observan en  $s$  sitios de un conjunto no numerable  $D \in \mathbb{R}^d$ . El principal objetivo en la aplicación de estos modelos es generalmente la predicción del proceso en sitios no visitados.
- **Lattice (Enrejado):** La región de estudio es continua, pero el fenómeno solo está definido en un conjunto contable de sitios  $s_1, s_2, \dots, s_n$ . La región puede ser regular o irregular. Sus aplicaciones más importantes en este tipo de datos son de tipo epidemiológico.
- **Procesos puntuales:** Un proceso puntual es una colección de puntos o eventos que ocurren aleatoriamente en una región.

Otro tipo de datos a considerar son los espacio-temporales, los cuales son valores obtenidos de un mismo espacio geográfico en diferentes instantes.

## 2.3. Sistemas de lattice

Para los sistemas de Lattice existen 3 principales características que pueden presentarse de la siguiente forma:

- **Sistema de datos agregados:** Las regiones espaciales pueden ser divididas en 2 formas, regular e irregular. Se entiende como regular aquellas regiones divididas de forma uniforme, es decir todos los sitios son de la misma forma y su conexión con el resto de los sitios es idéntica (Ej: Mapa de Bits, división de rejilla.). Los irregulares son aquellos cuya división de la región es diferente en cada sitio, y la relación de vecinos también varía (Ej: División política de México.).
- **Sitios individuales :** Según la naturaleza del espacio los sitios pueden ser puntos o regiones. Una situación que puede ilustrar los sitios por puntos es el siguiente: Supongamos que deseamos encontrar un patrón espacial en las coordenadas de donde se originan los temblores. En este caso las coordenadas representan un punto en particular. Los problemas cuyos sitios individuales son regiones pueden ser ejemplificados si suponemos que deseamos reconstruir una imagen generada por un mapa de bits.
- **Variables aleatorias asociadas:** Esta característica corresponde al tipo de datos que se obtienen de la región espacial, y pueden ser discretos o continuos.

Es importante señalar que los datos utilizados para la elaboración de esta tesis, provienen de un sistema de regiones irregulares (división política de la República Mexicana) con una variable aleatoria asociada de tipo discreta (muertes por SIDA)

## 2.4. Vecino más cercano (Nearest Neighbor)

Existen dos enfoques para analizar a los procesos estocásticos espaciales en rejilla, debido a que hay dos definiciones no equivalentes para el sistema de *vecino más cercano*.

El enfoque más utilizado y que sustenta el teorema de *Hammersley-Clifford* es el de probabilidad condicional, propuesto por Bartlett (1955).

Si restringimos nuestra atención momentaneamente a los lattice rectangulares con sitios etiquetados por las parejas de enteros  $(i,j)$  con una variable aleatoria  $\{X_{i,j}\}$ , el enfoque propuesto por Bartlett requiere que la probabilidad condicional de  $\{X_{i,j}\}$  dado todos los valores en el resto del lattice, solo dependa de los valores en los cuatro sitios mas cercanos a  $(i,j)$ . La probabilidad de  $\mathbf{x}$  esta dada por:

$$P(\mathbf{x}) = \prod_{i=1}^n P(x_{i,j} | x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1}) \quad (2.2)$$

A pesar de parecer muy intuitivo este enfoque, acarrea numerosas desventajas tales como:

- No existe un método obvio para deducir la estructura de probabilidad conjunta asociada a este modelo.
- Esta sujeto a condiciones muy restrictivas de consistencia.

Pero si suponemos un modelo con una clase mas amplia de probabilidad condicional donde la distribución condicional de  $X_{i,j}$  pueda depender de mas sitios, las desventajas mencionadas anteriormente dejan de estar asociadas al modelo.[3]

## 2.5. Teorema de Hammersley-Clifford

La importancia en la modelación espacial de este teorema radica en que limita la complejidad de las distribuciones condicionales, permitiendo especificar las distribuciones condicionales de forma local. Antes de enunciar el teorema de Hammersley-Clifford es importante exponer algunas definiciones.

**Definición 1.** Sea  $j (\neq i)$  un sitio dentro de un esquema espacial y  $X_j$  la variable asociada al sitio,  $j$  es vecino de  $i$  si y solo si la forma funcional de  $P(x_i | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$  depende de la variable  $X_j$

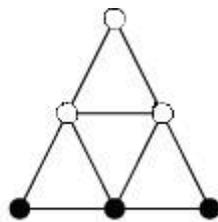


Figura 2.1: Ejemplo de Clique (marcado en color blanco) en un esquema de vecino más cercano

**Definición 2.** Si en un lattice rectangular la  $P(X_{ij}|\text{el resto de los sitios})$  depende de el y sus cuatro vecinos, es decir  $x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1}$  para cada sitio interno, entonces tenemos un *esquema de vecino mas cercano*

**Definición 3.** Un *clique* es el conjunto de sitios  $\mathcal{L}$  donde todos los elementos de  $\mathcal{L}$  son vecinos entre sí.

Para hacer mas clara la definición, mostraré un ejemplo. Supongamos que el espacio donde estamos trabajando es un lattice rectangular, y el sistema de vecinos es de primer orden, es decir, que los vecinos de un sitio unicamente serán los rectangulos a su derecha, su izquierda, arriba y abajo. En la figura (2.1) se muestra de color blanco un clique en particular del sistema expuesto en la misma figura.

Por último haremos dos suposiciones:

1. Supondremos que existe una cantidad finita de valores posibles en cada sitio.
2. **Condición de positividad.** El cero es un valor posible en todos los campos, y es equivalente a la propiedad global y local de Markov. Los propios autores del teorema pensaron que era una condición "no natural" y pospusieron la publicación en espera de relajar dicha condición.

Y bajo esto podemos definir:

$$Q(\mathbf{x}) = \ln \left\{ \frac{P(\mathbf{x})}{P(\mathbf{0})} \right\}. \quad (2.3)$$

Donde  $P(\mathbf{x})$  representa la probabilidad de una realización  $\mathbf{x}$  en particular, mientras que  $P(\mathbf{0})$  representa la probabilidad de una realización donde todos los valores de esta son cero.

El teorema de Hammersley-Clifford pretende encontrar la forma mas general en que  $Q(\mathbf{x})$  puede tomar para proporcionar una estructura de probabilidad válida al problema. La expresión :

$$\exp \{Q(\mathbf{x}) - Q(\mathbf{x}_i)\} = \frac{P(\mathbf{x})}{P(\mathbf{x}_i)} \quad (2.4)$$

para cualquier  $\mathbf{x} \in \Omega$  resulta de mucha utilidad para demostrar el teorema. Además para cualquier  $\mathbf{x} \in \Omega$ , escribimos  $\mathbf{x}_i$  para la realización:

$$(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n)$$

**Teorema 2.5.1.** (Hammersley-Clifford) *Para cada distribución de probabilidad  $P(\mathbf{x})$  sujeta a las restricciones 1. y 2. antes mencionadas , existe una expansión de  $Q(\mathbf{x})$ , única en  $\Omega$  y de la forma:*

$$Q(\mathbf{x}) = \sum_{1 \leq i \leq n} x_i G_i(x_i) + \sum_{1 \leq i < j \leq n} x_i x_j G_{ij}(x_i x_j) + \dots + x_1 x_2 \dots x_n G_{1,2,\dots,n}(x_1 x_2 \dots x_n) \quad (2.5)$$

Donde  $x_i G_i(x_i) \equiv Q(0, \dots, 0, x_i, 0, \dots, 0) - Q(0)$ .

Para cualquier  $1 \leq i < j < \dots < s \leq n$  función  $G_{i,j,\dots,s}$  en (2.5) debe ser no nula si y solo si los sitios  $i, j, \dots, s$  forman un cliques.

### Demostración

Tenemos que para cualquier  $\mathbf{x}$  en  $\Omega$ ,  $Q(\mathbf{x}) - Q(\mathbf{x}_i)$  solo debe depender de los valores vecinos de  $\mathbf{x}_i$  y de el mismo. Sin perdida de generalidad podemos considerar el sitio 1.

De la ecuación (2.5)obtenemos :

$$Q(\mathbf{x}) - Q(\mathbf{x}_1) = \left\{ \begin{array}{l} \sum_{1 \leq i \leq n} x_i G_i(x_i) + \sum_{1 \leq i} \sum_{< j \leq n} x_i x_j G_{i,j}(x_i, x_j) + \\ \sum_{1 \leq i} \sum_{< j} \sum_{< k \leq n} x_i x_j x_k G_{i,j}(x_i, x_j) + \dots \\ \dots + x_2 x_3 \dots x_n G_{1,2,3,\dots,n}(x_2 x_3 \dots x_n) - \\ \sum_{2 \leq i \leq n} G_i(x_i) - \sum \sum_{2 \leq j \leq i \leq n} x_i x_j G_{i,j}(x_i, x_j) - \dots \\ \dots - x_2 x_3 \dots x_n G_{2,3,\dots,n}(x_2 x_3 \dots x_n) \end{array} \right\} \quad (2.6)$$

Podemos simplificar la ecuación (2.6) usando el hecho que tanto  $Q(\mathbf{x})$  como  $Q(\mathbf{x}_1)$  poseen los mismos elementos salvo por aquellos que contienen a  $x_1$ . De esta forma podemos obtener la ecuación (2.7).

$$Q(\mathbf{x}) - Q(\mathbf{x}_1) = x_1 \left\{ \begin{array}{l} G_i(x_i) + \sum_{2 \leq j \leq n} x_j G_{1,j}(x_1, x_j) + \\ \sum_{2 \leq j} \sum_{< k \leq n} x_j x_k G_{1,j}(x_1, x_j) + \dots \\ \dots + x_2 x_3 \dots x_n G_{2,3,\dots,n}(x_2 x_3 \dots x_n) \end{array} \right\}$$

Ahora supongamos que el sitio  $l (\neq 1)$  no es un vecino de 1. Entonces  $Q(\mathbf{x}) - Q(\mathbf{x}_1)$  debe de ser independiente de  $x_l$  para todo  $\mathbf{x}$  en  $\Omega$ . Colocando  $x_i = 0$  para  $i \neq 1 \neq l$  inmediatamente vemos que  $G_{1l}(x_1, x_l) = 0$  en  $\Omega$ . Esto aplica para el resto de sitios que no son vecinos del sitio 1, y es fácil ver que el resto de las G-funciones que involucran a  $x_i$  y  $x_l$  son 0. Con esto podemos ver que las G-funciones son no nulas ( $G_{i,j,\dots,s}$ ) si los sitios  $i, j, \dots, s$  forman un clique.

Para completar la demostración usamos el hecho que  $Q(\mathbf{x}) - Q(x_i)$  depende solo de  $x_l$  si existe una G-función no nula que involucre a  $x_i$  y  $x_l$ . Esto lleva a que lo mismo es verdad de  $P(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$

## 2.6. Campos aleatorios de Markov

Los campos aleatorios son una colección de variables aleatorias indexadas por sitios. Los campos aleatorios de Markov son análogas a las cadenas de Markov pero en una dimensión

mayor. Es decir la variable aleatoria  $X_i$  no solo depende de  $X_{i-1}$  sino de todas las variables aleatorias que estén asociadas a los sitios que sean vecinos de la ubicación donde proviene  $X_i$ .

Procesos como los campos magnéticos o incluso una imagen pueden ser vistos como campos aleatorios de Markov. Uno de los problemas más famosos que usan campos aleatorios de Markov es conocido como "Modelo de Ising".

**Definición 4.** Sea  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  con una distribución de probabilidad dada. Sea  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , una gráfica donde  $\mathcal{V} = \{1, 2, \dots, n\}$  y  $\mathcal{E}$  esta definida tal que no existe unión entre el nodo  $i$  y  $j$  si solo si:

$$x_i \perp x_j | \mathbf{x}_{-ij}$$

es decir  $x_i$  es condicionalmente independiente de  $x_j$  dado  $\mathbf{x}_{-ij}$ . Entonces decimos que  $\mathbf{x}$  es un Campo Aleatorio de Markov con respecto a  $\mathcal{G}$ . [5]

Bajo esta definición, un campo aleatorio de Markov debe satisfacer las siguientes propiedades:

- $P(\mathbf{x}) > 0$
- $P(x_i | \mathbf{x}_{-i}) = P(x_i | \text{vecinos de } x_i)$

## 2.7. Lema de Brook

Este lema provee una técnica bastante útil para obtener las distribuciones conjuntas para  $x_i$  en ciertos modelos. Es claro que dado  $p(x_1, x_2, \dots, x_n)$  las llamadas condicionales completas  $p(x_i | x_j, j \neq i)$ ,  $i=1,2,\dots,n$  se pueden obtener y son únicas, el lema de Brook prueba que dadas las condicionales se puede obtener la conjunta y esta es única.

**Lema 2.7.1.** *La distribución conjunta dada las condicionales completas esta dada por:*

$$p(x_1, x_2, \dots, x_n) = \frac{p(x_1 | x_2, \dots, x_n)}{p(x_{10} | x_2, \dots, x_n)} \cdot \frac{p(x_2 | x_{10}, \dots, x_n)}{p(x_{20} | x_{10}, \dots, x_n)} \cdot \dots$$

$$\dots \cdot \frac{p(x_n | x_{10}, \dots, x_{n-1} \ 0)}{p(x_{n0} | x_{10}, \dots, x_{n-1} \ 0)} \cdot p(x_{10}, x_{20}, \dots, x_{n0}) \quad (2.7)$$

Donde  $\mathbf{x}_0 = (x_{10}, x_{20}, \dots, x_{n0})$  es un valor fijo en el soporte de  $p$ .

## 2.8. Automodelos o modelos condicionales autorregresivos

Los automodelos son una subclase muy útil de los campos de Markov y se definen como:

**Definición 5.** . Supongamos que tenemos  $n$  sitios etiquetados  $1, \dots, n$  con sus vecinos correspondientes, y usamos el Teorema de Hammersley-Clifford para proponer una clase válida de distribuciones de probabilidad para  $X_1, \dots, X_n$ . Llamamos **automodelos** a los esquemas donde  $Q(x)$  se define como:

$$Q(x) = \sum_{1 \leq i \leq n} x_i G_i(x_i) + \sum_{1 < i < j \leq n} \beta_{ij} x_i x_j \quad (2.8)$$

donde  $\beta_{ij} = 0$  excepto cuando los sitios  $i$  y  $j$  son vecinos.[2]

Es importante señalar los siguientes aspectos de los automodelos:

- La estructura de probabilidad de los automodelos depende solo de los cliques que contengan no mas de dos sitios.
- Las distribuciones de probabilidad condicional asociadas a los automodelos pertenecen a la familia exponencial.

$$\ln p_i(x_i; \dots) = A_i(\cdot)B(x_i) + C_i(x_i) + D_i(\cdot)$$

donde  $p(\cdot)$  denota la probabilidad condicional de  $X_i$  dado los otros sitios.

Por tanto los automodelos tienen una estructura de probabilidad condicional que satisface:

$$p_i(x_i; \dots)/p_i(0; \dots) = \exp \left[ \left\{ G_i(x_i) + \sum_{j=1}^n \beta_{i,j} x_j \right\} \right] \quad (2.9)$$

Los modelos pueden ser clasificados de acuerdo a la forma que tenga  $p_i(\cdot)$ . Dentro de estos esquemas espaciales se encuentran los llamados:

- auto-Normal
- auto-Logístico

- auto-Poisson
- auto-Binomial

Los automodelos con mayor relevancia para efectos de esta tesis es el modelo *auto-poisson*, y *auto-normal* y es por eso que a continuación se presentan las principales características de estos modelos.

### 2.8.1. Esquema auto-poisson

Supongamos que  $X_i$  tiene una distribución condicional Poisson con media  $\mu_i$  la cual depende de los valores de los sitios vecinos. Además  $\mu_i$  es de la forma:

$$\mu_i = \exp(\alpha_i + \sum_{i \neq j} \beta_{i,j} x_j x_i) \quad (2.10)$$

Como  $X_i$  es infinito debemos asegurarnos que  $Q(x)$  sea sumable sobre  $x$ . Esto exige que  $\beta_{i,j} \leq 0$  para toda  $i$  y  $j$ . Entonces tenemos que:

$$Q(\mathbf{x}) = \sum \alpha_i x_i - \ln(x!) + \sum \sum \beta_{i,j} x_i x_j \quad (2.11)$$

Es claro que  $\exp Q(\mathbf{x})$  es sumable cuando  $\beta_{i,j} = 0$  y por tanto también lo es cuando  $\beta_{i,j} \leq 0$ . Para mostrar la necesidad de esta condición, consideremos la distribución del par de variables  $(X_1, X_2)$  y el resto de los sitios es igual a cero. Entonces tenemos que la probabilidad de esta realización es:

$$\exp Q(x_1, x_2, 0, \dots, 0) = \exp \alpha_1 x_1 + \alpha_2 x_2 + \beta_{1,2} x_1 x_2 / (x_1! x_2!) \quad (2.12)$$

para enteros no-negativos  $x_1$  y  $x_2$ . Aquí requerimos que la suma de esta cantidad sobre  $x_1$  y  $x_2$  lleguen a converger, esto solo es posible si  $\beta_{1,2} \leq 0$ . En general se necesita que  $\beta_{i,j} \leq 0$  para toda  $i$  y  $j$ . Esta restricción es severa e implica una interacción "competitiva" entre las variables auto-Poisson.[3]

A pesar de que este esquema es sumamente restrictivo y solamente puede modelar dependencia negativa entre vecinos, existen modelos jerarquicos con observaciones Poisson que son posibles de capturar dependencia positiva entre las observaciones.[5]

### 2.8.2. Auto-normal

Este modelo es también conocido como Modelo Condicional Autoregresivo (CAR) bajo el caso normal y supone:

$$X_i|x_j, j \neq i \sim N \left( \sum_j b_{ij}x_j, \tau_i^2 \right) \quad i = 1, 2, \dots, n. \quad (2.13)$$

Estas condicionales son compatibles, entonces bajo el lema de Brook podemos obtener la distribución conjunta dada por:

$$p(x_1, x_2, \dots, x_n) \propto \exp \left\{ -\frac{1}{2} \mathbf{x}' D^{-1} (I - B) \mathbf{x} \right\} \quad (2.14)$$

donde  $B = \{b_{ij}\}$  y  $D$  es una matriz diagonal con  $D_{ii} = \tau_i^2$ . La expresión (2.14) sugiere una distribución conjunta normal multivariada para  $\mathbf{X}$  con media  $\mathbf{0}$  y matriz de varianza-covarianza

$$\Sigma_{\mathbf{x}} = (I - B)^{-1} D$$

. Tenemos asegurar que  $D^{-1}(I - B)$  sea simétrica. La condición resultante más simple es:

$$\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2} \quad (2.15)$$

Bajo la ecuación (2.15) es evidente que la matriz  $B$  no es simétrica. Si consideramos la matriz de proximidad  $W$  con entradas  $w_{ij} = 1$  si y solo si los sitios  $i$  y  $j$  son vecinos y  $0$  en otro caso, podemos suponer que  $b_{ij} = w_{ij}/w_{i+}$  y  $\tau_i^2 = \tau^2/w_{i+}$ . Con esto la ecuación (2.15) se cumple y (2.13) es igual a:

$$X_i|x_j \sim N \left( \sum_j w_{ij}x_j/w_{i+}, \tau^2/w_{i+} \right) \quad (2.16)$$

Y la ecuación (2.14) se convierte en:

$$p(x_1, x_2, \dots, x_n) \propto \exp \left\{ -\frac{1}{2\tau^2} \mathbf{x}' (D_w - W) \mathbf{x} \right\} \quad (2.17)$$

donde  $D_w$  es una matriz diagonal con  $(D_w)_{ii} = w_{i+}$ .

Con esta notación surge un segundo problema ya que  $(D_w - W)\mathbf{1} = \mathbf{0}$  y esto implica que  $\Sigma_x^{-1}$  es singular, de tal forma que  $\Sigma_x$  no existe, y la distribución definida en la ecuación (2.17) es impropia. Con la ayuda del álgebra podemos obtener:

$$p(x_1, x_2, \dots, x_n) \propto \exp \left\{ -\frac{1}{2\tau^2} \sum_{i \neq j} w_{ij} (x_i - x_j)^2 \right\} \quad (2.18)$$

## 2.9. Procesos espacio-temporales

Este tipo de procesos se definen de una forma muy similar a los procesos espaciales.

**Definición 6.** *Se dice que  $X_{(s,t)}$  es un proceso espacio-temporal si:*

- $\mathbf{s} = (x, y, z)$  denota las coordenadas (o valor asignado) al sitio observado.
- $\mathbf{t}$  es la variable que denota el estado en el tiempo en el cual fue observado el dato.
- $X_s$  esta es la variable de interes asociada al sitio  $\mathbf{s}$  en el tiempo  $\mathbf{t}$  y representa de forma numérica el comportamiento del fenómeno.

Estos procesos generalmente son complejos, dichos procesos son caracterizados por una gran cantidad de observaciones en el espacio y en el tiempo. La complejidad de estos procesos en situaciones del "mundo real" son usualmente intensificadas por la ausencia de ciertas suposiciones tales como normalidad, estacionariedad espacio-temporal, y linealidad. De esta manera una perspectiva conjunta para modelar procesos espacio-temporales, a pesar de que son fáciles de formular son difíciles de implementar.[11]

Es por esto que surge la necesidad de implementar una formulación jerárquica para poder modelar estas complicadas estructuras espaciales y temporales, descomponiendolas en modelos condicionales relativamente sencillos. En el siguiente capítulo se expone de forma general la teoría de los modelos jerárquicos bayesianos, con los cuales se pueden modelar de manera mas sencilla los procesos espacio-temporales.

# Capítulo 3

## Herramientas bayesianas

En este capítulo se expone el enfoque estadístico bayesiano para motivar la construcción de las herramientas bayesianas usadas en el modelo jerárquico propuesto en esta tesis, también presentadas en este capítulo. Para esta sección use como apoyo principal el libro de *Bayesian Choice* de Robert [10] así como el libro de *Simulation* de Ross[12].

### 3.1. Estadística Bayesiana

En la estadística bayesiana, a diferencia de la estadística clásica, se asume al parámetro desconocido de interés,  $\theta$ , como una variable aleatoria. Por esto el enfoque bayesiano está basado en la distribución posterior, que está dada por:

$$\pi(\theta|x) = \frac{l(\theta|x)\pi(\theta)}{\int l(\theta|x)\pi(\theta)d\theta} \quad (3.1)$$

Donde  $l(\theta|x)$  es la función de verosimilitud y  $\pi(\theta)$  es la función de distribución de  $\theta$  y  $X$  son las observaciones.

### 3.2. Modelos jerárquicos

Los modelos jerárquicos surgen al intentar modelar observaciones incorporando información apriori, denotada por  $(\pi(\theta))$  que se pueda descomponer.

**Definición 7.** *Un modelo bayesiano jerárquico es un modelo bayesiano estadístico  $(f(x|\theta), \pi(\theta))$  donde la distribución apriori  $\pi(\theta)$  se puede descomponer en distribuciones condicionales*

$$\pi_1(\theta|\theta_1), \pi_2(\theta_1|\theta_2), \dots, \pi_n(\theta_{n-1}|\theta_n)$$

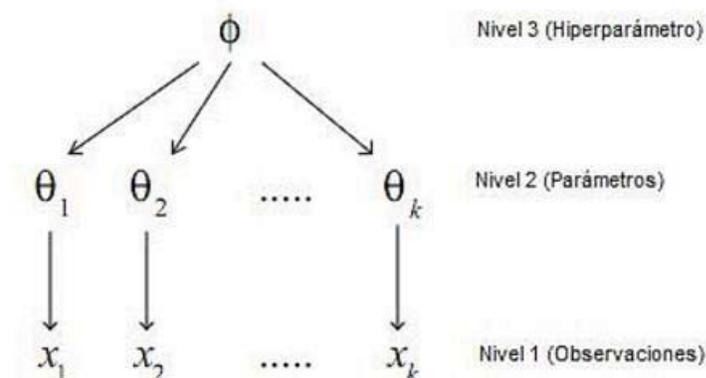


Figura 3.1: Representación gráfica de un modelo jerárquico de 3 niveles

y una distribución marginal  $\pi_{n+1}(\theta_n)$  tal que:

$$\pi(\theta) = \int_{\Theta_1 \times \dots \times \Theta_n} \pi_1(\theta|\theta_1), \pi_2(\theta_1|\theta_2), \dots, \pi_{n+1}(\theta_n) d\theta_1, \dots, d\theta_{n+1}$$

El parámetro  $\theta_i$  es llamado el hiperparámetro de nivel  $i$  ( $1 \leq i \leq n$ ) [10]

En la figura 3.1 se explica de manera gráfica la estructura de un modelo jerárquico de tres niveles. Generalmente los datos que son modelados de esta forma tienen una estructura en la que las observaciones son unidades agrupadas. Algunos ejemplos son:

- Datos de ciencias de la salud, donde es común que la información se recabe por estratos.
- Los estudiantes son agrupados en salones que a su vez son agrupados en escuelas, que a su vez son agrupados en distritos, y los datos que puedan generar los alumnos tienen la misma estructura jerárquica mencionada anteriormente.

En el caso de un modelo de 3 niveles, la estructura de este se podría explicar de la siguiente manera:

- Nivel 1 (Observaciones).- Las observaciones  $x_1, \dots, x_k$  provienen de estudios diferentes, pero en cierta manera están relacionados debido a la naturaleza del experimento.

$$\begin{aligned} p(x|\theta) &= p(x_1, \dots, x_k | \theta_1, \dots, \theta_k) \\ &= \prod_{i=1}^k p(x_i | \theta_i) \end{aligned} \quad (3.2)$$

- Nivel 2 (Parámetros).- Los parámetros  $\theta_1, \dots, \theta_k$  se suponen intercambiables, es decir asumimos que el subíndice  $i$  no proporciona información adicional. Los parámetros están asociados a las observaciones.

$$\begin{aligned} p(\theta|\phi) &= p(\theta_1, \dots, \theta_k|\phi) \\ &= \prod_{i=1}^k p(\theta_i|\phi) \end{aligned} \quad (3.3)$$

- Nivel 3 (Hiperparámetros).- El parámetro  $\phi$  describe alguna característica relevante de toda la población la cual se pretende estudiar.

$$p(\phi) \quad (3.4)$$

El principal problema al analizar los modelos jerárquicos es que la distribución inicial de los parámetros no está completamente especificada, sino que depende de hiperparámetros que a su vez tienen una distribución inicial propia. Por lo tanto el problema consiste en hacer inferencias sobre las características individuales de  $\theta_1, \dots, \theta_k$  y sobre los hiperparámetros.[15]

Con esto la distribución inicial para un modelo de tres niveles estaría dada por:

$$p(\theta\phi) = p(\theta|\phi)p(\phi) \quad (3.5)$$

### 3.2.1. Modelos jerárquicos espaciales

Estos modelos son el producto de distribuciones condicionales para datos condicionados en procesos espaciales y parámetros, el proceso espacial está condicionado mediante los parámetros que definen las dependencias entre sitios y los mismos parámetros.

Para construir estos modelos consideremos 3 variables aleatorias  $U$ ,  $V$  y  $W$ , cuya descomposición de la distribución conjunta de la tripleta  $(U, V, W)$  está dada por:

$$\pi(U, V, W) = \pi(W|U, V)\pi(V|U)\pi(U).$$

Partiendo de esta descomposición, supongamos que  $X$  es el proceso espacial de interés no observado, y si las observaciones  $Y$  son condicionales con respecto a  $X$ , podemos considerar el modelo jerárquico

$$\pi(Y, X, \theta_X, \theta_Y) = \pi(Y|X\theta_X, \theta_Y)\pi(X|\theta_X)\pi(\theta_X\theta_Y)$$

En esta descomposición hay tres niveles jerárquicos. En el nivel 1 se especifica como las observaciones  $Y$  son generadas mediante  $X$ , dando la distribución de  $Y$  condicionada por el proceso  $X$  y los parámetros del modelo  $(\theta_X, \theta_Y)$ . En el nivel 2 se explicita la distribución del proceso de interés  $X$ , condicionado a los parámetros  $(\theta_X)$ . Para el tercer nivel se expone la incertidumbre que tenemos de los parámetros  $(\theta_X, \theta_Y)$ . Esta metodología proporciona bastante libertad para modelar, permitiendo incorporar de manera simultánea incertidumbre y conocimiento apriori del fenómeno observado.[23]

Debido a que en muchas ocasiones la distribución final tiene dimensiones muy grandes y una estructura muy compleja, se recurre a herramientas computacionales tales como "Cadenas de Markov métodos de Monte Carlo"(MCMC) para que puedan ser analizadas estas distribuciones. En la siguiente sección se muestra una breve introducción a esta herramienta de simulación.

### 3.3. MCMC

La simulación es una técnica numérica en la que se pretende imitar un proceso del mundo real o de un sistema a través del tiempo, principalmente por medio de una computadora. Por su parte la simulación estocástica involucra procesos aleatorios para analizar el comportamiento de un problema, y es a este tipo de simulación al que pertenece MCMC. Este es una simulación basada en las técnicas de Monte Carlo usando cadenas de Markov.[12]

#### 3.3.1. Método de Monte Carlo

El método de Monte Carlo es usado para obtener una aproximación numérica de integrales cuyo valor no es inmediato. La idea básica es calcular una integral en términos del valor esperado de alguna función con respecto a alguna distribución de probabilidad.

#### 3.3.2. Cadenas de Markov

Una cadena de Markov es un proceso estocástico  $\{\theta^t : t \in T\}$  tal que, dado el estado actual, el pasado y el futuro del proceso independientes, es decir:

$$P(\theta^{(n+1)} \in E | \theta^{(n)} \in E_n, \theta^{(n-1)} \in E_{(n-1)}, \dots, \theta^0 \in E_0) = P(\theta^{(n+1)} \in E | \theta^{(n)} \in E_n)$$

Los métodos de Monte Carlo vía cadenas de Markov permiten generar de manera iterativa, observaciones de distribuciones que serían muy complicados simular de forma directa. El método consiste en construir una cadena de Markov sencilla de simular y cuya distribución de equilibrio corresponda a la distribución final de interés. Por tanto la cadena que se debe construir debe tener las siguientes propiedades:

- **Homogénea:** Esto indica que  $P(\theta^{(n+1)} \in E | \theta^{(n)} \in E_n)$  no depende de  $n$ .
- **Irreducible:** Con esto se refiere a que desde cualquier estado se puede acceder a otro estado.
- **Aperiódica:** Una cadena de Markov irreducible es aperiódica si para algún  $n \geq 0$  y algún estado  $j$ ,

$$P(X_n = j | X_0 = j) > 0 \quad y \quad P(X_{n+1} = j | X_0 = j) > 0$$

- **Reversible:** Cuando una cadena de Markov es aperiódica e irreducible se puede mostrar que:

$$\pi_j = \lim_{n \rightarrow \infty} P(X_n = j), \quad j = 1, 2, \dots, N$$

Se dice que una cadena es reversible si:  $\pi_i P_{ij} = \pi_j P_{ji}$  para toda  $i \neq j$ , donde  $P_{ij}$  es la probabilidad de pasar del estado  $i$  al  $j$ .

Si suponemos  $\theta^{(1)}, \theta^{(2)}, \dots$  una cadena de Markov con las propiedades antes mencionadas, con espacio de estados  $\Theta$  y distribución de equilibrio<sup>1</sup>  $p(\theta|x)$ . Entonces conforme  $t \rightarrow \infty$ :

- $\theta^t \xrightarrow{\mathcal{D}} \theta$ , donde  $\theta \sim p(\theta|x)$
- $\frac{1}{t} \sum_{i=1}^t g(\theta^{(i)}) \rightarrow \mathbb{E}(g(\theta)|x)$  c.s.

Con esto el reto que se plantea consiste en construir algoritmos que generen cadenas de Markov cuya distribución de equilibrio coincida con la distribución de interés. Dos algoritmos muy usados en la práctica son *Metropolis-Hasting* y *Gibbs Sampler*, el cual es un caso particular del primero.[16]

---

<sup>1</sup>La distribución de equilibrio se conoce como la distribución a la que converge la cadena de Markov cuando la cadena tiende a infinito

### 3.3.3. Metropolis-Hasting

Este algoritmo permite construir una cadena de Markov al definir las probabilidades de transición como se indican a continuación.

Sea  $Q(\theta^*|\theta)$  una distribución de transición arbitraria definida de la siguiente manera:

$$\alpha((\theta^*, \theta)) = \min \left\{ \frac{p(\theta^*|x)Q(\theta|\theta^*)}{p(\theta|x)Q(\theta^*|\theta)}, 1 \right\} \quad (3.6)$$

La idea es generar un valor de la distribución auxiliar y aceptarla con la probabilidad  $\alpha$ . Este mecanismo de corrección garantiza la convergencia de la cadena de distribución de equilibrio. Suponga que la cadena está en el estado  $\theta$  y se genera un valor  $\theta^*$  de una distribución propuesta  $Q(\cdot|\theta)$ . Un nuevo valor  $\theta^*$  es aceptado con la probabilidad  $\alpha$ . La cadena puede permanecer en un mismo estado a lo largo de muchas iteraciones. En la práctica, se acostumbra monitorear esto para calcular el porcentaje medio de iteraciones para los cuales los nuevos valores son aceptados.[16]

El algoritmo de *Metropolis-Hasting* se especifica de la siguiente forma:

- Definir un valor inicial  $\theta^{(0)}$  y fijar  $t=1$ .
- Generar una observación  $\theta^*$  de  $Q(\theta^*|\theta^{(t)})$ .
- Generar una variable aleatoria  $u \sim U(0, 1)$
- Si  $u \leq \alpha(\theta^*, \theta^{(t)})$ , hacer  $\theta^{t+1} = \theta^*$ ; en otro caso, hacer  $\theta^{(t+1)} = \theta^{(t)}$
- Hacer  $t=t+1$  y regresar al paso 2.

Este procedimiento genera una cadena de Markov con distribución de transición:

$$P(\theta^{(t+1)}|\theta^{(t)}) = \alpha(\theta^{(t+1)}, \theta^{(t)})Q(\theta^{(t+1)}|\theta^{(t)}). \quad (3.7)$$

### 3.3.4. Gibbs Sampler

Este algoritmo permite simular una cadena de Markov  $\theta^{(1)}, \theta^{(2)}, \dots$  con distribución de equilibrio  $p(\theta|x)$ . Cada nuevo valor de la cadena se obtiene al generar muestras de distribuciones cuya dimensión es menor que  $d$ , y que en la mayoría de los casos tiene una forma mas

sencilla que la de  $p(\theta|x)$ .

Sea  $\theta = \theta_1, \dots, \theta_n$  una partición del vector  $\theta$ , donde  $\theta_i \in \mathbb{R}^{d_i}$  y  $\sum_{i=1}^n d_i = d$ . En este caso  $\theta_i$  es un vector, pero en general cada componente  $\theta_i$  puede ser un escalar, un vector o una matriz. Las siguientes densidades para cada  $\theta_i$  son conocidas como densidades condicionales completas.

$$\begin{aligned} & p(\theta_1|\theta_2, \dots, \theta_n, x) \\ & \vdots \\ & p(\theta_i|\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n, x) \\ & \vdots \\ & p(\theta_n|\theta_1, \dots, \theta_{n-1}, x) \end{aligned}$$

Estas densidades pueden identificarse fácilmente al observar la forma de la distribución final  $p(\theta|x)$ . De hecho para cada  $i=1,2,\dots,n$ .

$$p(\theta_i|\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n, x) \propto p(\theta|x) \quad (3.8)$$

donde  $p(\theta|x)$  es vista como una función de  $\theta_i$ .

Este algoritmo funciona de la manera siguiente:

- Dar un valor  $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_n^{(0)})$ . Fijar  $t=1$ .
- Generar una observación  $\theta_1^{t+1}$  de  $p(\theta_1|\theta_2^{(t)}, \dots, \theta_n^{(t)}, x)$ .
- Generar una observación  $\theta_2^{t+1}$  de  $p(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_n^{(t)}, x)$ .
- $\vdots$
- Generar una observación  $\theta_n^{t+1}$  de  $p(\theta_n|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{n-1}^{(t+1)}, x)$ .
- Fijar  $t=t+1$  y regresar al paso 2.

La sucesión obtenida  $\theta^{(1)}, \theta^{(2)}, \dots$  es una realización de una cadena de Markov cuya distribución de transición esta dada por:

$$p(\theta^{(t+1)}|\theta^{(t)}) = \prod_{i=1}^n p\left(\theta_i^{(t+1)}|\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t+1)}, \dots, \theta_n^{(t+1)}\right)$$

[16]

# Capítulo 4

## Análisis exploratorio de los datos

### 4.1. Descripción de los datos

Los datos usados para este problema fueron obtenidos del Sistema de Mortalidad (SIS-MOR) del Instituto Mexicano del Seguro Social (IMSS) y Consejo Nacional de Población (CONAPO), dentro del periodo comprendido entre 1998 a 2007. Las variables usadas fueron:

- **Defunciones por SIDA (SIDA H) (SIDA M):** Es la variable a modelar, se usaron los datos de la población derechohabiente del IMSS fallecida a causa de SIDA, reportados cada 6 meses, y divididos según el sexo y el estado de la república en donde se reportó su muerte. La razón de usar exclusivamente las defunciones ocasionadas por el SIDA reportadas en el IMSS es debido a que no fue posible encontrar en el resto de las instituciones tanto públicas como privadas de salud, las defunciones que ocurrieron en dichos hospitales. Es importante mencionar que el IMSS atiende el 40 % del total personas con VIH y que se atienden en alguna institución pública de salud, por lo cuál podemos decir que este subconjunto es representativo para mostrar el comportamiento de las defunciones de SIDA en el país.[24]
- **Defunciones(Def):** Son el número de muertes registradas durante el periodo antes mencionado según la entidad federativa donde fue reportada la muerte con una periodicidad anual.
- **Crecimiento Poblacional Social (CPS):** Es la diferencia entre la migración y emigración reportada por estado de la república anualmente.
- **Población (Pob H) (Pob M):** Número de personas que residen en cada entidad federativa, divididos según el sexo y reportados anualmente.

- **Hospitales (Hosp)**: Número de clínicas del IMSS de nivel 1 (Unidades Médicas Familiares) que existen en cada estado de la república.
- **Nacimientos (Nac)** : Son el número de nacimientos registrados durante el periodo antes mencionado segun la entidad federativa donde fue reportado con una periodicidad anual

## 4.2. Análisis global de las variables

Con este análisis exploratorio se pretende encontrar mayor información de la interacción de las variables entre si. Con base en la matriz de correlación podemos identificar que las variables con mayor dependencia lineal son **Nac**, **Pob H**, **Pob M** y **Def**, en tanto que las variables con mayor correlación con las variables dependientes son **Def** y **Pob** de ambos sexos. Esto coincide con la dependencia evidente que existe con las defunciones, ya que al aumentar la variable SIDA aumenta la variable Def, y lo mismo ocurre con la Pob, ya que esta es la población en riesgo.

	SIDA H	SIDA M	Hosp	CPS	Def	Nac	Pob H	Pob M
SIDA H	1.00	0.86	0.61	-0.46	0.79	0.69	0.75	0.76
SIDA M	0.86	1.00	0.64	-0.39	0.77	0.70	0.74	0.74
Hosp	0.61	0.64	1.00	-0.36	0.71	0.66	0.69	0.68
CPS	-0.46	-0.39	-0.36	1.00	-0.54	-0.36	-0.38	-0.40
Def	0.79	0.77	0.71	-0.54	1.00	0.96	0.97	0.98
Nac	0.69	0.70	0.66	-0.36	0.96	1.00	0.98	0.98
Pob H	0.75	0.74	0.69	-0.38	0.97	0.98	1.00	0.99
Pob M	0.76	0.74	0.68	-0.40	0.98	0.98	0.99	1.00

Cuadro 4.1: Matriz de correlación de las variables involucradas en el estudio realizado en esta tesis.(La descripción de las variables se encuentran al inicio de este capítulo.)

### 4.2.1. Análisis de la matriz de correlación

Con este análisis pretendo encontrar relaciones no lineales entre variables y puntos atípicos, ya que son características que no se pueden identificar en el cuadro de correlación.

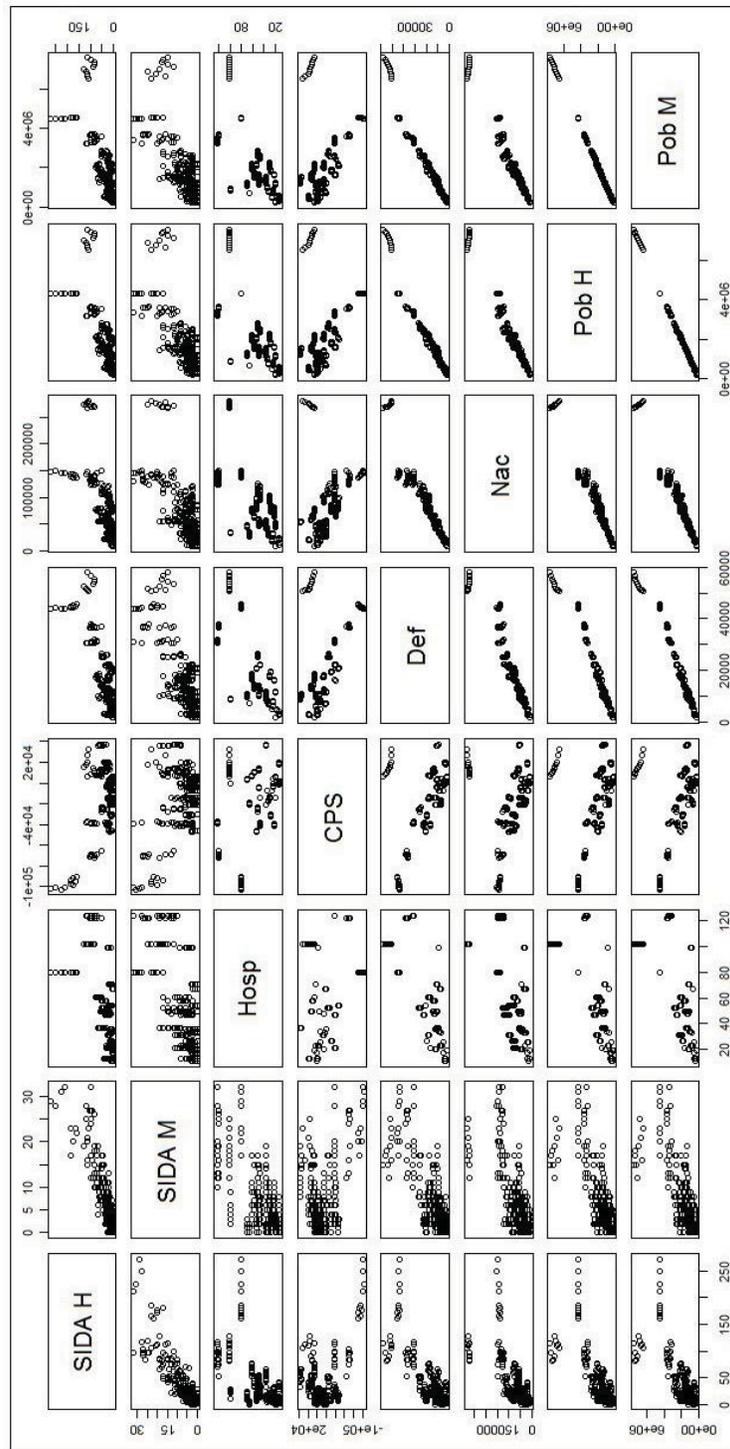


Figura 4.1: Correlograma de las variables que se usaron para realizar el trabajo en esta tesis(La descripción de las variables se encuentran al inicio de este capítulo.)

Al observar la gráfica correspondiente a **SIDA H** y **SIDA M** es bastante evidente que existe una agrupación de datos que se sale del comportamiento lineal creciente. Dichos datos pertenecen al Distrito Federal.

Otro dato bastante interesante es proporcionado por la gráfica correspondiente a **Def** y **Hosp** el cuál muestra como es de esperarse una tendencia lineal creciente muy evidente, exceptuando dos grupos de puntos, el primero Yucatán y el segundo es Edo de México. Ambos muestran un menor número de muertos por hospital que el resto.

También la gráfica de **CPS** y **SIDA H** muestra una evidencia muy clara (polinómica) que a mayor Sida en hombres hay un menor crecimiento poblacional.

En la gráfica que relaciona a **SIDA H** con **Def** es claro el comportamiento lógico, que a mayor defunciones hay más muertes de SIDA en hombres, pero existe un grupo de puntos correspondientes al D.F. que se comportan diferente, porque en este estado las **Def** se mantienen constantes pero **SIDA H** aumenta.

Existen seis gráficas muy parecidas en cuanto al comportamiento de sus puntos. Estas gráficas son:

- **SIDA H** con **Nac**
- **SIDA H** con **Pob H**
- **SIDA H** con **Pob M**
- **SIDA M** con **Nac**
- **SIDA M** con **Pob H**
- **SIDA M** con **Pob M**

Las seis presentan un crecimiento (exponencial) (las primeras tres gráficas lo muestran muy marcado) y también un grupo de puntos aislado del resto, estos son propiedad de Edo de México.

En el covariograma podemos ver que las siguientes cuatro gráficas tienen un comportamiento creciente igual que el mismo grupo atípico correspondiente a Edo de México.

- **Hosp** con **Def**

- Hosp con Pob H
- Hosp con Pob M
- Hosp con Nac

Por último existe un grupo de cuatro gráficas donde se muestra un comportamiento lineal decreciente exceptuando un conjunto de datos de Edo de México

- CPS con Def
- CPS con Pob H
- CPS con Pob M
- CPS con Nac

Después de realizar este análisis, resulta interesante observar con mayor detalle la relación que hay entre el número de muertes por SIDA y Hospitales en cada estado. A diferencia de lo mostrado en el covariograma (4.1) donde se cruzan de forma individual las variables SIDA H y SIDA M con hospitales cada año, en la figura (4.2), se observa de forma mucho más clara la relación que existe entre los Hospitales y las muertes por SIDA, haciendo evidente el hecho que a mayor número de hospitales hay mas registro de muertes por SIDA. Es importante recordar lo observado en la matriz de correlación (4.1) donde la correlación observada entre la población de hombres y mujeres rebasa en ambos casos el valor de 0.68, indicando con esto una fuerte correlación entre variables.

### 4.2.2. Dendograma

El dendograma presentado en la figura 4.3 fue obtenido mediante el método de agrupación completa, aunque realizando este gráfico por otros métodos (simple y promedio) el resultado obtenido para aglomerar las primeras tres categorías fueron idénticos. Esta figura nos refuerza la idea observada en el covariograma de considerar a México como un estado atípico, puesto que los tres principales grupos obtenidos serian:

1. DF, Jalisco, Veracruz
2. México
3. Los 28 estados restantes.

El estado de México es atípico debido a la gran cantidad de población que alberga, y por tanto todas las variables demográficas también son mucho mayores que el resto.

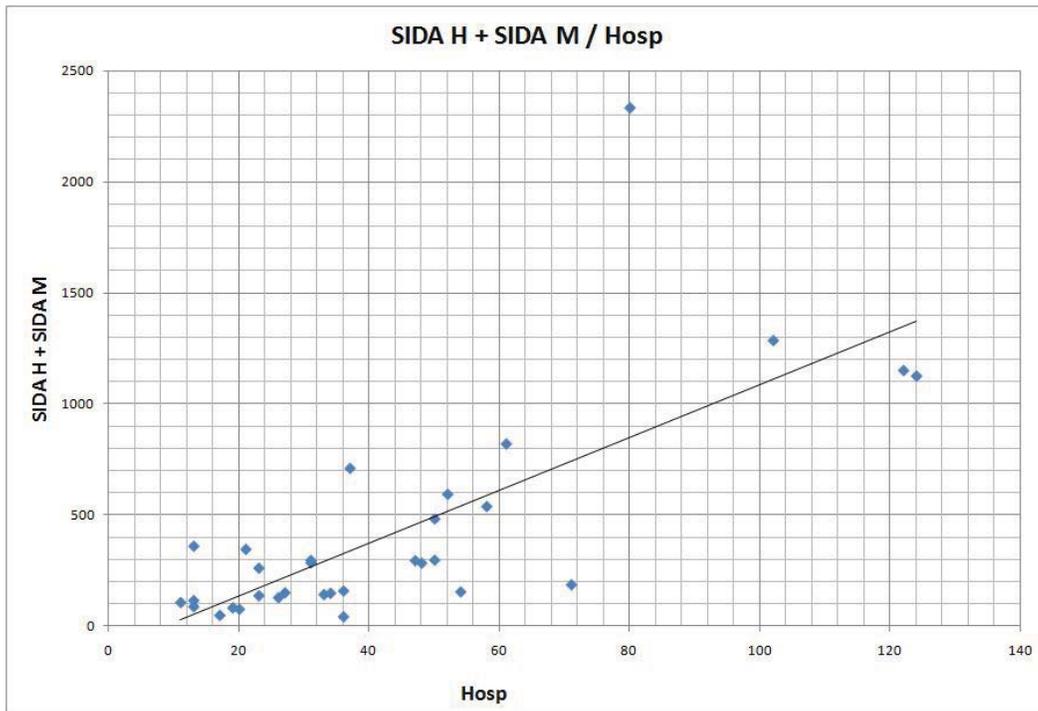


Figura 4.2: Gráfica que relaciona el número total de muertes por SIDA en el periodo de 1998 a 2007 entre el promedio de hospitales que existieron en los años antes mencionados.

## 4.3. Defunciones por SIDA

A continuación se presenta un análisis descriptivo de las muertes por SIDA en hombres y mujeres, obteniendo resultados concordantes con otros estudios realizados del SIDA.

### 4.3.1. Serie de tiempo

En la figura(4.4) se muestran dos series de tiempo respecto al número de muertes por año estandarizada por la población total, según el sexo. La gráfica presentada no muestra evidencia de tendencia alguna, debido a que el periodo graficada es muy corto, pero si podemos observar que el comportamiento del SIDA no es igual en hombres y mujeres, situación que concuerda con las conclusiones de otros estudios de la enfermedad [CONASIDA(2000)].

### 4.3.2. Proporción hombre / mujer

Es importante analizar la relación entre las muertes por género ya que la epidemia del SIDA se comporta de forma diferente. El porcentaje de hombres muertos por SIDA es del 84.1% del total mientras que las mujeres aportan el 15.9% del total de muertes. Pero como se puede observar en la figura 4.4 la proporción Hombre/Mujer esta disminuyendo, posiblemente debido a que el impacto de la epidemia en mujeres va en aumento, y por tanto las muertes de estas también. Por su parte en la figura 4.5 no se logra apreciar una relación evidente entre el total de muertes por estado y la proporción.

Además en la figura 4.9 podemos observar que la proporción de Hombre/Mujer muertos a causa del SIDA es mayor en los estados del sur que en los estados del norte del país.

### 4.3.3. Análisis por estado

Con este análisis se busca obtener una idea general sobre el comportamiento que tiene en cada estado las muertes por SIDA, principalmente la proporción Hombre/Mujer así como el porcentaje de muertes según el total de población en riesgo.

Las series de tiempo al graficarse por estado (mostradas en el apéndice), nos permite observar un comportamiento similar en cada estado al presentado en la figura 4.4.

En tanto que la figura 4.7 muestra que hay una mayor concentración de muertes en los estados que tienen costa en el Golfo de México, y en los estados en el norte del país. En

la figura 4.8 el mapa muestra un comportamiento muy similar ya que parece indicar que las muertes de mujeres tiene una mayor incidencia en los estados del Golfo de México, la península de Yucatán y de Baja California. Con esto los estados con mayor incidencia por muertes de SIDA en hombres son:

- **Quintana Roo:** Es el estado que ha sido más afectado por el SIDA. Esto puede ser consecuencia de la gran cantidad de turistas, en su mayoría jóvenes, que visitan el estado, en especial Cancún, el cual es el primer destino turístico del país (SECTUR; 2008) A este hecho se debe agregar la mala vigilancia que existe para identificar turistas portadores del virus en el país.
- **D.F. :** Con el análisis previo del covariograma y el dendrograma, es fácil observar que la capital del país es un estado atípico. Algunas posibles razones por las que la epidemia está afectándolo son su gran densidad de población, la gran cantidad de centros nocturnos y de prostitución, así como a la inmigración de enfermos a los hospitales del D.F..
- **Baja California:** En este estado, las mujeres son más afectadas en proporción a los hombres afectados. Una posibilidad de ser uno de los estados más afectados por el virus, yace en que es un estado fronterizo, y el riesgo podría radicar en la poca vigilancia para identificar a los extranjeros portadores del virus que cruzan la frontera.
- **Nuevo León:** Aquí puede haber diversos factores que afecten a este estado, entre ellos está el gran crecimiento industrial que ha tenido desde el 2000 (INEGI;2008), así como su importancia como estado fronterizo.
- **Baja California Sur:** Las posibles causas del impacto que ha tenido el virus en este estado, pueden deberse a la gran cantidad de sitios turísticos, entre ellos Los Cabos y Loreto (SECTUR; 2008), y posiblemente también debido a compartir frontera con Baja California.
- **Jalisco:** En este estado se encuentra uno de los destinos turísticos preferidos por homosexuales (Puerto Vallarta) más importantes a nivel mundial, esto puede generar una gran cantidad de infectados.
- **Yucatan:** En Yucatán el crecimiento poblacional social es negativo, por lo tanto en base a lo que puede observarse en el covariograma esto propicia una mayor cantidad de muertes por SIDA. Por último es importante agregar que Yucatán es el estado de la zona sur del país con más hospitales, y si consideramos que Quintana Roo tiene muy pocos hospitales y es un estado vecino, es posible que muchas muertes que deberían corresponder a Quintana Roo se registren en Yucatán.

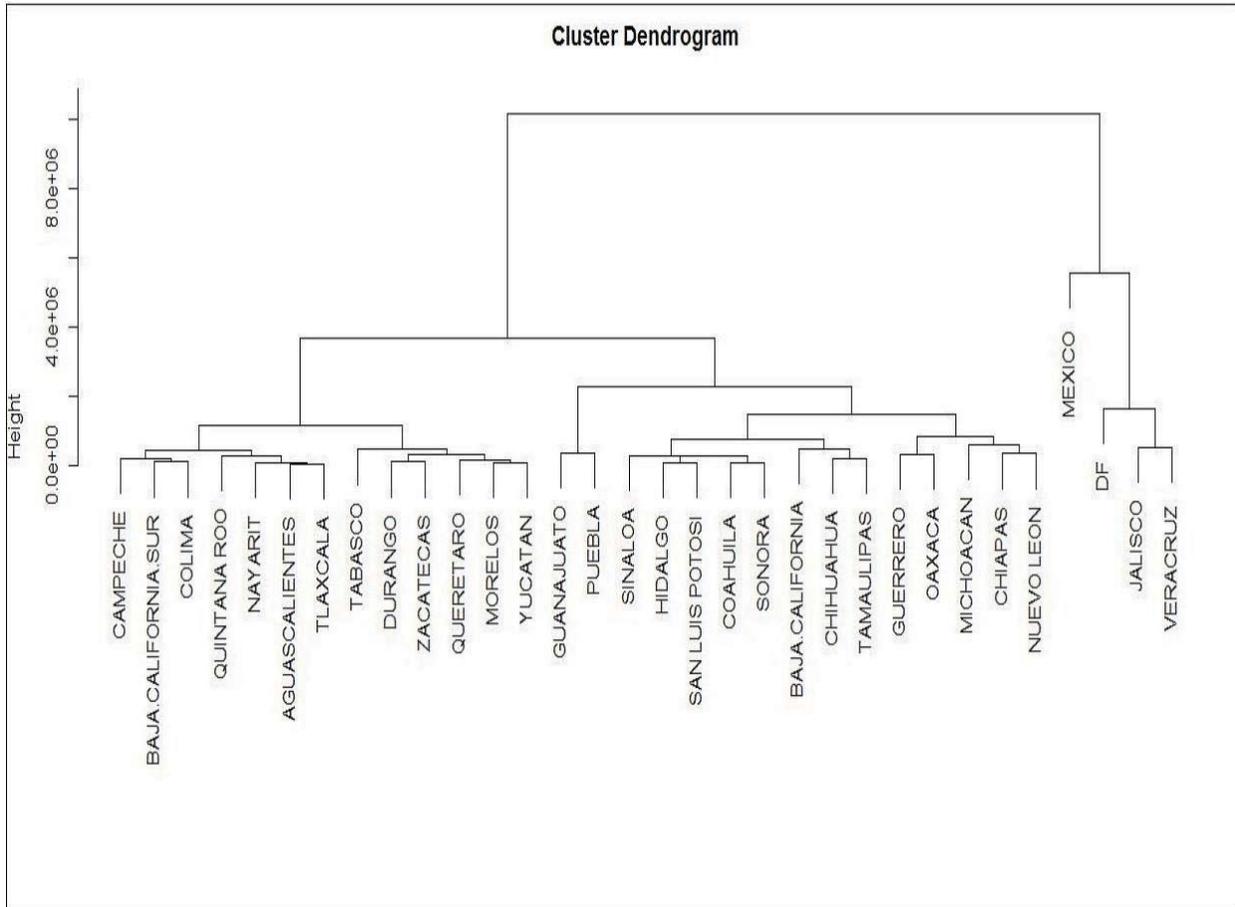


Figura 4.3: Dendrograma que agrupa los estados según las variables analizadas bajo el método de agrupación completa.

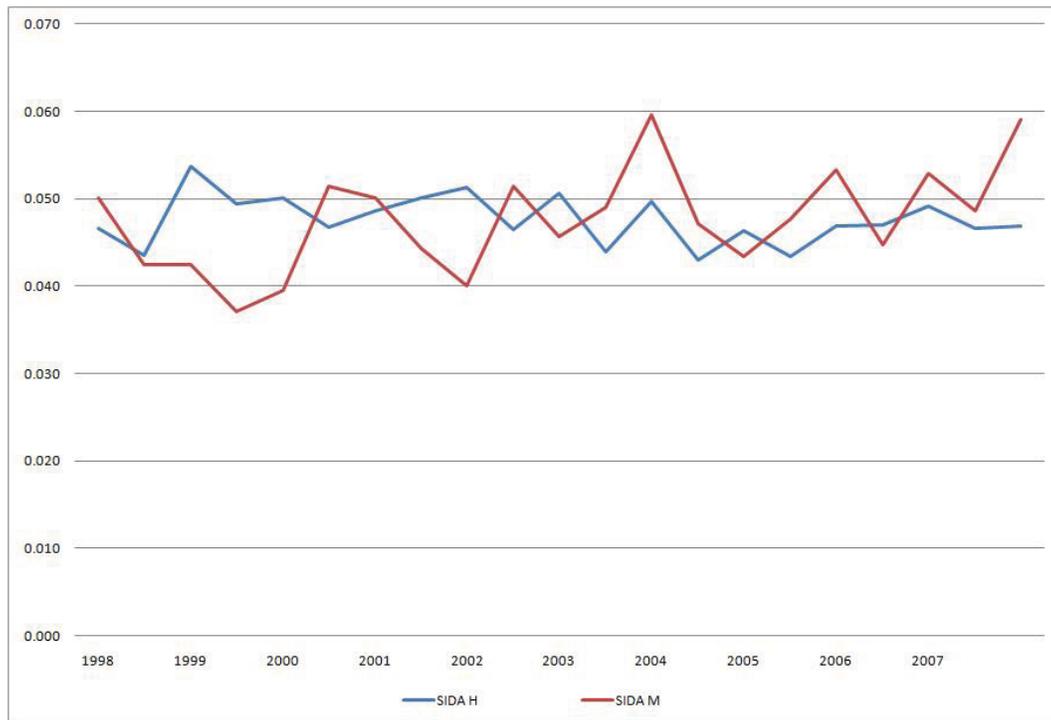


Figura 4.4: Serie de tiempo que indica las muertes anuales causadas por SIDA desde 1998 a 2007 de hombres y mujeres

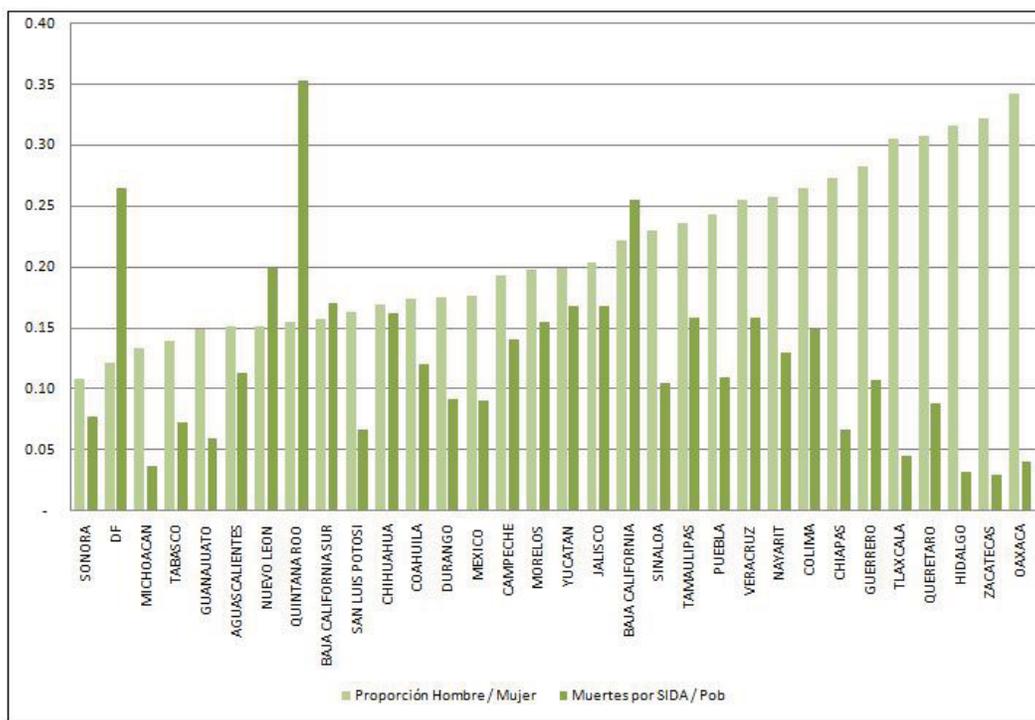


Figura 4.5: Proporción de hombres infectados con VIH, en relación a las mujeres, divididos por Estado, entre los años 1998 y 2007

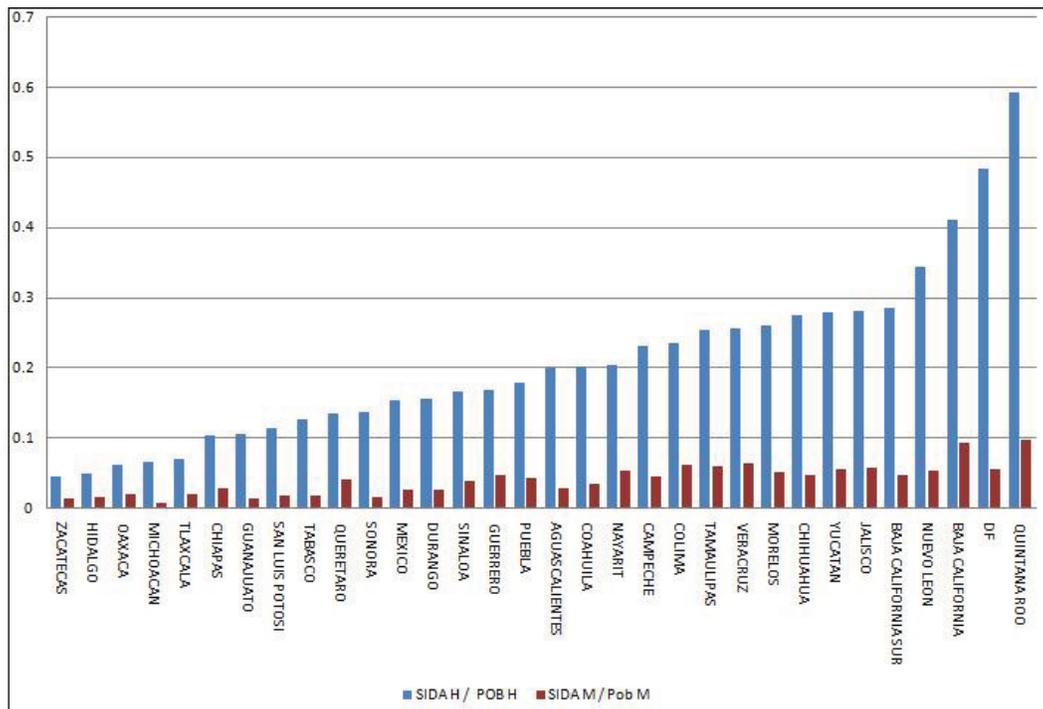


Figura 4.6: Proporción de muertes causadas por VIH, durante el periodo comprendido entre 1998 a 2007, en cada estado, según la población promedio en este tiempo, según el género.

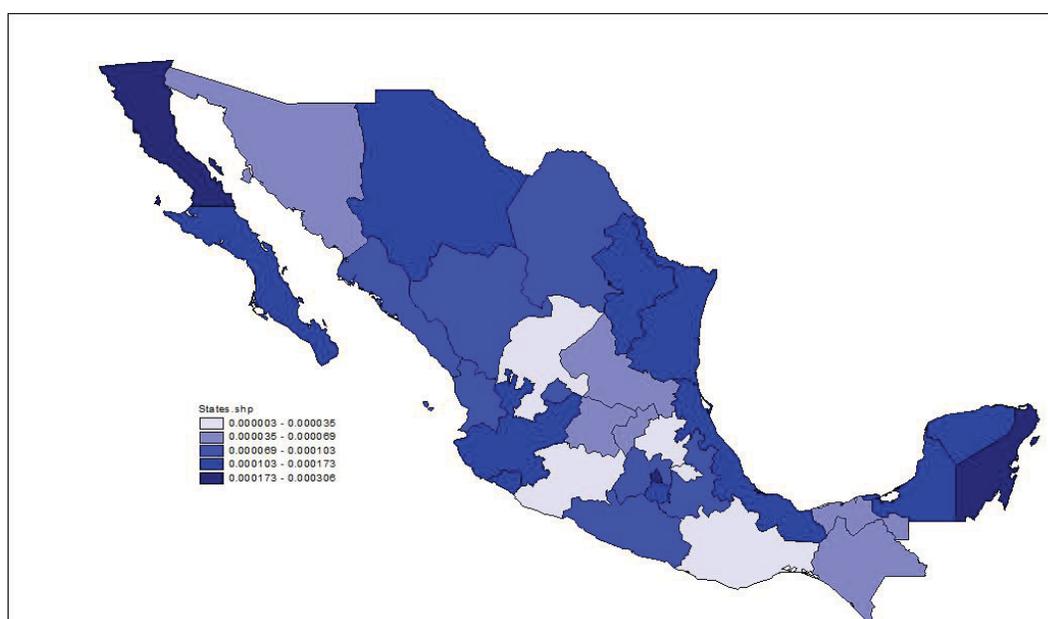


Figura 4.7: Mapa que muestra la proporción de muertes ocasionadas por VIH en hombres, divididas entre el promedio de población del mismo género que habita en cada estado , durante el periodo 1998 a 2007

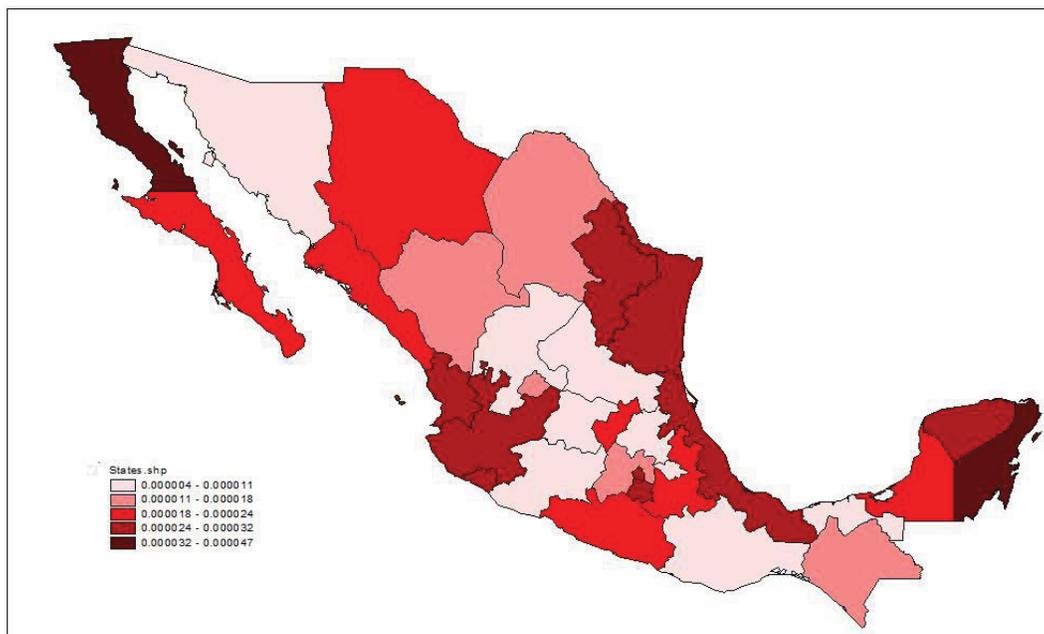


Figura 4.8: Mapa que muestra la proporción de muertes ocasionadas por VIH en hombres, divididas entre el promedio de población del mismo género que habita en cada estado, durante el periodo 1998 a 2007

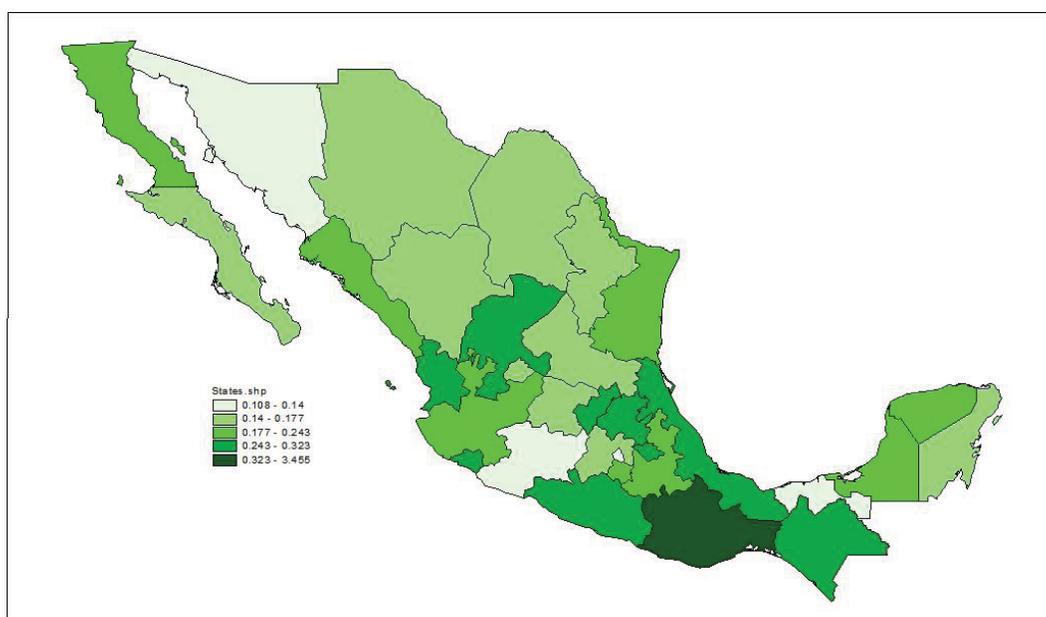


Figura 4.9: Este mapa muestra la proporción por estado de hombres entre mujeres que fallecieron a causa de SIDA durante el periodo comprendido entre 1998 a 2007



# Capítulo 5

## Análisis espacio temporal

### 5.1. Descripción del modelo

#### 5.1.1. Variables

En esta sección se expone la notación usada para definir el modelo, así como una descripción explícita de las variables en su forma matricial.

La variable respuesta que corresponde a los muertos de SIDA según género, año y estado, fueron definidas de la siguiente forma:

- $Y_{its}$  = Número de muertos por SIDA del sexo  $s$ , en el estado  $i$ , bajo el año  $t$ . Donde  $i = 1, 2, \dots, 32$  esta asociado de forma alfabetica a cada una de los estados de la Republica Mexicana,  $t = 1, 2, \dots, 10$  representa al lapso de tiempo 1998 al 2007 dividido de forma anual y  $s = 1, 2$  con 1 para hombres y 2 para mujeres. El vector que representa a  $Y_{ts}$  tiene la forma:

$$Y_{ts} = \begin{pmatrix} Y_{1ts} \\ Y_{2ts} \\ \vdots \\ Y_{32ts} \end{pmatrix}_{[32 \times 1]} \quad (5.1)$$

- $X_{irt}$  = Es la covariable  $r$  asociada al estado  $i$  en el tiempo  $t$ . Donde  $r = 1, 2, \dots, c$ . La matriz que representa a  $X_t$  es:

$$X_t = \begin{pmatrix} 1 & X_{1,1,t} & X_{1,2,t} & \cdots & X_{1,c,t} \\ 1 & X_{2,1,t} & X_{2,2,t} & \cdots & X_{2,c,t} \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & X_{32,1,t} & \cdots & & X_{32,c,t} \end{pmatrix}_{[32 \times c+1]} \quad (5.2)$$

## 5.2. Modelo jerárquico

Para modelar el comportamiento de muertes por SIDA, use un modelo jerárquico con 5 niveles.

- Nivel I (Observaciones)

$$Y_{its} \sim \text{Poisson}(\mu_{its}) \quad (5.3)$$

- Nivel II (Parámetros)

$$\log(\mu_{its}) = X_{it}\beta_s + \psi_{its} \quad (5.4)$$

- Nivel III (Hiperparámetros)

$$\forall \beta_{rs} \sim N(0, 10,000) \quad (5.5)$$

$$\psi_{its} = \mathbf{B}_s \psi_{ts} + \gamma_{1s} \psi_{i,t-1,s} + \gamma_{2s} \psi_{i,t+1,s} \quad (5.6)$$

- Nivel IV (Hiperparámetros)

$$B_s = \{b_{ij}\} \quad \forall b_{i,j} \sim N(0, \lambda_{i,j}^{-1}) \quad \gamma_{is} \sim N(0, \tau_i^{-1}) \quad i = 1, 2. \quad (5.7)$$

- Nivel V (Hiperparámetros)

$$\lambda_{i,j} \sim \Gamma(0.5, 0.0005) \quad \tau_i \sim \Gamma(0.5, 0.005) \quad (5.8)$$

En el Nivel I se asume que las observaciones son producto de una Poisson, ya que son valores positivos y discretos. El vector descrito como  $Y_s$  tiene la forma:

$$Y_s = \begin{pmatrix} Y_{1s} \\ Y_{2s} \\ \vdots \\ Y_{10s} \end{pmatrix}_{[(32*10) \times 1]} \quad (5.9)$$

mientras que para  $\mu_s$  el vector esta conformado como se muestra inmediatamente.

$$\mu_s = \begin{pmatrix} \mu_{1s} \\ \mu_{2s} \\ \vdots \\ \mu_{10s} \end{pmatrix}_{[(32*10) \times 1]} \quad (5.10)$$

Para el nivel II se busco modelar la posible relación funcional entre las  $\mu_{its}$  y la matriz de covariables  $X_t$  mediante una función lineal de parametros  $\underline{\beta}$  más la interacción espacial del estado  $i$  denotada por  $\psi$ . Partiendo de la probabilidad condicional:

$$p(Y_{its} | \mu_{its}) \sim \text{Poisson}(\mu_{its}) \quad (5.11)$$

el modelo lineal usado es:

$$\mu_{its} = \exp(\beta_{0s} + \beta_{1s}X_{i1t} + \beta_{2s}X_{i2t} + \dots + \beta_{cs}X_{ict} + \psi_{its})$$

que equivalentemente es:

$$\log(\mu_{its}) = \beta_{0s} + \beta_{1s}X_{i1t} + \beta_{2s}X_{i2t} + \dots + \beta_{cs}X_{ict} + \psi_{its} \quad (5.12)$$

y bajo su forma matricial, la representación es como se indica en la ecuación (5.4). La forma de  $\beta_s$  es:

$$\beta_s = \begin{pmatrix} \beta_{0s} \\ \beta_{1s} \\ \vdots \\ \beta_{cs} \end{pmatrix}_{[(c+1) \times 1]} \quad (5.13)$$

Para evitar redundancia en la matriz  $X_t$  se eliminaron las covariables con un alto grado de correlación dejando únicamente dentro del modelo las variables:

- Hosp
- CPS
- Pob H

En el nivel III se especifica la distribución de las  $\beta's$ , asumiendo independencia entre ellas. Además se expone la parte espacial del modelo bajo la distribución de  $\psi_{its}$ . Para especificar la forma matricial de la ecuación (5.6) es conveniente describir las siguientes matrices y vectores.

$$\psi_s = \begin{pmatrix} \psi_{1s} \\ \psi_{2s} \\ \vdots \\ \psi_{10s} \end{pmatrix}_{[(10 \times 32) \times 1]} \quad (5.14)$$

donde los elementos de la matriz  $\psi_s$  son denotados por  $\psi_{ts}$  que es un vector de la forma:

$$\psi_{ts} = \begin{pmatrix} \psi_{1ts} \\ \psi_{2ts} \\ \vdots \\ \psi_{32ts} \end{pmatrix}_{[(32) \times 1]} \quad (5.15)$$

los cuales representan los residuales del nivel II asociados al estado  $i$  en el tiempo  $t$  del sexo  $s$ .

$$B_s = \begin{pmatrix} b_{1,1,s} & b_{1,2,s} & \cdots & b_{1,32s} \\ b_{2,1,s} & b_{2,2,s} & \cdots & b_{2,32s} \\ \vdots & & \ddots & \vdots \\ b_{32,1,s} & b_{32,2,s} & \cdots & b_{32,32,s} \end{pmatrix}_{[32 \times 32]} \quad (5.16)$$

Debido a que la matriz  $B_s$  denota los parámetros de interacción espacial entre los estados, existen ciertas reglas que se deben considerar:

1.  $b_{i,i,s} = 1$
2.  $b_{i,j,s} = b_{j,i,s}$
3.  $b_{i,j,s} = 0$ . A menos que  $j$  sea vecino de  $i$ .

Esto se puede deducir partiendo que en un lattice de vecino mas cercano, la variable asociada al sitio  $i$ , solo depende de las variables de sus vecinos.  $\{\delta_i\}$ . Además bajo este modelo se asume que la relación de los estados en el tiempo no cambia.

Por último la parte temporal esta asociada a las  $\gamma_k$ , donde  $k=1,2$ .  $\gamma_1$  es el parámetro que asocia la relación temporal del estado  $i$  en el tiempo  $t-1$ , mientras que  $\gamma_2$  es para  $t+1$ .

Una vez definidos los elementos que interactuan en el nivel III es importante señalar que el proceso que se asume en el modelo tanto para las  $\psi$ 's como para las  $\gamma$ 's es un modelo condicional autorregresivo normal descrito en la sección (2.8.2).

En el cuarto nivel, se definen las distribuciones asociadas a los parámetros que interactuan en la parte espacial del modelo, mas la varianza asociada a las  $\beta_{is}$ .

En el último nivel se especifican las distribuciones de las varianzas de los parámetros espaciales. La varianza fué especificada grande debido al estado de ignorancia que se tiene acerca de la interacción de las variables.

Una vez definido el modelo, para llegar a elaborar los mapas de riesgo es necesario obtener la distribución de los parámetros. Y es a través del método de Gibbs sampler como la distribución de los parámetros puede ser encontrada.

### 5.3. Estimación de parámetros

Para poder facilitar la obtención de las distribuciones de los parámetros desconocidos por esta via, fue usado el paquete WinBUGS.

Dentro de la simulación el primer paso fue especificar los valores iniciales que tomarian los parámetros. Estos fueron determinados de la siguiente manera:

- $\beta_r = 0 \quad \forall r = 1, 2, 3, 4.$
- $\mu_{it} = 0 \quad \forall i = 1, 2, \dots, 32, \quad t = 1, 2, \dots, 10.$

La razón de asignar estos valores radica en el hecho que problemas similares expuestos como ejemplos en modelos jerárquicos espacio-temporales tales como "Scottish Lip Cancer" asignan a estos valores para evitar problemas al muestrear la distribución de los parámetros.[1]

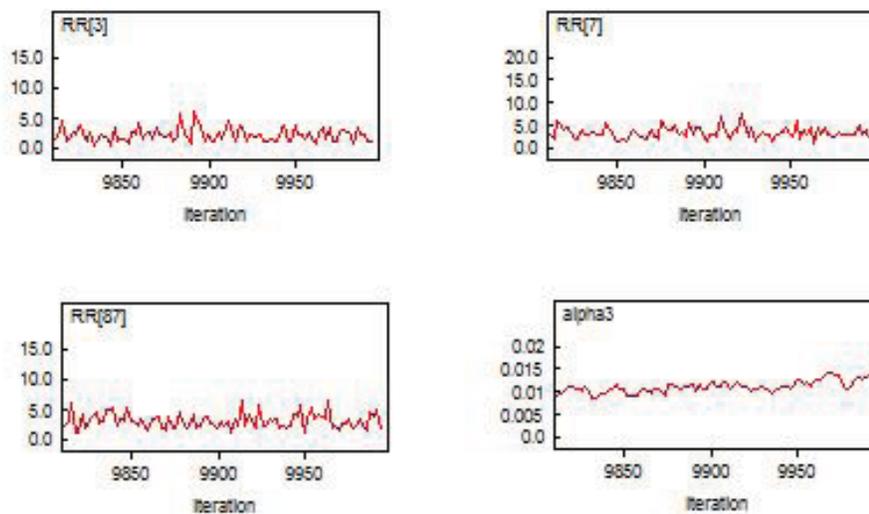


Figura 5.1: Cadenas generadas por la simulación correspondientes a la  $\mu$  de Baja California Sur en 1998, a la  $\mu$  de Coahuila en el mismo año, la  $\mu$  de Quintana Roo en el 2000, y a  $\beta_3$ .

Una vez definido los valores iniciales se realizó la simulación con 10,000 iteraciones. Ya simuladas se checó que las cadenas generadas hayan llegado a un punto donde la cadena sea estable. En la figura (5.1) se muestran a formade ejemplo 4 cadenas generadas en sus últimas 200 iteraciones. El resto de las variables tienen un comportamiento similar.

Una vez observando que las cadenas están en una fase estacionaria y que el error de Monte Carlo es menor al 5% de la desviación estándar de la muestra es útil observar las distribuciones observadas.

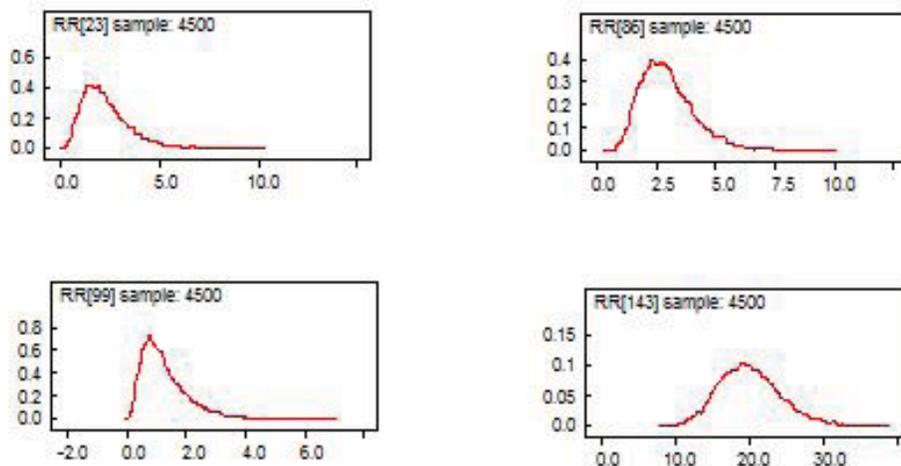


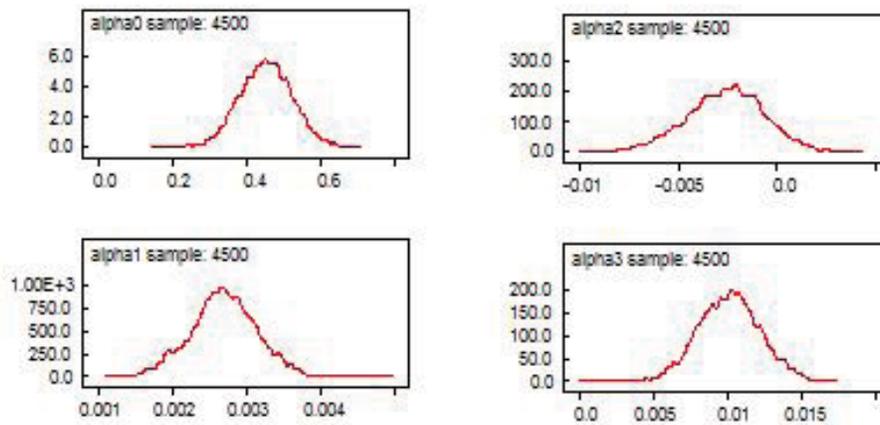
Figura 5.2: Distribución de las  $\mu$ 's correspondientes a Quintana Roo en 1998, Querétaro en el 2000, Baja California Sur en 2001 y el Estado de México en 2002

## 5.4. Resultados

Para obtener las distribuciones se desecharon las primeras 1000 iteraciones y se tomaron los valores de las iteraciones pares. La figura (5.2) muestra ejemplos de 4 distribuciones de  $\mu$ 's en los estados en diferentes años. Las distribuciones resultantes en la mayoría de los parámetros asociados a las  $\mu$ 's de los estados tienen una similitud bastante grande con la distribución Gamma, situación que es lógica tratándose de eventos que se distribuyen Poisson.

Por su parte la distribución de las  $\beta$ 's son presentadas en la figura(5.3). Con la ayuda de las estadísticas presentadas por WinBUGS expuestas en la tabla(??) podemos decir que las hipótesis  $\beta_0 = 0$ ,  $\beta_1 = 0$  y  $\beta_3 = 0$  son rechazadas, mientras que  $\beta_2 = 0$  no se rechaza, permitiéndonos asumir que la variable **CPS** no tiene un impacto lineal significativo en la variable **SIDA H**.

	2.5 %	97.5 %
$\beta_0$	0.31	0.59
$\beta_1$	0.001	0.003
$\beta_2$	-0.006	0.001
$\beta_3$	.001	.014

Cuadro 5.1: Valores de los cuantiles 2.5 % y 97.5 % de las distribuciones de las  $\beta$ 'sFigura 5.3: Distribuciones resultantes de las  $\beta$ 's

## 5.5. Mapas de riesgo

La presentación de los mapas de riesgo implica la decisión de un estimador puntual bayesiano que sea consistente con el problema. El estimador más natural para presentar los resultados es el M.A.P. (maximum a posteriori) presentado por York, Besag y Mollie en [4]. Por tanto los mapas presentados a continuación representan la moda de la distribución de las  $\mu$ 's divididas entre **Pob H** conocido como *riesgo relativo*. La escala de azul de claro a oscuro denotan los siguientes intervalos:

- $(0, 1 \times 10^{-6})$
- $(1 \times 10^{-6}, 2 \times 10^{-6})$
- $(2 \times 10^{-6}, 3 \times 10^{-6})$
- $(3 \times 10^{-6}, 4 \times 10^{-6})$
- $(4 \times 10^{-6}, 5 \times 10^{-6})$

Para ver la evolución de las muertes a lo largo del tiempo en los estados, se presentan en color verde los mapas de diferencia del riesgo relativo, denotados por:

$$RR_{i,t} - RR_{i,t-1}$$

,

donde  $RR_{i,t}$  representa el riesgo relativo asociado al estado  $i$  en el año  $t$ . La escala de colores fue usada para denotar de color oscuro aquellos estados que tuvieron un aumento representativo de riesgo relativo con respecto al año anterior, de color claro aquellos cuyo riesgo relativo disminuyó de un año a otro y de color intermedio los estados que se mantuvieron constantes. La escala de claro a oscuro comprende los siguientes intervalos:

- $(-3 \times 10^{-6}, -1 \times 10^{-6})$
- $(-1 \times 10^{-6}, 1 \times 10^{-6})$
- $(1 \times 10^{-6}, 3 \times 10^{-6})$

## 5.6. Series de tiempo

En esta sección se presentan las series de tiempo de los estados con mayor riesgo relativo a lo largo de los 10 años analizados. En las gráficas se presenta el riesgo relativo multiplicado por 1,000,000, esto con el objetivo de tener una escala que permitiera ver valores fácilmente identificables. Con ayuda de las gráficas presentadas se puede sugerir un análisis más cuidadoso en su evolución con respecto al tiempo para los siguientes estados:

- Baja California: La gráfica sugiere un aumento en el riesgo relativo al avance de los años.
- D.F.: La serie de tiempo nos sugiere una fuerte evidencia de que el riesgo relativo está decreciendo de forma significativa.
- Veracruz: Este estado parece presentar un crecimiento constante en el riesgo relativo.
- Quintana Roo: La tendencia del riesgo relativo que parece presentar el estado es creciente.

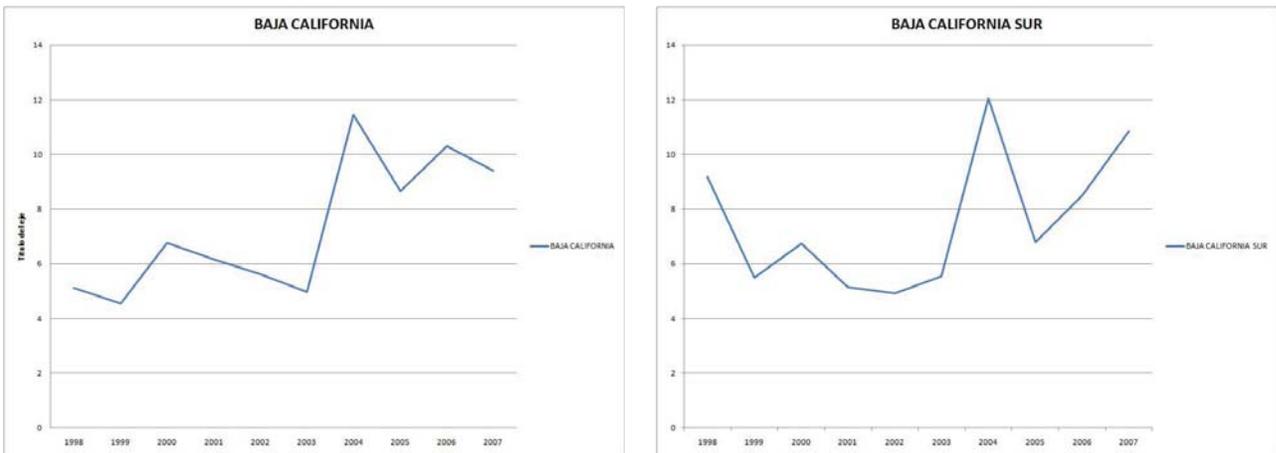


Figura 5.4: Series de tiempo que corresponden a las modas del riesgo relativo en Baja California y Baja California Sur.

Para poder afirmar de forma contundente las sospechas que generaron las series de tiempo es indispensable usar las distribuciones finales del riesgo relativo en dichos estados.

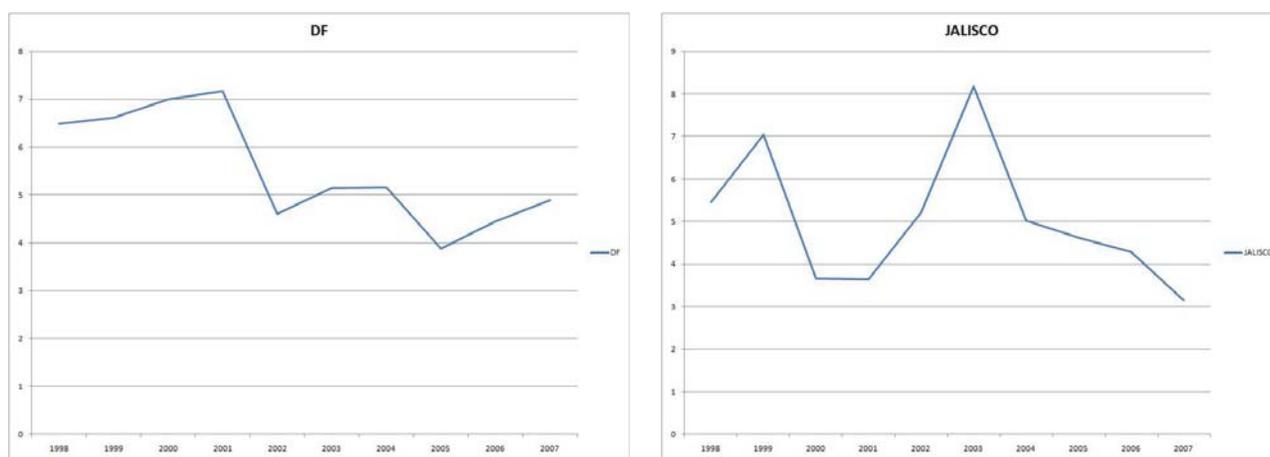


Figura 5.5: Series de tiempo que corresponden a las modas del riesgo relativo en el DF y Jalisco.

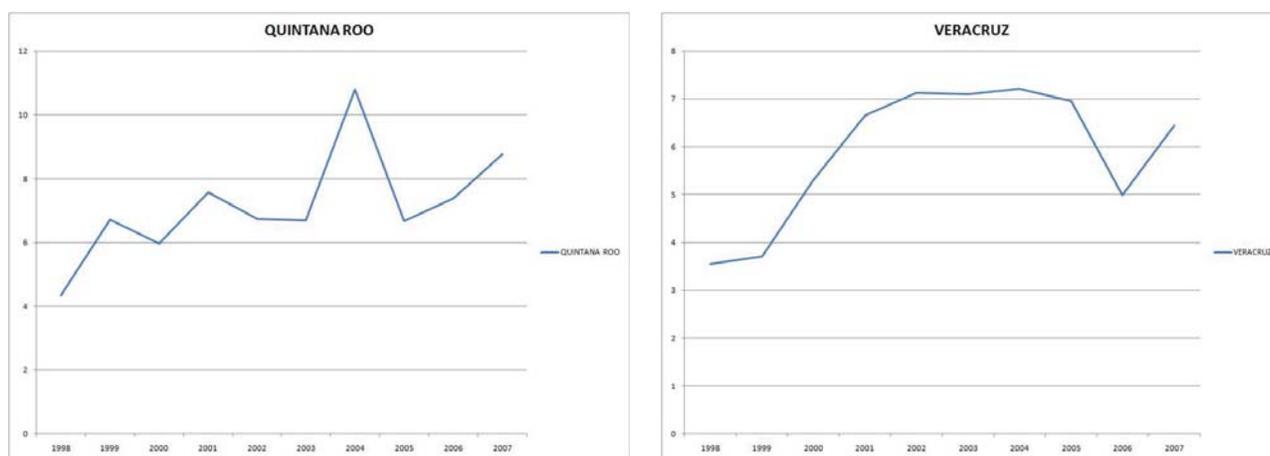


Figura 5.6: Series de tiempo que corresponden a las modas del riesgo relativo en Quintana Roo y Veracruz.

## 5.7. Observaciones y comparaciones

En los mapas de riesgo se pueden observar cuatro zonas con un particular comportamiento. A continuación se exponen dichos comportamientos regionales.

- La región que atrae mayormente la atención es aquella asociada a la Península de Yucatán. Los estados involucrados son Campeche, Quintana Roo y Yucatán. A lo largo

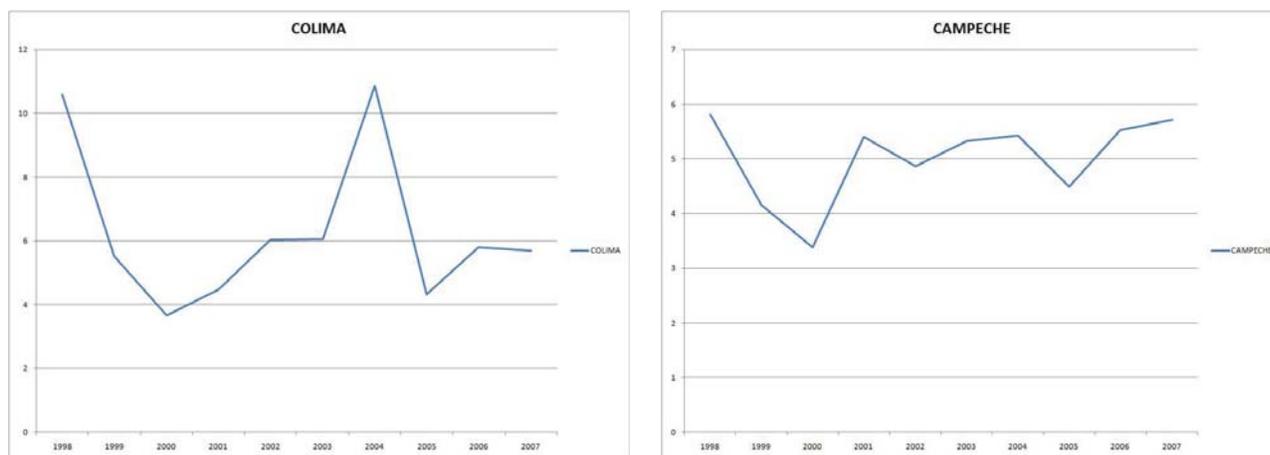


Figura 5.7: Series de tiempo que corresponden a las modas del riesgo relativo en Colima y Campeche.

del tiempo se puede observar que en esta zona, el riesgo no es idéntico para los tres estados, y a pesar de que Quintana Roo es el estado con mayor riesgo de esta zona, en los años de 1998 y 2002 los estados vecinos poseen un mayor riesgo.

- La segunda región con mayor riesgo es la península de Baja California, los dos estados pertenecientes a esta localidad presentan un comportamiento sumamente similar, exceptuando un riesgo relativo poco mayor por parte de Baja California.
- La región comprendida por Jalisco, Colima y Nayarit presentan un comportamiento muy similar a la Península de Yucatán salvo que el riesgo relativo parece ser un poco menor.
- Por su parte Veracruz es un estado que presenta riesgo relativo relativamente mayor al resto de los estados vecinos, siendo el estado más riesgoso en la zona del golfo.
- Por su parte el D.F. es un estado cuyo comportamiento aparentemente no es relevante.
- Una característica importante que se puede observar es que en el último año el norte del país presenta un comportamiento atípico, ya que en el resto de los años no presenta un riesgo relativo significativo, pero en el 2007 marca un aumento bastante notable.



Figura 5.8: Mapa de riesgo relativo para la República Mexicana en 1998.

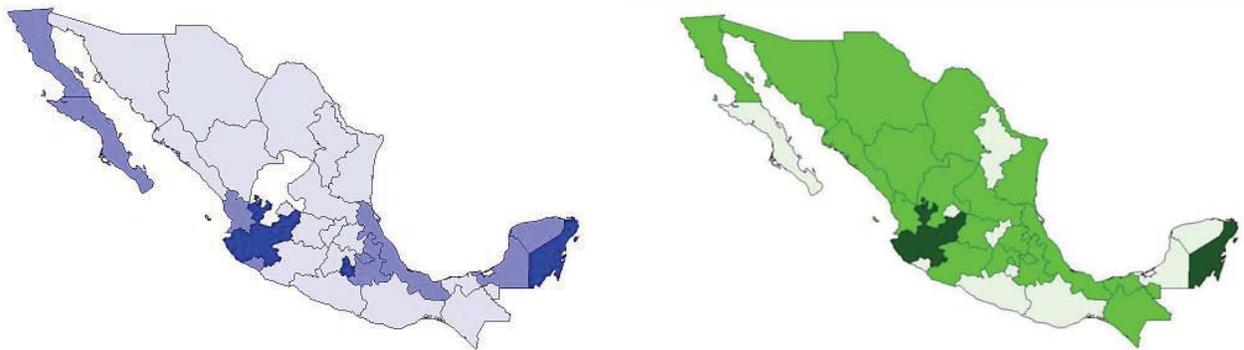


Figura 5.9: Mapas de riesgo relativo y de diferencia de riesgo relativo para la República Mexicana en 1999.

## 5.8. Analisis por estado

En base a los mapas y series de tiempo presentadas, podemos observar ciertos estados de la república que ameritan un análisis mas detallado de su comportamiento tanto espacial como temporal, en esta sección se presenta un análisis estadístico mas elaborado para 5 estados que con ayuda de las herramientas antes mencionadas presentan comportamientos

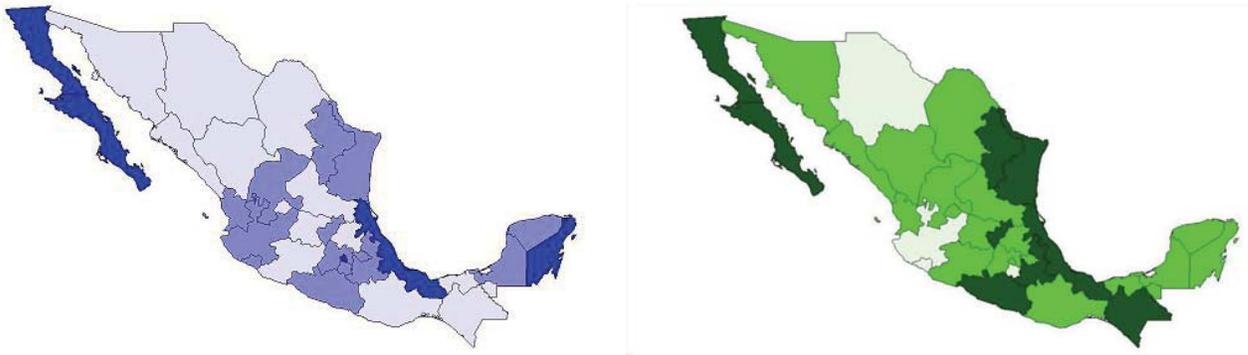


Figura 5.10: Mapas de riesgo relativo y de diferencia de riesgo relativo para la República Mexicana en 2000.

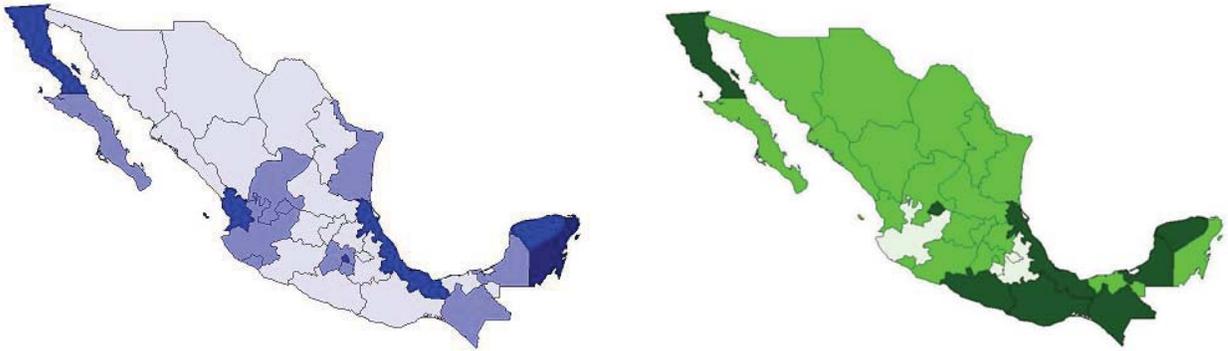


Figura 5.11: Mapas de riesgo relativo y de diferencia de riesgo relativo para la República Mexicana en 2001.

diferentes al resto.

### 5.8.1. Veracruz

Con la ayuda de la tabla de cuantiles (5.4) los mapas de riesgo relativo y las series de tiempo, se puede apreciar que:

- Con el 95 % de probabilidad se puede afirmar que temporalmente existe un aumento en el riesgo relativo de 1998 a 2001. Después de ese periodo no hay evidencia para decir que el riesgo relativo ha aumentado ni disminuido, ya que la moda de los años comprendidos entre 2001 a 2007 siempre están comprendidos en el intervalo del 95 % de probabilidad del riesgo relativo.

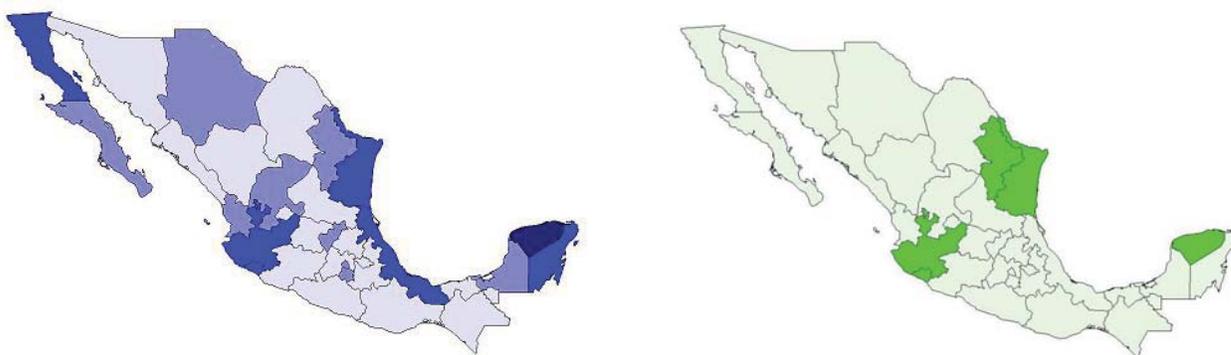


Figura 5.12: Mapas de riesgo relativo y de diferencia de riesgo relativo para la República Mexicana en 2002.



Figura 5.13: Mapas de riesgo relativo y de diferencia de riesgo relativo para la República Mexicana en 2003.

- Los estados vecinos de mayor relevancia a analizar son Puebla, debido a su gran frontera que comparte y Tamaulipas. Con ayuda de la tabla 5.4 podemos ver si el comportamiento de Veracruz es ajeno o no al de sus vecinos. En dicha tabla se muestran en negritas aquellos valores en los que la moda riesgo relativo correspondiente sugieren con un 95 % de probabilidad que el comportamiento en ese año del estado vecino no es diferente al de Veracruz. De esta forma es fácil ver que Puebla no presenta un comportamiento diferente a Veracruz en el periodo de 1998 a 2000, pero en el resto de los años el comportamiento si es diferente. Por su parte Tamaulipas a diferencia de Puebla presenta en la mayoría de los años estudiados un comportamiento similar al de Veracruz con excepción de 2001 y 2006.

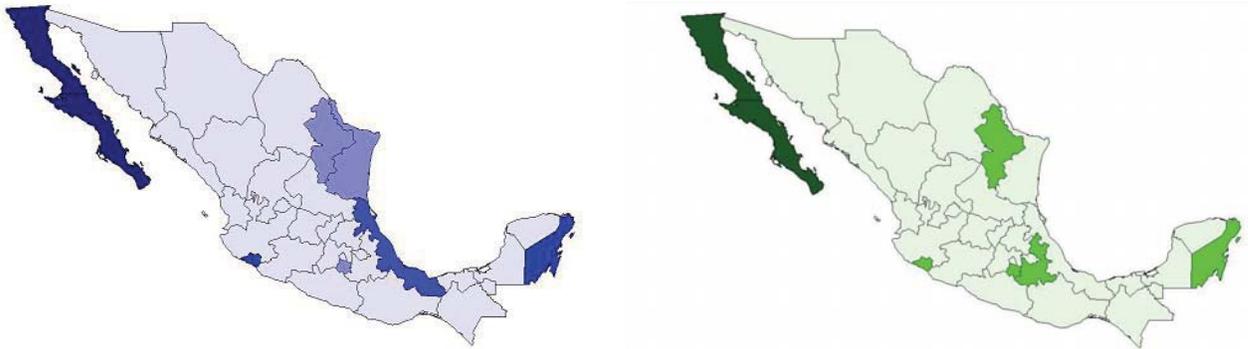


Figura 5.14: Mapas de riesgo relativo y de diferencia de riesgo relativo para la República Mexicana en 2004.

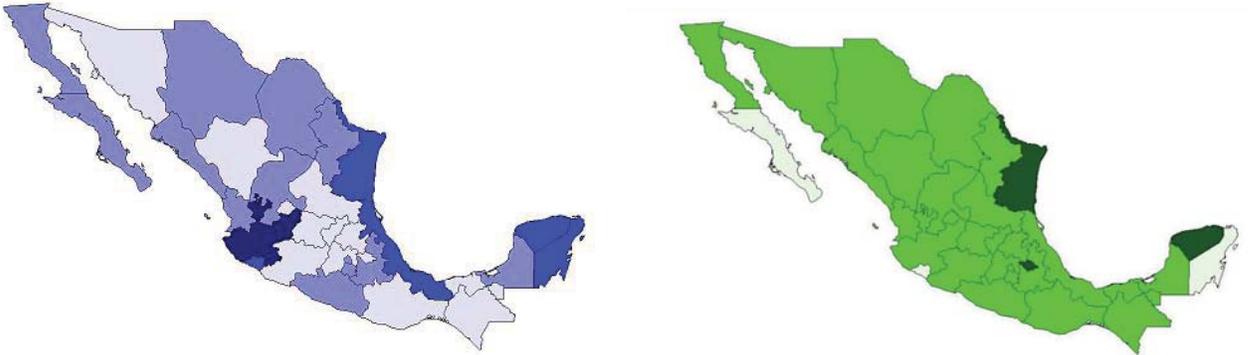


Figura 5.15: Mapas de riesgo relativo y de diferencia de riesgo relativo para la República Mexicana en 2005.

### 5.8.2. D.F.

Basandonos en las herramientas estadísticas presentadas incluyendo la tabla (5.3) podemos decir que:

- Contrastando las distribuciones finales del riesgo relativo del D.F. para todos los años podemos decir que con un 95 % de probabilidad que existe un decrecimiento en el riesgo relativo.
- Con ayuda de los cuatiles presentados en la tabla (5.3) podemos decir con un 95 % de probabilidad que no hay evidencia para pensar que el estado de Morelos presenta un comportamiento diferente al D.F. en 8 de los 10 años que se analizan. Por su parte el estado de México presenta evidencia de no comportarse de la misma forma que el D.F.

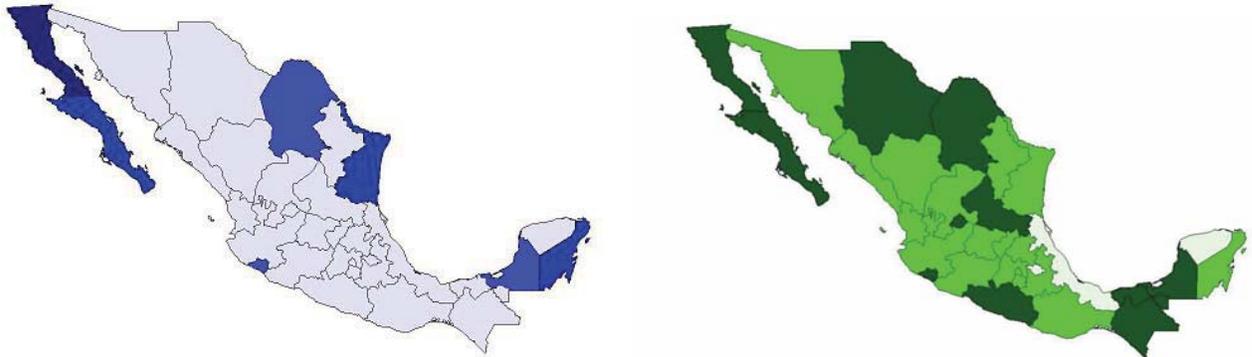


Figura 5.16: Mapas de riesgo relativo y de diferencia de riesgo relativo para la República Mexicana en 2006.

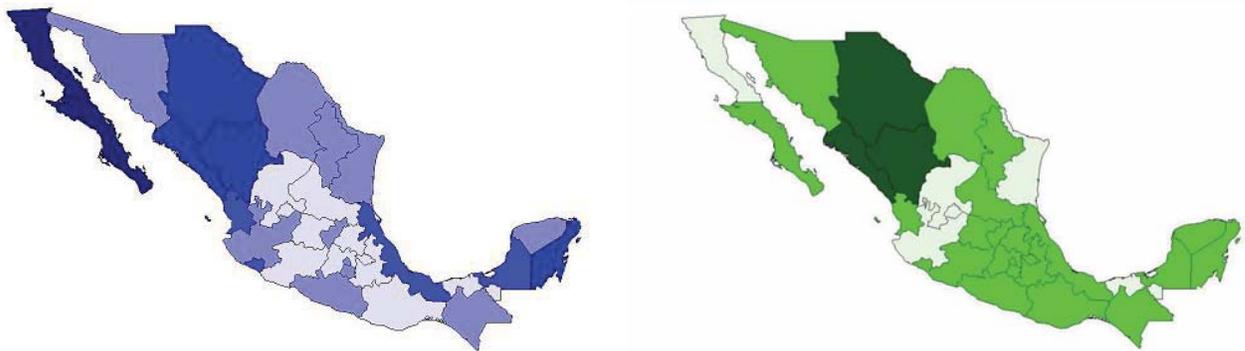


Figura 5.17: Mapas de riesgo relativo y de diferencia de riesgo relativo para la República Mexicana en 2007.

### 5.8.3. Quintana Roo

Bajo la ayuda de las herramientas estadísticas podemos afirmar que:

- Con un 95 % de probabilidad podemos decir que hay evidencia para decir que el riesgo relativo para Quintana Roo es constante para todos los años.
- Además se puede observar que el comportamiento de Quintana Roo no es ajeno al resto de sus estados vecinos ya que no hay evidencia con un 95 % para decir que el comportamiento es diferente en los estados vecinos, por lo cual debería considerarse la región completa como una zona alta de riesgo.

Año	Veracruz			Puebla	Tamaulipas
	Moda	2.5 %	97.5 %	Moda	Moda
1998	3.56	2.11	5.42	<b>4.54</b>	<b>2.93</b>
1999	3.71	2.22	5.60	<b>3.57</b>	<b>2.55</b>
2000	5.31	3.43	7.65	<b>4.85</b>	<b>4.29</b>
2001	6.65	4.48	9.27	2.42	3.37
2002	7.13	4.83	9.97	2.62	<b>6.11</b>
2003	7.09	4.84	9.86	3.59	<b>5.87</b>
2004	7.21	4.91	10.09	4.78	<b>6.59</b>
2005	6.95	4.74	9.67	3.01	<b>8.46</b>
2006	4.99	3.20	7.28	2.72	7.89
2007	6.44	4.27	9.14	2.62	<b>4.49</b>

Cuadro 5.2: Tabla de cuantiles y moda de las distribuciones finales correspondientes al riesgo relativo multiplicado por 1,000,000 de Veracruz y la moda de sus estados vecinos mas representativos desde 1998 hasta 2007.

#### 5.8.4. Jalisco

Bajo la ayuda de la tabla (5.5) podemos afirmar lo siguiente.

- Como lo sugiere la serie de tiempo del riesgo relativo de Jalisco, este estado no presenta una tendencia definida, situación que es fácil de verificar con los cuantiles presentados en la tabla (5.5).
- Al analizar a sus estados vecinos podemos ver que el comportamiento del estado con respecto al riesgo relativo no es ajeno al de Nayarit y Colima, ya que en la mayoría de los casos no hay evidencia para rechazar la suposición de un comportamiento idéntico entre estos estados. EN la tabla (5.5) se muestran en negritas todos aquellos valores que no rechazan dicha hipótesis.

#### 5.8.5. Baja California

- En este estado si hay evidencia de un aumento de riesgo relativo. Si consideramos el periodo de 1998 y 1999 podemos decir que a partir del año 2004 que existe dicha evidencia con 95 % de probabilidad que ha ocurrido un incremento en el riesgo relativo de dicho estado.
- En comparación con su esta vecino mas importante, Baja California Sur, es claro que no hay evidencia para suponer que existe un comportamiento diferente entre ambos.

Año	D.F.			Morelos	México
	Moda	2.5 %	97.5 %	Moda	Moda
1998	6.48	4.36	9.03	<b>7.91</b>	3.45
1999	6.61	4.48	9.22	<b>6.47</b>	2.60
2000	6.99	4.73	9.67	3.94	3.23
2001	7.18	4.93	9.81	4.43	3.55
2002	4.61	2.90	6.77	<b>3.62</b>	2.83
2003	5.15	3.30	7.45	<b>3.84</b>	2.20
2004	5.16	3.29	7.38	<b>5.33</b>	2.54
2005	3.88	2.36	5.84	<b>2.75</b>	<b>2.84</b>
2006	4.45	2.78	6.52	<b>3.12</b>	2.74
2007	4.89	3.08	7.13	<b>3.60</b>	2.51

Cuadro 5.3: Tabla de cuantiles y moda de las distribuciones finales correspondientes al riesgo relativo del D.F. multiplicado por 1,000,000 y la moda de sus estados vecinos desde 1998 hasta 2007.

Año	Quintana Roo			Campeche	Yucatán
	Moda	2.5 %	97.5 %	Moda	Moda
1998	4.34	1.30	9.79	<b>3.43</b>	<b>5.81</b>
1999	6.72	2.57	13.49	<b>6.77</b>	<b>4.15</b>
2000	5.97	2.22	12.09	<b>5.85</b>	<b>3.38</b>
2001	7.57	2.89	14.68	<b>7.63</b>	<b>5.40</b>
2002	6.73	2.57	13.32	<b>6.76</b>	<b>4.86</b>
2003	6.70	2.61	13.35	<b>6.88</b>	<b>5.32</b>
2004	10.81	4.82	20.02	<b>12.71</b>	<b>5.42</b>
2005	6.68	2.55	13.32	<b>6.72</b>	<b>4.50</b>
2006	7.39	2.95	14.60	<b>7.78</b>	<b>5.52</b>
2007	8.77	3.39	16.96	<b>8.92</b>	<b>5.71</b>

Cuadro 5.4: Tabla de cuantiles y moda de las distribuciones finales correspondientes al riesgo relativo multiplicado por 1,000,000 de Quintana Roo y la moda de sus estados vecinos desde 1998 hasta 2007.

Año	Jalisco			Colima	Nayarit
	Moda	2.5 %	97.5 %	Moda	Moda
1998	5.46	3.46	7.97	10.59	<b>4.02</b>
1999	7.05	4.70	9.94	<b>5.51</b>	<b>4.92</b>
2000	3.67	2.18	5.69	<b>3.67</b>	<b>4.95</b>
2001	3.64	2.14	5.61	<b>4.47</b>	<b>5.54</b>
2002	5.21	3.29	7.65	<b>6.03</b>	<b>4.93</b>
2003	8.18	5.56	11.32	<b>6.06</b>	4.50
2004	5.02	3.14	7.39	10.87	<b>3.76</b>
2005	4.63	2.85	6.91	<b>4.33</b>	<b>3.58</b>
2006	4.28	2.61	6.46	<b>5.80</b>	<b>4.43</b>
2007	3.15	1.79	4.97	5.69	6.99

Cuadro 5.5: Tabla de cuantiles y moda de las distribuciones finales correspondientes al riesgo relativo multiplicado por 1,000,000 de Jalisco y la moda de sus estados vecinos mas importantes desde 1998 hasta 2007.

Año	Baja California			Baja California Sur
	Moda	2.5 %	97.5 %	Moda
1998	5.12	2.37	9.08	9.19
1999	4.55	2.13	8.05	<b>5.49</b>
2000	6.75	3.50	11.10	<b>6.75</b>
2001	6.17	3.18	10.43	<b>5.13</b>
2002	5.61	2.79	9.45	<b>4.91</b>
2003	4.97	2.39	8.71	<b>5.54</b>
2004	11.45	6.88	17.46	<b>12.04</b>
2005	8.66	4.85	13.76	<b>6.78</b>
2006	10.30	6.04	15.76	<b>8.49</b>
2007	9.41	5.32	14.87	<b>10.84</b>

Cuadro 5.6: Tabla de cuantiles y moda de las distribuciones finales correspondientes al riesgo relativo multiplicado por 1,000,000 de Baja California y la moda de Baja California Sur, su estado vecino mas importante, desde 1998 hasta 2007.

# Capítulo 6

## Conclusiones

Tener un conocimiento del comportamiento espacio-temporal de la epidemia de SIDA permite tomar acciones preventivas para que eventualmente se pueda controlar esta enfermedad en nuestro país.

Mediante este análisis se pudieron detectar ciertas regiones con diversas características con comportamientos que requieren una acción diferente para controlar en esa zona la epidemia.

Existen ciertas suposiciones hechas en el análisis exploratorio de los datos que con ayuda del modelo ajustado se pueden afirmar. La principal suposición confirmada por el modelo es que alguna de las zonas turísticas de mayor importancia en nuestro país (Península de Yucatán, Península de Baja California, Veracruz y Jalisco) presentan un riesgo mayor que aquellas que no lo son.

La problemática que se presenta en Veracruz, no es ajena a Tamaulipas, por tanto los factores que ocasionan el aumento del riesgo relativo en estos estados, casi seguro son los mismos, por tanto deben trabajar en conjunto para identificarlos y generar medidas de prevención.

Las dos zonas del país con mayor riesgo relativo y que presentan un comportamiento similar son :

- Península de Yucatán
- Península de Baja California

En ambos casos todos los estados pertenecientes a esta región presentan un riesgo relativo mucho mas alto que el resto de los estados. Además estas dos regiones concentran la mayor cantidad de centros turisticos para jóvenes, comparten frontera con otros países y estan en los extremos del país. Por esto es importante que se ponga especial énfasis en crear medidas de prevención eficientes a la par de los programas de turismo y migración.

Jalisco junto con sus dos estados vecinos, Nayarit y Colima son otra región que presentan un riesgo realtivo alto. Su gran concentración de centros turísticos puede colocar dentro del mismo escenario que las Penínsulas antes mencionadas. Bajo esta suposición, las medidas de prevensión aplicadas de forma conjunta con las de turismo podrian beneficiar la situación de esta región.

Bajo este modelo podemos tener evidencia que el D.F. no es un estado cuya influencia en la epidemia sea alarmante, situación que en los análisis preliminares parecia evidenciar que era un estado con un riesgo relativo alto. Incluso con el analisis de sus distribuciones finales se pudo observar evidencia de un decrecimiento en dicho riesgo. Este efecto puede ser originado por la gran cantidad de hospitales que existen, junto el avance que ha existido en los tratamientos para mantener a las personas con VIH viviendo por mas tiempo con la enfermedad.

Después de obtener los resultados ajustados bajo este modelo es claro evidenciar que los modelos jerarquicos espaciales proporcionan una herramienta importante para disminuir las muertes por SIDA, al identificar a las zonas con riesgos realtivos altos y a los estados involucrados en las problematicas de cada región, encontrando similitudes geográficas y comportamiento parecido en el tiempo.

# Apéndice A

## Apéndice

En esta sección se presentan las series de tiempo que corresponden al número de hombres muertos por SIDA, reportados en cada estado, durante el período de 1998 a 2007.

En las gráficas presntadas, no se usa la misma escala, por lo cual es importante verificar la escala en la que cada derie de tiempo esta graficada para no obtener interpretaciones erroneas.

La mayoría de los estados no presentan un comportamiento definido, pero podemos destacar los siguientes estados:

- Con un comportamiento ascendente:
  - Chihuahua
  - Quintana Roo
  - Sonora
  - Tamaulipas
  - Veracruz
  
- Con un comportamiento descendente:
  - D.F
  - Queretaro
  - Tlaxcala

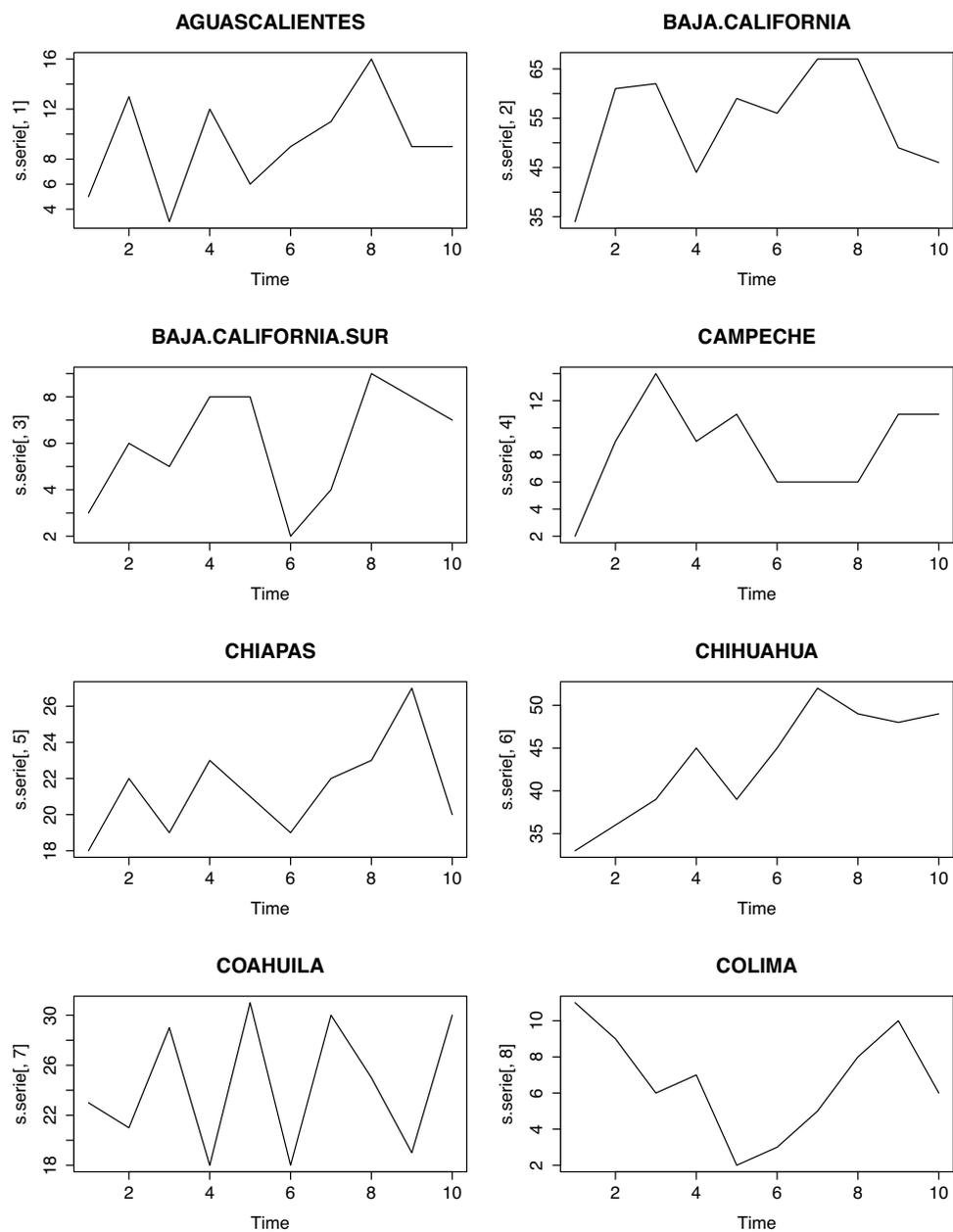


Figura A.1: Series de tiempo que corresponden al número de hombres muertos por SIDA, durante el período de 1998 a 2007.

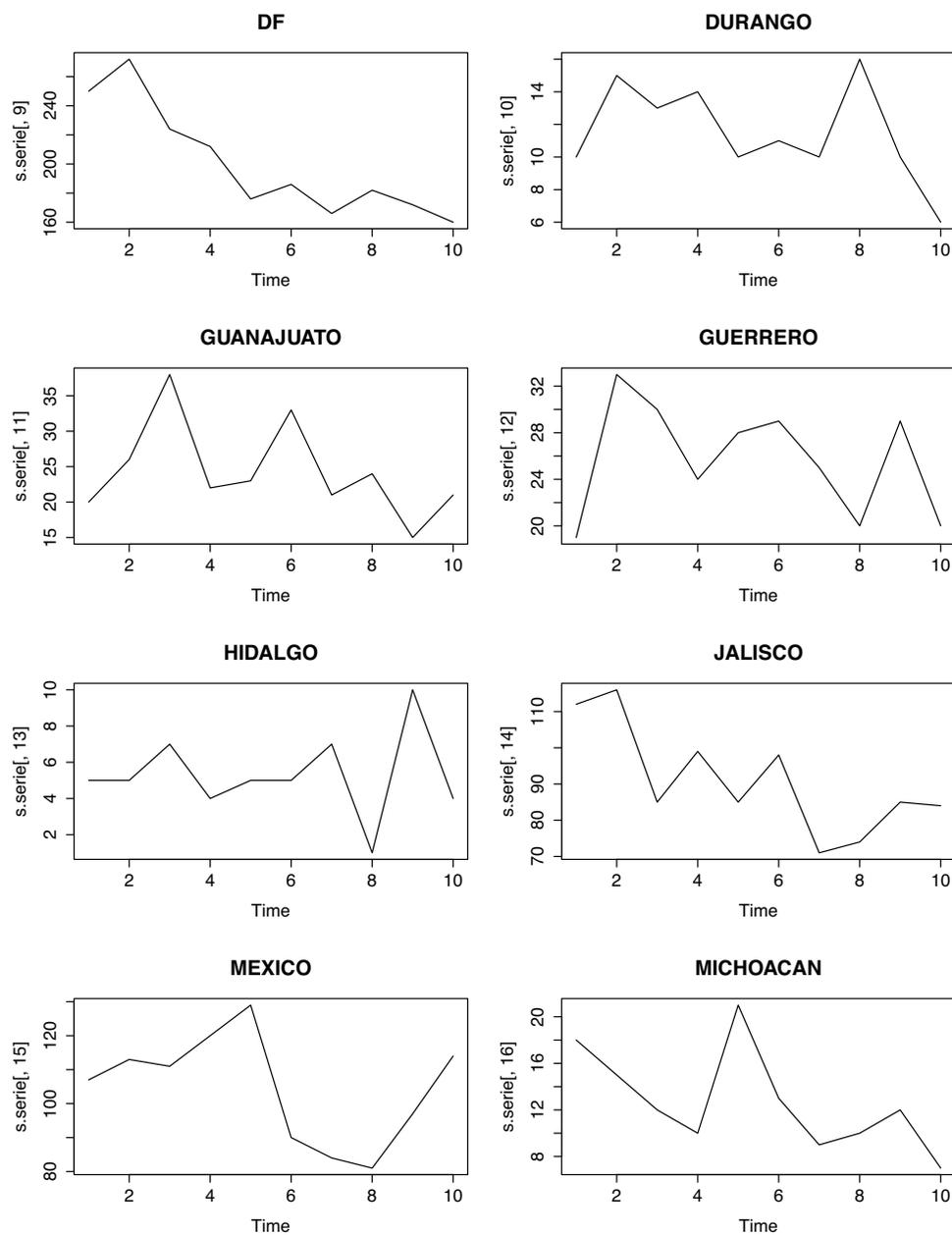


Figura A.2: Series de tiempo que corresponden al número de hombres muertos por SIDA, durante el período de 1998 a 2007.

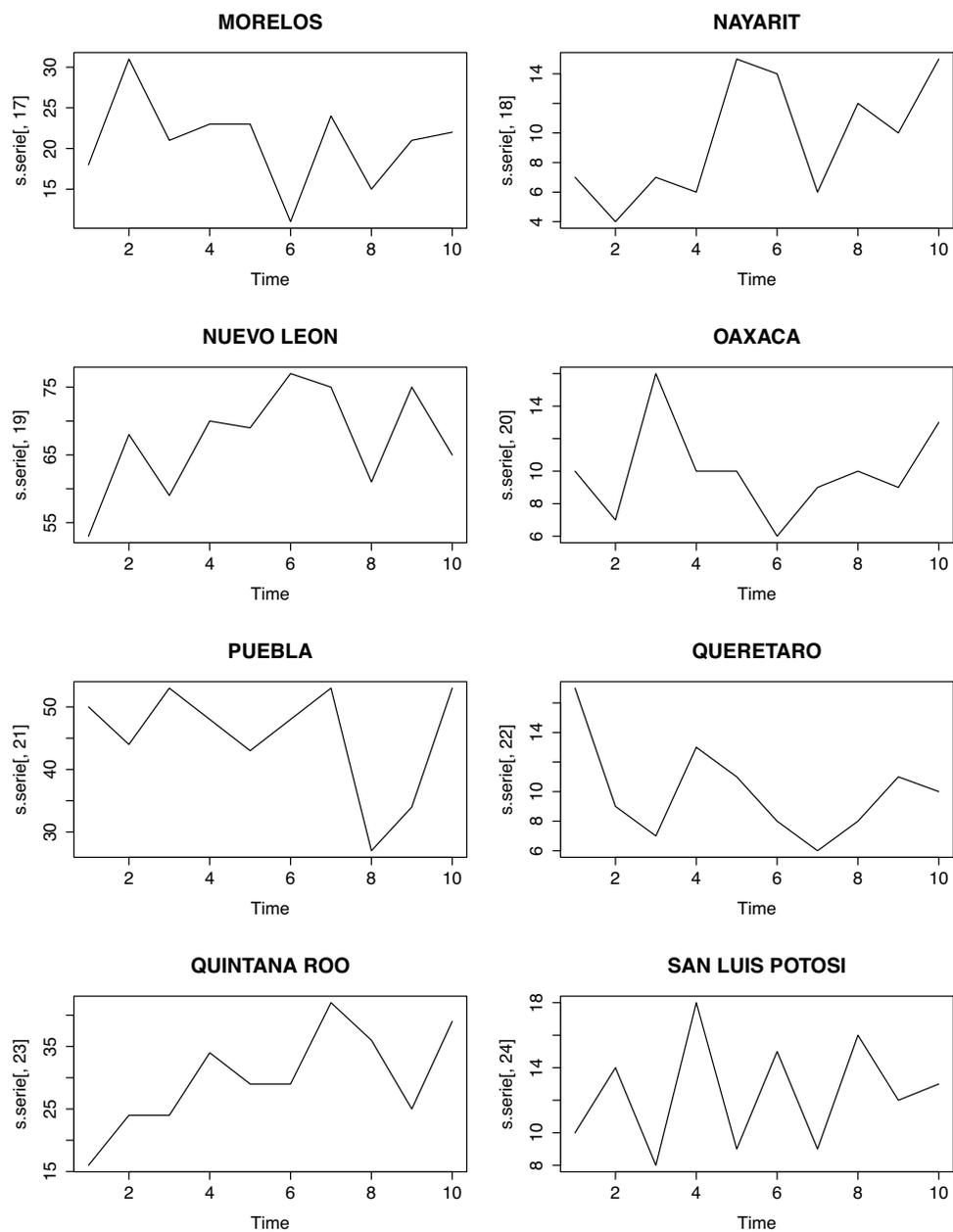


Figura A.3: Series de tiempo que corresponden al número de hombres muertos por SIDA, durante el período de 1998 a 2007.

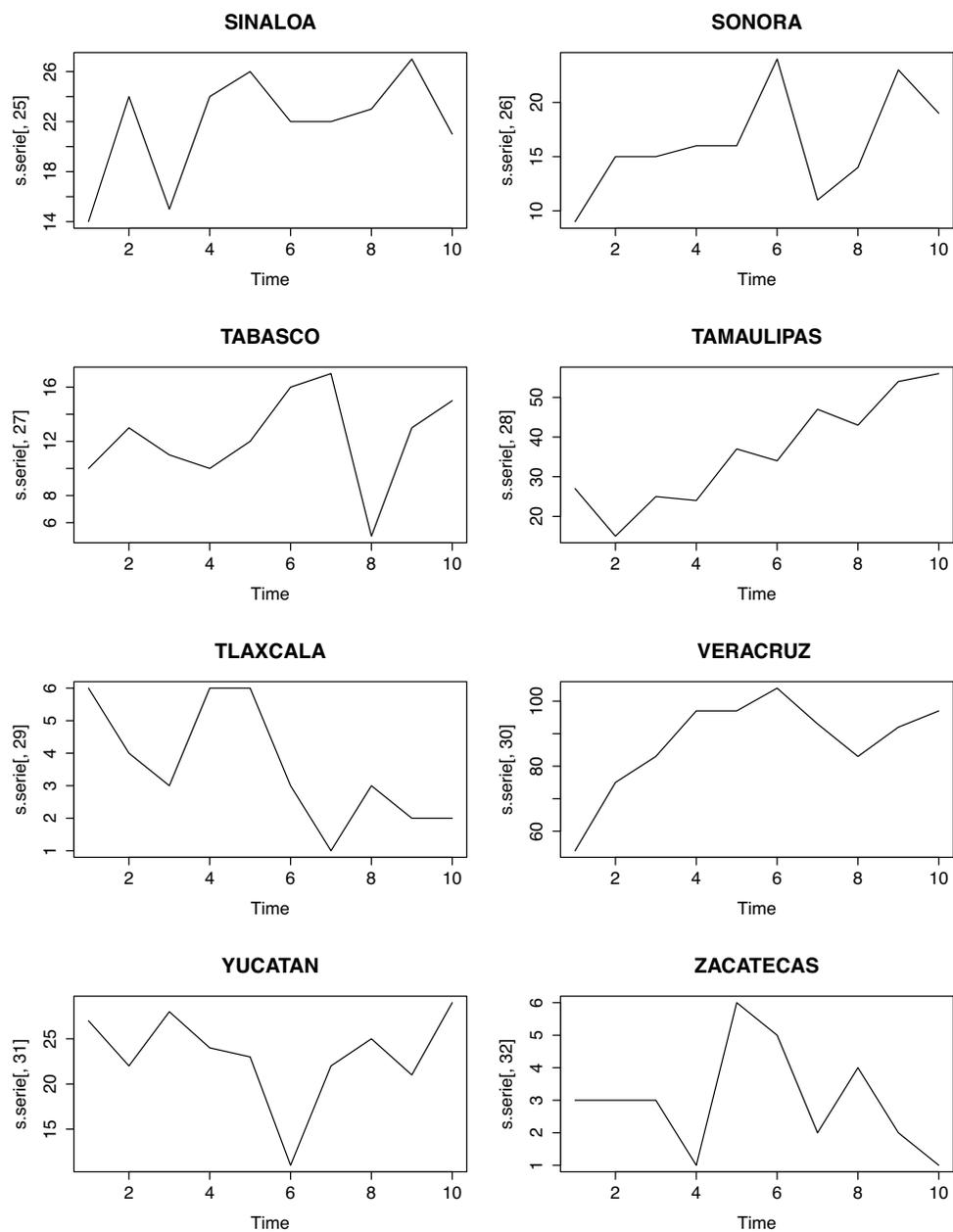


Figura A.4: Series de tiempo que corresponden al número de hombres muertos por SIDA, durante el período de 1998 a 2007.



# Bibliografía

- [1] Waller and Gotway. *Applied Spatial Statistics for Public Health Data*. Wiley, 2004.
- [2] N.A.C. Cressie. *Statistics for Spatial Data*. Wiley , 1993.
- [3] J. Besag. Spatial interactions and the statistical analysis of lattice systems (with discussion). *In Journal of the Royal Statistical Society, Series B*, 36, 192-236, 1974.
- [4] J. Besag, J.C. York and A. Mollie. Bayesian image restoration, with two applications in spatial statistics (with discussion). *In Annals of Institute of Statistical Mathematics*, volume 43, pp. 1-59, 1991.
- [5] Rue and Held. *Gaussian Markov Random Fields*. Chapman & Hall.
- [6] Schabenberger, Gotway. *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC, 2005
- [7] Higareda, Ruiz-Velazco. *Estimación de incidencia futura de casos de SIDA. Una aplicación estadística* UNAM, 2005.
- [8] Albert. *Bayesian Computation with R*. Springer, 2007.
- [9] W.R. Gilks, S. Richardson, D.J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1996.
- [10] Robert. *The Bayesian Choice*. Springer, 2001.
- [11] Arab, Hooten, Wikle. *Hierarchical Spatial Models*
- [12] Ross *Simulation* Elsevier, 2006.
- [13] *Informe sobre la epidemia mundial de SIDA 2008*; ONUSIDA, 2009, <http://www.unaids.org>.

- [14] Sepulveda, García, Morales. *SIDA Información básica para personal de salud* Secretaria de Salud 2000.
- [15] Gutierrez-Peña E. *Apuntes de Analisis bayesiano de datos*. IIMAS, UNAM
- [16] Monroy Cruz E., Gutierrez-Peña E. *WinBUGS: Un software para la inferencai bayesiana* UNAM, 2006.
- [17] Sudipto, Bradley, Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall, 2004.
- [18] Lawson, Browne, VidalRodeiro. *Disease Mapping with WinBUGS and MLwiN*. Wiley, 2003.
- [19] Besag, Knorr-Held. *Modelling Risk from a Disease in Time and Space* 1997.
- [20] Marco Minozzo *Hierarchical Spatial Factor Models for Poisson Count Data*
- [21] A. López-Qúilez *Una visión integral del espacio y el tiempo* Universidad de Valencia.
- [22] Besag, Green, Higdon, Mengersen *Bayesian Computation and Stochastic Systems* Statistical Sciense, Vol 10, No. 1, 3-66, 1995.
- [23] Gaetan, Guyon. *Spatial Statistics and Modelling* Springer, 2010.
- [24] *El VIH en México 2009*; CENSIDA, 2010, <http://www.censida.salud.gob.mx>.