



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Análisis Espacial De Las Enfermedades
Cardiovasculares En México: Modelos
Predictivos

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuario

PRESENTA:

Luis Eduardo González Medina

TUTOR

Dra. Nancy Raquel Mejía Domínguez





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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

<p>1. Datos del alumno González Medina Luis Eduardo (55) 2743 4071 Universidad Nacional Autónoma de México Facultad de Ciencias Actuaría 311154511</p>
<p>2. Datos del tutor Dra. Nancy Raquel Mejía Domínguez</p>
<p>3. Datos del sinodal 1 Dr. Carlos Díaz Ávalos</p>
<p>4. Datos del sinodal 2 Mtro. Bernardo Flores López</p>
<p>5. Datos del sinodal 3 Dra. Lizbeth Naranjo Albarrán</p>
<p>6. Datos del sinodal 4 Dra. Silvia Ruíz Velasco Acosta</p>
<p>7. Datos del trabajo escrito Análisis espacial de las enfermedades cardiovasculares en México: modelos predictivos 69 p 2021</p>

AGRADECIMIENTOS

Investigación realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM IA206717. Agradezco a la DGAPA-UNAM la beca recibida. También se recibió apoyo del proyecto PAPIIT - TA101320 financiado a Dra. Nancy Raquel Mejía Domínguez.

A la Red de Apoyo a la Investigación (RAI).

A la Facultad de Ciencias de la Universidad Nacional Autónoma de México por la formación y el mundo de posibilidades que me brindó.

A mi tutora la Dra. Nancy Raquel Mejía Domínguez por invitarme a formar parte de su proyecto, por las enseñanzas, la guía, la paciencia y la motivación que me brindó durante la realización de este proyecto.

A mis sinodales: Dr. Carlos Díaz Ávalos, Mtro. Bernardo Flores López, Dra. Lizbeth Naranjo Albarrán y Dra. Silvia Ruíz Velasco por su tiempo, disposición, comentarios, sugerencias y entusiasmo en hacer de esta tesis un mejor trabajo.

A mis compañeros de investigación Bernardo, Alejandro y Erick por su ayuda, aportaciones, intercambio de ideas y ánimo que me brindaron siempre para realizar esta tarea.

A mis padres Diana y Pedro por brindarme todo en todos sentidos. Por su ejemplo, amor, paciencia, esfuerzo y sacrificios que me permitieron terminar mis estudios de licenciatura. Lo que he logrado y lo que lograré siempre será por y para ustedes. Significan todo en mi vida.

A mi hermano Rodrigo por desafiarme cada día a ser mejor, por estar para mí y apoyarme bajo cualquier circunstancia, por ser un ejemplo de vida y por hacerme pasar tan buenos ratos. No podría pedir un mejor compañero.

A mis abuelitos Tere y Pedro por quererme tanto, guiarme, apoyarme, consentirme y por el sentimiento de que los hago sentir orgullosos; el orgullo es mío por ser su nieto. Los quiero mucho.

A mi abuelita Elsa† que sé que estaría orgullosa de mí. Te quiero mucho para siempre, estés en donde estés.

A mi tío Lalo por su amistad, por siempre estar al pendiente de mí y por compartir tantas cosas conmigo. Jamás se me van a olvidar tus desmañadas para llevarme a la escuela.

A mis amigos de la Universidad Berni, Roque y Tatos por desvelarse conmigo estudiando cada semestre, por ayudarme y por convertirse en tres hermanos más para mí. También a Gaby, Erick y Ger gracias por su amistad y apoyo siempre. La Facultad no hubiera sido lo mismo sin todos ustedes, me hicieron disfrutarla al máximo.

A Silvia, Nati, Bernardo, Carmen, Alejandro†, Rocío, Karla y Carlos por abrirme las puertas de su casa y adoptarme como uno más de ustedes. Siempre me hicieron sentir como de su familia.

A César y Joss por crecer conmigo, por su lealtad, por las anécdotas y por su apoyo incondicional. Esta etapa, como todas las demás, no hubiera sido tan divertida sin ustedes; su amistad es invaluable.

ÍNDICE GENERAL

Introducción	1
1 FUNDAMENTOS DE LA ESTADÍSTICA ESPACIAL	3
1.1 Campos aleatorios de Markov	3
1.2 Cadenas de Markov Monte Carlo	8
1.3 Introducción a los datos espaciales	10
1.4 Autocorrelación espacial en los modelos estadísticos	14
1.5 Modelos autorregresivos condicionales y modelos autorregresivos simultáneos	18
1.6 Aproximación de Laplace anidada integrada, INLA	21
2 ENFERMEDADES SIN VECTOR DE TRANSMISIÓN	27
3 APLICACIÓN DEL MODELO PREDICTIVO PARA DA- TOS DE ÁREA	33
3.1 Modelo predictivo para las enfermedades cardio- vasculares en México	34
3.2 Resultados	38
3.3 Discusión	49
A CONCEPTOS DE PROBABILIDAD Y PROCESOS ESTO- CÁSTICOS	53
B ESTADÍSTICA BAYESIANA	61
C CÓDIGO DE R	65
Referencias bibliográficas	67

INTRODUCCIÓN

De acuerdo con la OMS, las enfermedades cardiovasculares son un conjunto de trastornos del corazón y de los vasos sanguíneos. Estas enfermedades en los últimos años han mostrado un crecimiento en México y en el mundo. Según la Encuesta Nacional de Salud y Nutrición (ENSANUT) 2018 se estima que al año ocurren aproximadamente ciento cuarenta y dos mil defunciones por este motivo, siendo así la primera causa de muerte en el país, lo cual despierta la curiosidad por saber si esto sucede estadísticamente de igual forma en cualquier región o no.

Los datos sobre las enfermedades cardiovasculares pueden registrarse según el lugar donde vive la persona que la padece. Los datos espaciales pueden analizarse estadísticamente como cualquier otro tipo, sin embargo, cuentan con la singularidad de tener coordenadas que pueden aportar cierta información adicional con la que se pueden tener diversos resultados.

Para este trabajo se utilizaron datos de la ENSANUT 2012, en donde a las personas encuestadas se les preguntó si algún médico les diagnosticó un infarto, angina de pecho, insuficiencia cardíaca, hipertensión y/u otra enfermedad del corazón. Dado que en México cada municipio tiene diferentes características y goza de autonomía administrativa y de participación en los proyectos de desarrollo comunitario se creyó útil e importante desarrollar esta tesis a dicho nivel geográfico.

Uno de los principales retos para estimar el riesgo relativo de padecer alguna enfermedad cardiovascular en cada municipio de México fue la cantidad de datos faltantes, ya que en la ENSANUT solo ochocientos cincuenta de dos mil cuatrocientos cincuenta y siete municipios fueron encuestados. Esta circunstancia es la principal motivación para la realización de esta tesis, ya que mediante técnicas de estimación y análisis de datos de estas enfermedades se pueden tomar decisiones en relación al diagnóstico, evaluación y planeación de programas de salud a nivel nacional.

El objetivo del presente estudio fue estimar el riesgo relativo de padecer alguna enfermedad cardiovascular en cada municipio mexicano para personas mayores a 20 años por medio de la estadística espacial, utilizando el modelo de Besag et al (1991) ajustado por medio de aproximaciones de Laplace. Los resultados se representaron en mapas que permiten visualizar de mejor forma el riesgo relativo en cada municipio, para cada grupo de edad y sexo.

Esta tesis está dividida en tres capítulos: en el primero se presentan las herramientas fundamentales de la estadística espacial e información teórica relacionada a la autocorrelación espacial en los modelos estadísticos, los modelos autorregresivos condicionales y simultáneos, y un apartado dedicado a la aproximación de Laplace anidada integrada, INLA. En el capítulo posterior se habla de la problemática general a nivel mundial y nacional, factores de riesgo y transiciones epidemiológicas de las enfermedades no transmisibles, en particular de las enfermedades cardiovasculares. Por último, se explica la aplicación del modelo de Besag (1991), se reportan los resultados obtenidos y una discusión sobre los mismos. Las referencias y el código de R pueden ser consultados al final del trabajo como apéndice.

FUNDAMENTOS DE LA ESTADÍSTICA ESPACIAL

En este capítulo se desarrolla una introducción, basada en diferentes fuentes y autores, a algunos de los fundamentos y bases que dan vida a la estadística espacial. Lo referente a campos aleatorios de Markov, teoría sobre la cual se puede desarrollar la estadística espacial bajo ciertos tipos de datos fue escrita con base en definiciones de Norris (1997). Para la definición del modelo general espacial y los tipos de datos con los cuales se desarrolla la estadística espacial se utilizaron los textos de Casella y Robert (1999), Gelman et al. (2004), Cressie (1993) y Winkler (1990).

Se busca ahondar en las características de los datos lattices y entender su naturaleza. Se define su estructura, se da un ejemplo para construir el concepto de *vecindad* y se introduce a las posibles *medidas de distancia* con las cuales se estudian este tipo de datos. Además, se plantea la importancia de la autocorrelación espacial en los modelos estadístico-espaciales y se induce a las autorregresiones condicionales. Por último, se definen los modelos condicional y simultáneamente autorregresivos. Del mismo modo, se usaron diversas fuentes: Cressie (1993), Goodchild (1986), el manual de Diggle et al. (2010), la publicación de Wall (2004) y el documento de Rue y Martino (2009).

1.1 CAMPOS ALEATORIOS DE MARKOV

Un campo aleatorio (RF, por su nombre en inglés), en términos generales, es un proceso estocástico que toma valores en un espacio euclidiano y está definido sobre un espacio parametral de dimensión uno, por lo menos (Adler y Taylor, 1993). En esta sección se definen primero otros conceptos y después se entra formalmente al concepto de campo aleatorio que nos interesa ya que a partir de ellos, como con cualquier medida aleatoria, es posible modelar datos espaciales.

Cadenas de Markov

Consideremos un proceso estocástico $\{X_n, n = 0, 1, 2, \dots\}$ que toma un número finito o contable de posibles valores. Si $X_n = i$, se dice que el proceso está en el estado i a tiempo n . Supóngase que siempre que el proceso esté en estado i , hay una probabilidad π_{ij} que luego estará en el estado j . Supongamos que

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \pi_{ij}$$

y que

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_i = i_1, X_0 = i_0) = \pi_{ij}$$

para todos los estados $i_0, i_1, \dots, i_{n-1}, i, j$ y para todo $n \geq 0$. Este proceso estocástico es conocido como *cadena de Markov*. La ecuación anterior puede ser interpretada como que, para una cadena de Markov, la distribución condicional de cualquier estado futuro X_{n+1} , dado los estados pasados X_0, X_1, \dots, X_{n-1} y el estado presente X_n , es independiente de los estados pasados y depende únicamente del estado presente. Lo anterior, es conocido como la *propiedad de Markov*. El valor π_{ij} representa la probabilidad de que el proceso, en el estado i , después hará la transición al estado j . Como las probabilidades son no negativas y como el proceso tiene que hacer la transición a algún estado, se tiene que

$$\pi_{ij} \geq 0, \quad i, j \geq 0; \quad \sum_{j=0}^{\infty} \pi_{ij} = 1, \quad i = 0, 1, \dots$$

Denotemos a π como la matriz de transición de probabilidad de un paso π_{ij} , entonces

$$\pi = \begin{pmatrix} \pi_{00} & \pi_{01} & \pi_{02} & \dots \\ \pi_{10} & \pi_{11} & \pi_{12} & \dots \\ \vdots & & & \\ \pi_{i0} & \pi_{i1} & \pi_{i2} & \dots \\ \vdots & \vdots & & \vdots \end{pmatrix}$$

Veamos el caso de las cadenas de Markov a tiempo discreto, es decir, $T \subset \mathbb{Z}$. Sea \mathcal{X} el espacio de estados numerable o finito.

Sigamos el proceso estocástico $X = \{X_n : n \in \mathbb{T}\}$ con valores (S, S) y digamos que

$$\pi_{ij}^{n,n+1} = \mathbb{P}(X_{n+1} = j | X_n = i), \quad i, j \in S$$

es la probabilidad de transición a un paso del estado i al estado j al tiempo n , mientras que el caso homogéneo en el tiempo, independiente de la posición n ,

$$\pi_{ij} = \pi_{ij}^{0,1}$$

es la probabilidad de transición a un paso. Por otro lado, sea $v = \{v_i, i \in S\}$ el vector de distribución inicial. Entonces, las distribuciones finito dimensionales, en el caso homogéneo en el tiempo, se pueden escribir como

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = v_{i_0} \pi_{i_0, i_1} \pi_{i_1, i_2} \dots \pi_{i_{n-1}, i_n}.$$

Luego, sea $CM_S(v, \pi)$ la cadena de Markov homogénea en el tiempo (discreto), con distribución inicial v , matriz de transición π y valores en S . Sea \mathcal{X} una $CM_S(v, \pi)$, entonces para toda $n, m \geq 0$

1. $\mathbb{P}(X_n = j) = (v\pi^n)_j$
2. $\mathbb{P}(X_n = j | X_0 = i) = \mathbb{P}(X_{n+m} = j | X_m = i) = \pi_{ij}^{(n)}$,

donde

$$(v\pi)_j = \sum_{i \in S} v_i \pi_{ij}, \quad (\pi^2)_{ik} = \sum_{j \in S} \pi_{ij} \pi_{jk}.$$

Por otro lado, para el caso de las cadenas de Markov a tiempo continuo con espacio de estados S numerable, consideremos la matriz $Q = \{q_{ij}\} = (q_{ij} : i, j \in S)$ tal que sus entradas satisfacen que

1. $0 \neq -q_{ii} < \infty$ para toda i .
2. $q_{ij} \geq 0$ para toda $i \neq j$.
3. $\sum_{j \in S} q_{ij} = 0$ para toda i .

Entonces, en cada renglón se puede escoger a los elementos fuera de la diagonal como entradas no negativas sujetos a que $q_i = \sum_{j \neq i} q_{ij} < \infty$, con $q_{ii} = -q_i$.

Por otro lado, a la matriz Q definida como q_{ij} se le conoce como el *generador infinitesimal* de la cadena de Markov a tiempo continuo. Se escribe como

$$Q = \lim_{h \rightarrow 0} \frac{\pi(h) - \pi(0)}{h},$$

es decir, la derivada derecha en cero de la función matricial $t \mapsto \pi(t)$.

Otros conceptos

Independencia condicional

Sea $x_A = \{x_i : i \in A\}$ y $x_{-A} = \{x_i : i \notin A\}$, y $\pi(\cdot)$ la densidad. Se dice que dos variables aleatorias x y y son independientes ($x \perp y$) si

$$\pi(x, y) = \pi(x)\pi(y),$$

es decir, la distribución conjunta es igual al producto de las marginales.

Por otro lado, dos variables x y y son *condicionalmente independientes* de z ($x \perp y | z$) si

$$\pi(x, y | z) = \pi(x | z)\pi(y | z).$$

Notemos que x y y pueden ser marginalmente dependientes, a pesar de que sean condicionalmente independientes de z (Diggle et al., 2010).

Para los siguientes conceptos, las definiciones se obtuvieron del texto de Rue y Held (2005) y Cressie (1993).

Gráfica no dirigida

Una gráfica no dirigida $G = (\mathcal{V}, \mathcal{E})$ es un conjunto de vértices $\mathcal{V} = \{1, \dots, n\}$ y un conjunto de aristas $\mathcal{E} = \{\{i, j\} : i, j \in \mathcal{V}\}$, es

decir, cada lado en \mathcal{E} está asociado a un par no ordenado de vértices.

Sistema de vecindades y vecino

Un *sitio* s se refiere a un punto específico en el tiempo o a un punto en particular en un espacio de dos o más dimensiones, o áreas particulares en una región geográfica. Luego, se dice que un sitio $s \neq i$ es vecino de un sitio i si $(i, s) \in \mathcal{E}$. Y un conjunto de sitios que consta de un sitio o de sitios que son vecinos entre sí es llamado un clique.

Con lo anterior, podemos definir que una colección $\partial = \{\partial(s) : s \in D\}$ de subconjuntos de D es llamado un sistema de vecindades si $s \notin \partial(s)$ y $s \in \partial(i)$ si y sólo si $i \in \partial(s)$, en donde D es un conjunto de sitios.

Distribución normal multivariada

Se dice que una variable aleatoria $\mathbf{X} = (X_1, \dots, X_n)^T$ se distribuye Normal con media μ y matriz simétrica positiva definida de covarianzas Σ si la forma de su distribución es

$$\pi(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right\}, \mathbf{x} \in \mathbb{R}^n.$$

Lo denotamos como $X \sim \mathcal{N}(\mu, \Sigma)$.

Campos aleatorios

Dado un espacio de probabilidad (Ω, \mathcal{F}, P) , un campo aleatorio es una colección de variables aleatorias que toman valores en X indexadas por elementos en un espacio topológico T . Es decir, un campo aleatorio F es una colección

$$\{F_t : t \in T\}$$

donde cada F_t es una variable aleatoria que toma valores en X .

Luego, el campo aleatorio π es de Markov con respecto al sistema de vecindades ∂ si para todo $\mathbf{x} \in \Omega$,

$$\pi(S_s = x_s | S_r = x_r, r \neq s) = \pi(S_s = x_s | S_r = x_r, r \in \partial(s)).$$

Y se dice que un vector $x \in \mathbb{R}^n$ es un campo aleatorio de Markov Gaussiano (*GMRF*) con respecto al gráfico etiquetado no dirigido $G = (\mathcal{V}, \mathcal{E})$, con media μ y con matriz de precisión simétrica positiva definida Q , si su densidad tiene la forma

$$\pi(x) = (2\pi)^{-n/2} |Q|^{1/2} \exp \left\{ -\frac{1}{2} (x - \mu)^T Q (x - \mu) \right\}$$

1.2 CADENAS DE MARKOV MONTE CARLO

En estadística bayesiana existen varios métodos de simulación que permiten obtener información sobre alguna distribución de manera numérica. Las *cadena de Markov Monte Carlo* (MCMC, por su nombre en inglés) son una sucesión de variables aleatorias que se pueden considerar que evolucionan en el tiempo, con probabilidad de transición dependiendo del conjunto particular en donde la cadena esté. Estos métodos se basan en la idea de construir una cadena de Markov de tal forma que converja a la distribución deseada, para después muestrear de esta simulando una trayectoria de la cadena. Se puede definir la cadena en términos de su kernel de transición.

Kernel de transición

Un *kernel de transición* es una función K definida en $\mathcal{X} \times \mathcal{B}(\mathcal{X})$ tal que

1. Para toda $x \in \mathcal{X}$, $K(x, \cdot)$ es una medida de probabilidad.
2. Para toda $A \in \mathcal{B}(\mathcal{X})$, $K(\cdot, A)$ es medible.

Nota: cuando \mathcal{X} es discreta, el kernel de transición es una matriz K con elementos

$$\pi_{xy} = \mathbb{P}(X_n = y | X_{n-1} = x), \quad x, y \in \mathcal{X}.$$

En el caso continuo, el kernel también denota la densidad condicional $K(x, x')$ de la transición $K(x, \cdot)$, es decir,

$$\mathbb{P}(X \in A|x) = \int_A K(x, x') dx'.$$

Para garantizar que el algoritmo sea lo suficientemente preciso, es necesario que la cadena de Markov resultante cumpla con ciertas propiedades.

Irreducibilidad

Dada una medida ϕ , se dice que una cadena de Markov X_t con kernel de transición $K(x, y)$ es ϕ -rreducible si para todo $A \in \mathcal{X}^2$ con $\phi(A) > 0$, existe una n tal que $K^n(x, A) > 0$ para toda $x \in X$, semejantemente, $\mathbb{P}(\tau_A < \infty) > 0$, con $\tau_A = \inf\{n \geq 1; X_n \in A\}$. Esto se refiere a que todos los estados de la cadena se comunican, es decir, que hay una probabilidad mayor que cero de llegar a cualquier estado desde cualquier estado.

Lo anterior no asegura la convergencia de la cadena, pues solo garantiza que todo estado será visitado por ella, es por eso que se necesita de otra propiedad para tener un barrido «estable» del espacio.

Recurrencia

En un espacio finito X , un estado $\omega \in X$ es transitorio si el número esperado de visitas,

$$\mathbb{E}_\omega(\eta_\omega) = \mathbb{E}_\omega \left(\sum_{n=1}^{\infty} \mathbb{1}_\omega(X_n) \right),$$

es finito, y es recurrente si $\mathbb{E}_\omega(\eta_\omega) = \infty$.

Periodicidad

Una cadena ϕ -rreducible tiene un ciclo de longitud d si existe un conjunto pequeño C , un entero M y una medida de probabilidad μ tal que d es el máximo común denominador de $\{m \geq 1; \exists \delta_m > 0 \text{ tal que } C \text{ es pequeño para } \mu > \delta_{mM}\}$. Y se dice que la cadena es aperiódica si $d = 1$.

Anteriormente se mencionó el concepto de *conjunto pequeño*. Un conjunto C es pequeño si existe $m \in \mathbb{N}$ y una medida positiva m tal que

$$K^m(x, A) \geq \epsilon(A), \quad \forall x \in C, \forall A \in \mathcal{X}.$$

En los algoritmos MCMC, generalmente, se busca que las cadenas construidas tengan una distribución estacionaria. Este concepto se refiere a la marginal del proceso evaluada en cada tiempo y puede ser expresada en términos del kernel de transición.

Una medida π es *invariante* o *estacionaria* para el kernel de transición $K(.,.)$ y para su cadena asociada si

$$\pi(B) = \int_{\mathcal{X}} K(x, B)\pi(dx), \quad \forall B \in \mathcal{X}.$$

Por último, se necesita la condición de invertibilidad. Se dice que una cadena de Markov estacionaria es invertible si sus distribuciones finito-dimensionales son invariantes al flujo del tiempo, es decir, si la distribución $X_n|X_{n+1}$ es igual a la de $X_{n+1}|X_n$.

1.3 INTRODUCCIÓN A LOS DATOS ESPACIALES

El análisis espacial es un término que se podría definir como la representación de técnicas y modelos, que usan la referencia espacial asociada a cada dato u objeto que es especificado dentro del sistema estudiado (Haining, 2003). En esta sección se introducirá al modelo espacial general y a los tres tipos de datos que se usan en la estadística espacial, es decir, datos *geoestadísticos*, *puntuales* y *lattices* o de *área*.

Para la realización de este documento se usaron datos lattices, por lo que no se busca profundizar en los conceptos de geoestadística y datos puntuales, sin embargo, en el próximo capítulo sí

se darán las características de los datos de área. Ésta es entonces, una sección a nivel intuitivo para darse una idea de lo que son estos datos y/o para qué han sido utilizados en la práctica.

Modelo espacial general

Sea $s \in \mathbb{R}^d$ una ubicación de datos en un espacio euclídeo d -dimensional y supóngase que $Z(s)$ es una variable aleatoria en una ubicación s . Si se hace variar s sobre el conjunto de índices (sitios) $D \subset \mathbb{R}^d$ se obtiene el campo aleatorio

$$\{Z(s) : s \in D\}$$

y sus realizaciones se denotan como

$$\{z(s) : s \in D\}.$$

Tipos de datos espaciales

Datos geoestadísticos

La geoestadística es una disciplina híbrida aplicada a la ingeniería minera, geología, matemáticas y estadística (Cressie, 1993), y tiene como objetivo proporcionar descripciones cuantitativas de variables distribuidas en el espacio o en el tiempo-espacio, en donde se asume que dichos valores tienen una correlación entre sí (Chilès y Delfiner, 1999).

Aquí nos ocupamos de modelar datos observando realizaciones (incompletas) de un proceso aleatorio $\{Z(s) : s \in D\}$, donde el conjunto de índices D permite que s varíe continuamente a lo largo de una región del espacio euclidiano d -dimensional. Además de necesitar las observaciones $\{z(s_1), \dots, z(s_n)\}$, se requiere conocer las ubicaciones espaciales $\{s_1, \dots, s_n\}$ (Cressie, 1993).

Patrones puntuales

Los patrones puntuales se diferencian de los dos anteriores tipos de datos porque las variables de interés son las localizaciones

y no las mediciones. Sea el proceso $\{Z(s_i) : s_i \in D\}$, en donde D es un conjunto de eventos aleatorios en donde ocurren las realizaciones de un proceso puntual

$$X = \cup \{Z(s_i) : s_i \in D\} \text{ (Cressie, 1993).}$$

Datos lattices

Los datos lattices, de área o de rejilla nacen cuando en el proceso estocástico $\{Z(s_i) : s_i \in D\}$, D es un conjunto contable de sitios espaciales en la que los datos son observados, este conjunto es llamado *lattice* o *rejilla* y es complementado con la información que brindan los vecinos, por lo que, matemáticamente hablando, los sitios son vértices que están conectados por aristas, dando lugar a una estructura de vecindad (Cressie, 1993).

La figura 1 muestra un mapa numerado, por orden alfabético, de los estados de la República Mexicana. A continuación se adjunta la lista con índices y estado para la identificación de ellos.

- | | | |
|------------------------|----------------|---------------------|
| 1. Aguascalientes | 12. Guerrero | 23. Quintana Roo |
| 2. Baja California | 13. Hidalgo | 24. San Luis Potosí |
| 3. Baja California Sur | 14. Jalisco | 25. Sinaloa |
| 4. Campeche | 15. México | 26. Sonora |
| 5. Chiapas | 16. Michoacán | 27. Tabasco |
| 6. Chihuahua | 17. Morelos | 28. Tamaulipas |
| 7. Ciudad de México | 18. Nayarit | 29. Tlaxcala |
| 8. Coahuila | 19. Nuevo León | 30. Veracruz |
| 9. Colima | 20. Oaxaca | 31. Yucatán |
| 10. Durango | 21. Puebla | 32. Zacatecas |
| 11. Guanajuato | 22. Querétaro | |



Figura 1: Mapa de México numerado por estado en orden alfabético

Sin ayuda del mapa, el conjunto de índices $\{1, 2, \dots, 32\}$ por sí solo no brinda información espacial alguna y por lo tanto no se le puede catalogar como un lattice espacial. Sin embargo, si a esos índices les agregamos información como latitud x y longitud y , sí es posible formar un lattice, es decir, un conjunto de la forma

$$D = \{(x_i, y_i) : i = 1, 2, \dots, 32\}.$$

Vecindad

Desde el lattice espacial D la información de la vecindad puede ser especificada, por ejemplo, basados en la distancia euclídeana. Supongamos que tenemos las coordenadas de la capital de cada estado de la Figura 1. Llamémosle *vecino* de la capital de la entidad i a cada capital del estado que esté a una distancia j de i . Supongamos que $j = 200\text{km}$, entonces bajo este razonamiento, las capitales de los estados de Hidalgo (13), México (15), Morelos (17), Puebla (21), Querétaro (22) y Tlaxcala (29) son vecinos de la Ciudad de México (7).

Sean

$$\partial_i = \{k : k \text{ es un vecino de } i\}, \quad i = 1, \dots, 32.$$

y

$$D_{\partial} = \{(i; \partial_i) : i = 1, \dots, 32\}.$$

El conjunto D_{∂} es también un lattice espacial, pues aunque no contiene la información exacta de la ubicación de los sitios, sí se puede construir un modelo de dependencia espacial entre los estados; en términos matemáticos D_{∂} es un gráfico.

Medida de distancia

Es importante aclarar que para un análisis espacial basado en distancia, las medidas pueden variar de un estudio a otro. Por ejemplo, medir la distancia entre eventos juega un papel importante cuando los eventos viven en una red lineal. Un segmento de línea en el plano con extremos u y v se puede escribir en forma paramétrica como $[u, v] = \{tu + (1 - t)v : 0 \leq t \leq 1\}$.

Sea $|L|$ la longitud total de todos los segmentos de línea. La distancia entre dos puntos u y v en L generalmente se calcula mediante la distancia del camino más corto $d_L(u, v)$ que es el mínimo de la longitud de todos los caminos posibles entre u y v . Sin embargo, se han obtenido diferentes distancias posibles.

También, se podría utilizar la distancia Euclideana $[(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2}$. En algún otro tipo de estudio, la distancia podría estar dada por unidades de tiempo o de tiempo-costo de un lugar a otro. Por lo tanto, es tarea del investigador definir la medida de acuerdo a su estudio.

1.4 AUTOCORRELACIÓN ESPACIAL EN LOS MODELOS ESTADÍSTICOS

Autocorrelación espacial

Tobler (1970) en la *Primera Ley de la Geografía* formula que "todo está relacionado entre sí, pero las cosas más cercanas están más relacionadas que las lejanas". A partir de ese pensamiento es que se define a la autocorrelación espacial como el grado de similaridad que existe entre objetos o actividades en cierta

ubicación con otros que se ubican cerca, y se puede interpretar a la misma como un índice descriptivo, que mide aspectos de la distribución de las cosas en el espacio, pero que al mismo tiempo puede verse como un proceso causal que mide el grado de influencia que ejerce algo sobre sus vecinos.

En cuanto al análisis espacial, este se tiene que trabajar bajo dos tipos distintos de información: características espaciales y la ubicación espacial. El primer tipo se refiere a los atributos espaciales y a las covariables, por ejemplo en un contexto médico-social, si se cuenta o no con seguro social, si son de sexo masculino o femenino, edad, estatura, peso, etc. Mientras que el segundo tipo se refiere a la localización descrita por su posición en un mapa, es decir, al municipio en el que se encuentran, por ejemplo.

La autocorrelación espacial es una propiedad del proceso espacial estudiado. La forma de estimarla es a través de la comparación de dos conjuntos de similaridades por medio de técnicas que ayudan a estimar dichas estadísticas. Esto es, si las características que son similares en la ubicación también tienden a ser similares en los atributos, entonces se dice que el patrón muestra correlación espacial positiva. Por otro lado, se dice que el patrón muestra correlación espacial negativa, cuando las características cercanas tienden a no ser similares en atributos, mientras que con las lejanas sí lo son. Por último, cuando los atributos son independientes de la ubicación, se dice que la autocorrelación espacial es cero.

Autorregresiones condicionales

Tradicionalmente, las autorregresiones condicionales se han utilizado para estimar o predecir procesos estocásticos con correlación espacial en datos observados de una gráfica predefinida o en una estructura lattice. Consideremos un vector aleatorio $X = (X_1, \dots, X_n)$ en donde cada componente es univariado y está localizado en un sitio fijo $i \in \{1, \dots, n\}$, y X es markoviano.

La distribución conjunta con densidad $\pi(x)$ para X se define como

$$\pi(x) = \pi(x_1) \cdot \pi(x_2|x_1) \cdot \pi(x_3|x_1, x_2) \cdot \dots \cdot \pi(x_n|x_1, x_2, \dots, x_{n-1})$$

que aplicando la propiedad de Markov, se simplifica a

$$\pi(x) = \pi(x_1) \cdot \pi(x_2|x_1) \cdot \pi(x_3|x_2) \cdot \dots \cdot \pi(x_n|x_{n-1}).$$

dando como resultado a la factorización que se le conoce como autorregresivo de primer orden, el cual puede ser modificado para órdenes mayores.

En un contexto espacial, los índices $i \in \{1, \dots, n\}$ son arbitrarios y pueden ser permutados (Diggle et al., 2010). En este sentido, es más natural especificar la distribución condicional completa $\pi(x_i|x_{-i})$, la distribución condicional de X_i en un sitio i , dados los valores $X_j = x_j$ en otros sitios $j \neq i$. Bajo la misma idea, asumiendo la propiedad de Markov, la distribución condicional $\pi(x_i|x_{-i})$ depende solo de algunos componentes de x_{-i} , es decir, de los de vecinos del sitio i , donde $x_{-i} = \{x_j : j \neq i\}$.

Uno de los modelos más estudiados es el caso gaussiano, en donde $\pi(x_i|x_{-i})$ se distribuye Normal y $\pi(x)$ se distribuye normal multivariado. Primero se definirán algunos conceptos que son necesarios para la comprensión de estos modelos y más adelante se definirá el caso descrito anteriormente.

Especificación condicional

Una especificación condicional define a los condicionales completos $\pi(x_i|x_{-i})$ como una distribución Normal con momentos

$$\mathbb{E} [x_i|x_{-i}] = \mu_i + \sum_{j \neq i} \beta_{i,j} (x_j - \mu_j) \quad \text{y} \quad \text{precisión}(x_i|x_{-i}) = k_i > 0. \quad (1)$$

Comparando la Ecuación (1) con la distribución de un GMRF, podemos tomar μ como la media, $Q_{ij} = k_i$, $\beta_{i,j} = \frac{Q_{i,j}}{Q_{i,i}}$ y así obtener los mismos condicionales completos. Como Q es simétrica, entonces para todo i

$$k_i \beta_{i,j} = k_j \beta_{j,i}. \quad (2)$$

Entonces, si $\beta_{i,j} \neq 0$, $\beta_{j,i}$ también es distinta de cero. Las aristas en el gráfico G están definidas como el conjunto $\{\{i, j\} : \beta_{i,j} \neq 0\}$. Q , además de ser simétrica, es positiva y definida, y al ser una propiedad de la distribución conjunta, es difícil validarla en términos locales. Sin embargo, podemos escoger una parametrización con la cual Q sea una matriz simétrica positiva definida, tal que, $Q_{i,i} > \sum_j |Q_{i,j}|$ para toda i . Por lo tanto, para toda i se tiene que

$$\sum_j |\beta_{i,j}| < 1.$$

Luego, podemos identificar una Normal con los mismos condicionales completos, teniendo en cuenta que la densidad conjunta es única. Una técnica para obtener esta distribución es la resultante del *Lema de Brooke* consultada en Rue y Held (2005).

Lema de Brooke

Sea $\pi(x)$ la densidad de $x \in \mathbb{R}^n$ y sea $\Omega = \{x \in \mathbb{R}^n : \pi(x) > 0\}$. Sean $x, x' \in \Omega$, entonces

$$\frac{\pi(x)}{\pi(x')} = \prod_{i=1}^n \frac{\pi(x_i | x_1, \dots, x_{i-1}, x'_{i+1}, \dots, x'_n)}{\pi(x'_i | x'_1, \dots, x'_{i-1}, x_{i+1}, \dots, x_n)} \quad (3)$$

$$= \prod_{i=1}^n \frac{\pi(x_i | x'_1, \dots, x'_{i-1}, x_{i+1}, \dots, x_n)}{\pi(x'_i | x'_1, \dots, x'_{i-1}, x_{i+1}, \dots, x_n)}. \quad (4)$$

Demostración del Lema de Brooke

Por probabilidad condicional se tiene que

$$\frac{\pi(x_n | x_1, \dots, x_{n-1})\pi(x_1, \dots, x_n)}{\pi(x'_n | x_1, \dots, x_{n-1})\pi(x_1, \dots, x'_n)} = \frac{\pi(x_1, \dots, x_n)}{\pi(x_1, \dots, x'_n)}$$

por lo que

$$\pi(x_1, \dots, x_n) = \frac{\pi(x_n | x_1, \dots, x_{n-1})}{\pi(x'_n | x_1, \dots, x_{n-1})} \pi(x_1, \dots, x'_n)$$

Aplicando de manera inductiva sobre la entrada x_{n-1} se tiene

$$\begin{aligned}\pi(x_1, \dots, x_n) &= \frac{\pi(x_n | x_1, \dots, x_{n-1})}{\pi(x'_n | x_1, \dots, x_{n-1})} \\ &\quad * \frac{\pi(x_{n-1} | x_1, \dots, x_{n-2}, x'_n)}{\pi(x'_{n-1} | x_1, \dots, x_{n-2}, x'_n)} \\ &\quad * \pi(x_1, \dots, x_{n-2}, x'_{n-1}, x'_n)\end{aligned}$$

Por lo tanto,

$$\frac{\pi(x)}{\pi(x')} = \prod_{i=1}^n \frac{\pi(x_i | x_1, \dots, x_{i-1}, x'_{i+1}, \dots, x'_n)}{\pi(x'_i | x'_1, \dots, x'_{i-1}, x_{i+1}, \dots, x_n)}$$

1.5 MODELOS AUTORREGRESIVOS CONDICIONALES Y MODELOS AUTORREGRESIVOS SIMULTÁNEOS

Una de las formas para modelar a los datos espaciales lattice, sin ignorar su forma discreta, es definiendo una estructura de vecindad basada en la forma de la rejilla. Así, en lugar de considerar la distancia entre los centros de las regiones, es usado un sistema en donde se definen las regiones vecinas basadas en, por ejemplo, dónde sus bordes se tocan o no.

Una vez definida esta vecindad, los modelos se asemejan a los modelos autorregresivos utilizados en series de tiempo. Dos tipos de modelos populares que incorporan esta información discreta de las vecindades, son los modelos autorregresivos condicionales (CAR) y simultáneos (SAR) (Wall, 2004).

Para ambos modelos, consideremos $\{Z(s_i) : s_i \in (s_1, s_2, \dots, s_n)\}$ un proceso gaussiano, en donde $\{s_1, s_2, \dots, s_n\}$ pertenecen al lattice espacial D . Por lo cual, decimos que los sitios $\{s_1, \dots, s_n\}$ forman un lattice en D si esos sitios forman una partición de D .

Modelos autorregresivos condicionales, CAR

Como se dijo anteriormente, una de las formas para modelar el proceso $\{Z(s_i) : s_i \in (s_1, s_2, \dots, s_n)\}$, es por medio de los modelos CAR, introducidos por Besag (1974), los cuales son principalmente populares en el diseño de modelos espaciales jerárquicos para ser usados sobre datos de área (Arab et al., 2006). El modelo es el siguiente:

$$Z(s_i)|Z(s_{-i}) \sim \mathcal{N} \left(\mu_i + \sum_{j=1} c_{ij}(Z(s_j) - \mu_j), \tau_i^2 \right),$$

en donde:

- $Z(s_{-i}) = \{Z(s_j) : i \neq j\}$.
- $\mathbb{E} [Z(s_i)] = \mu_i$.
- τ_i^2 : la varianza condicional.
- c_{ij} : constantes tales que $c_{ii} = 0$ par toda $i = 1, \dots, n$.

Entonces, si n es finita, C es la matriz que contiene a las constantes y $T = \text{diag}(\tau_1^2, \dots, \tau_n^2)$. Por lo tanto, la distribución de $Z = (Z(s_1), Z(s_2), \dots, Z(s_n))^T$ se expresa como

$$Z \sim \mathcal{N}(\mu, (I - C)^{-1} T).$$

Modelos autorregresivos simultáneos, SAR

Una forma de implementar estos procesos es por medio de los modelos, lo cuales fueron desarrollados por Whittle (1954) como una clase de modelos espaciales para datos de área y son un subconjunto de los MRFs. El modelo es el siguiente:

$$Z(s_i) = \mu_i + \sum_{j=1}^n b_{ij} (z_j - \mu_j) + \epsilon_i$$

en donde,

- $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T \sim \mathcal{N}(0, \Lambda)$ con Λ una matriz diagonal.
- $\mathbb{E}[Z(s_i)] = \mu_i$.
- b_{ij} son constantes tales que $b_{ii} = 0$ para toda $i = 1, \dots, n$.

Luego, si tenemos un número finito de datos, podemos tomar a B como la matriz que contiene a los b_{ij} . Estos modelos son llamados *simultaneos* porque los errores ϵ_i estarán correlacionados con el proceso $\{Z(s_j) : j \neq i\}$.

Entonces, la distribución conjunta de $Z = (Z(s_1), Z(s_2), \dots, Z(s_n))^T$ es

$$Z \sim \mathcal{N}\left(\mu, (I - B)^{-1} \Lambda (I - B)^{-1T}\right), \quad (5)$$

en donde $\mu = (\mu_1, \dots, \mu_n)$ e I es la matriz identidad de tamaño n .

Notemos que los modelos CAR y SAR son equivalentes si y solo si sus matrices de covarianzas son iguales, es decir, si $(I - C)^{-1} T = (I - B)^{-1} \Lambda (I - B)^{-1T}$.

La forma de las matrices B y C normalmente es especificada por la forma del lattice, sin embargo, una forma de construir las es a partir de un parámetro que escale una matriz de vecindades definida W , la cual indica si las regiones son vecinas o no. Una forma común de hacerlo es definiendo $W = (w_{ij})$ como

$$(w_{ij}) = \begin{cases} 1 & \text{si el sitio } s_i \text{ comparte una arista con el sitio } s_j. \\ 0 & \text{si } i = j. \\ 0 & \text{en otro caso.} \end{cases}$$

Entonces, las matrices C y B , para los modelos CAR y SAR respectivamente, quedan de la forma

$$C = \rho_c W \quad \text{y} \quad B = \rho_s W,$$

en donde los parámetros ρ_c y ρ_s tienen que ser estimados; estos comúnmente son llamados parámetros de correlación espacial.

1.6 APROXIMACIÓN DE LAPLACE ANIDADA INTEGRADA, INLA

En esta sección se presenta el enfoque INLA, el cual será usado más adelante para las aplicaciones en esta tesis. Este método es usado para aproximar las marginales de la posterior del campo gaussiano $\pi(x_i|y)$, $i \in \{1, \dots, n\}$. Tal y como lo describen Rue y Martino (2009):

Primero, se aproxima la marginal posterior de θ usando la aproximación de Laplace. Después, se calcula la aproximación de Laplace, o la aproximación simplificada de Laplace, de $\pi(x_i|y, \theta)$, para los valores seleccionados de θ , para mejorar la aproximación gaussiana. Por último, mediante la integración numérica, se combinan los dos pasos anteriores.

Las marginales posteriores pueden ser escritas como

$$\begin{aligned}\pi(x_i|y) &= \int \pi(x_i|\theta, y)\pi(\theta|y)d\theta, \\ \pi(\theta_j|y) &= \int \pi(\theta|y)d\theta_{-j},\end{aligned}$$

y la característica clave de nuestro enfoque es usar este formulario para construir aproximaciones anidadas

$$\begin{aligned}\tilde{\pi}(x_i|y) &= \int \tilde{\pi}(x_i|\theta, y)\tilde{\pi}(\theta|y)d\theta, \\ \tilde{\pi}(\theta_j|y) &= \int \tilde{\pi}(\theta|y)d\theta_{-j},\end{aligned}$$

en donde, $\tilde{\pi}(\cdot|\cdot)$ es una densidad condicional aproximada. Las aproximaciones de $\pi(x_i|y)$ son calculadas aproximando $\pi(\theta|y)$ y $\pi(x_i|\theta, y)$, y usando integración numérica (suma finita) para integrar hacia afuera θ . El enfoque con aproximaciones de Laplace

se hace muy preciso cuando lo usamos para modelos gaussianos latentes.

Supongamos que tenemos la siguiente aproximación $\tilde{\pi}(\theta|y)$ de la marginal posterior de θ :

$$\tilde{\pi}(\theta|y) \propto \frac{\pi(x, \theta, y)}{\tilde{\pi}_G(x|\theta, y)} \Big|_{x=x^*(\theta)},$$

en donde $\tilde{\pi}_G(x|\theta, y)$ es la aproximación gaussiana del condicional completo de x , y $x^*(\theta)$ es el modelo del condicional completo de x , para una θ dada. La proporcionalidad de la expresión anterior, viene del hecho de que la constante normalizada $\pi(x, \theta|y)$ es desconocida; esta expresión es equivalente a la aproximación de la distribución marginal posterior de Laplace que sugiere que el error de aproximación es relativo y de orden $\mathcal{O}(n_d^{-\frac{3}{2}})$ después de renormalizar.

Consideremos a $\theta = (\theta_1, \dots, \theta_m) \in \mathcal{R}^m$, la cual siempre puede ser obtenida por reparametrización. Justo como lo proponen Rue y Martino (2009) se pueden seguir 3 pasos:

1. Paso 1: ubicar el modo de $\tilde{\pi}(\theta|y)$, optimizando el $\log(\pi(\tilde{\theta}|y))$ con respecto a θ . Esto puede ser realizado utilizando algún método de Newton que construya una aproximación de la segunda derivada del $\log(\tilde{\pi}(\theta|y))$ usando la diferencia entre los gradientes sucesivos. El gradiente es aproximado usando diferencias finitas. Sea θ^* la configuración modal.
2. Paso 2: cuando la configuración modal θ^* calcule la matriz Hessiana negativa $H > 0$, usando diferencias finitas. Sea $\Sigma = H^{-1}$, la cual sería la matriz de covarianzas para θ si la densidad fuera gaussiana. Para ayudar al proceso explorativo, usemos variables estandarizadas z en lugar de θ . Sea $\Sigma = V\Lambda V^T$ la eigendescomposición de Σ , y definamos θ via z , de la siguiente manera:

$$\theta(z) = \theta^* + V\Lambda^{\frac{1}{2}}z.$$

Si $\tilde{\pi}(\theta|y)$ es una distribución gaussiana, entonces z es $\mathcal{N}(0, 1)$. Esta reparametrización corrige la escala y rota-

ción, y simplifica la integración numérica.

3. Probando $\log(\tilde{\pi}(\theta|y))$ usando la z -parametrización. La figura 2.2 dibuja el procedimiento cuando $\log(\tilde{\pi}(\theta|y))$ es unimodal. (a) muestra un diagrama de $\log(\tilde{\pi}(\theta|y))$ para $m = 2$, la ubicación de la moda y el nuevo eje de coordenadas para z . Queremos explorar $\log(\tilde{\pi}(\theta|y))$ para localizar la mayor parte de la masa de probabilidad. El resultado de este procedimiento es presentado en (b), en donde cada punto es un punto donde el $\log(\tilde{\pi}(\theta|y))$ es considerado significativo y usado en la integración numérica

$$\tilde{\pi}(x_i|y) = \sum_k \tilde{\pi}(x_i|\theta_k, y) \tilde{\pi}(\theta_k|y) \Delta_k.$$

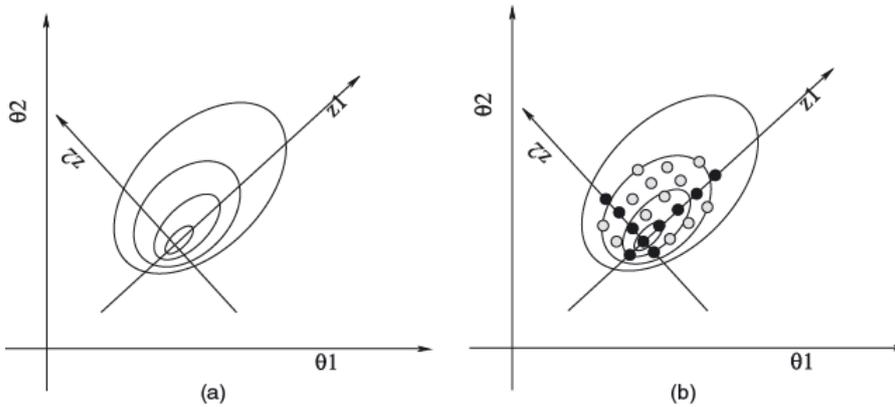


Figura 2: Gráfica de la exploración de la marginal posterior de θ . En (a) la moda está en el Hessiano y el sistema co-ordenado para z calculado; en (b) cada dirección coordinada es explorada por (\bullet) hasta que la log-densidad cae por debajo de cierto de límite; por último los puntos en gris son explorados. (Obtenida de Rue y Martino, 2009).

4. Aproximando $\pi(\theta_j|y)$: las marginales posteriores para θ_j puede ser obtenida directamente de $\tilde{\pi}(\theta|y)$ usando integración numérica. Este procedimiento, computacionalmente es exigente pues necesitamos evaluar $\tilde{\pi}(\theta|y)$ para un gran número de configuraciones, por lo que un enfoque más factible es usar los puntos que ya se calcularon durante los pasos 1–3 para construir un interpolante de $\log(\tilde{\pi}(\theta|y))$ y para calcular marginales utilizando la integración numérica de este interpolante.

Aproximando $\pi(x_i|\theta, y)$

La aproximación más sencilla de $\pi(x_i|\theta, y)$ es la aproximación gaussiana $\tilde{\pi}_G(x_i|\theta, y)$, donde la media $\mu_i(\theta)$ y la varianza marginal σ_i^2 son derivadas de usar la recursión

$$\sum_{ij} = \left(\frac{\delta_{ij}^2}{L_{ii}} - \frac{1}{L_{ii}} \right) \left(\sum_{k=i+1}^n L_{ki} \right) \Sigma_{kj}, \quad j \geq i, i = n \dots 1,$$

en donde $\Sigma (= Q^{-1})$ es la matriz de covarianzas, y $\delta_{ij} = 1$ si $i = j$ y $\delta_{ij} = 0$ en otro caso.

Mientras se exploró $\tilde{\pi}(\theta|y)$, ya se había calculado $\tilde{\pi}_g(x|\theta)$, así que solo las variaciones marginales deben ser adicionalmente calculadas. La aproximación gaussiana a menudo da resultados razonables, pero puede tener errores en la ubicación y/o errores debido a la falta de asimetría.

La forma natural de mejorar la aproximación gaussiana es calcular la aproximación de Laplace

$$\tilde{\pi}_{LA}(x_i|\theta, y) \propto \frac{\pi(x, \theta, y)}{\tilde{\pi}_{GG}(x_{-i}|x_i, \theta, y)} \Big|_{x_{-i}=x_{-i}^*(x_i, \theta)},$$

en donde $\tilde{\pi}_{GG}$ es la aproximación gaussiana de $x_{-i}|x_i, \theta, y$ y $x_{-i}^*(x_i, \theta)$ es la configuración modal. Vease, además, que $\tilde{\pi}_{GG}$ es diferente de la densidad condicional correspondiente a $\tilde{\pi}_G(x|\theta, y)$.

Dado que la expresión anterior implica que $\tilde{\pi}_{GG}$ debe ser recalculada para cada valor de x_i y θ , pues su matriz de precisión depende de x_i y θ , y al ser esto muy costoso, pues se requieren n factorizaciones de la matriz de precisión, hay dos formas que permiten que esto sea más factible computacionalmente.

La primera, consiste en evitar el paso de optimización informática $\tilde{\pi}_{GG}(x_{-i}|x_i, \theta, y)$ aproximando la configuración modal

$$x_{-i}^*(x_i, \theta) \approx E_{\tilde{\pi}_G}(x_{-i}|x_i).$$

El lado derecho se evalúa bajo la densidad condicional que se deriva de la aproximación gaussiana $\tilde{\pi}_G(x|\theta, y)$. Veamos que la media condicional se puede calcular mediante una actualización de rango 1 a partir de la media incondicional, usando la muestra del campo aleatorio de Markov

$$x^c = x - Q^{-1}A^T(AQ^{-1}A^T)^{-1}(Ax - e).$$

En el caso espacial el costo es $\mathcal{O}(n \log(n))$, para cada i , que proviene de resolver $Q^*(\theta)v = 1_i$, en donde 1_i es igual a 1 en la posición i y 0 en otro caso. Esta actualización de rango 1 se calcula solo una vez para cada i , ya que es lineal en x_i . Aunque sus configuraciones son ligeramente diferentes, Hsiao et al. (2004) mostró que desviarse del modo condicional no necesariamente degrada el error de aproximación. Otra característica positiva de la aproximación es que la media condicional es continua con respecto a x_i , que no lo es cuando se utiliza la optimización numérica para calcular $x_{-i}^*(x_i, \theta)$.

La siguiente modificación materializa la siguiente intuición: solo aquellos x_j que están cercanos a x_i deberían tener un efecto en la marginal de x_i . Si la dependencia entre x_j y x_i decae a medida que la distancia entre los nodos i y j aumenta, solo aquellos x_j en una región de interés alrededor de i , $R_i(\theta)$, determinan el marginal de x_i . La esperanza condicional en la aproximación $x_{-i}^*(x_i, \theta) \approx E_{\tilde{\pi}_G}(x_{-i}|x_i)$ implica que

$$\frac{E_{\tilde{\pi}_G}(x_j|x_i) - \mu_j(\theta)}{\sigma_j(\theta)} = a_{ij}(\theta) \frac{x_i - \mu_i(\theta)}{\sigma_i(\theta)},$$

para algún $a_{ij}(\theta)$ cuando $j \neq i$. Por lo tanto, una regla simple para construir el conjunto $R_i(\theta)$ es

$$R_i(\theta) = \{j : |a_{ij}(\theta)| > 0,001\}.$$

El ahorro computacional más importante usando $R_i(\theta)$ proviene del cálculo del denominador de la aproximación de Laplace, en donde sabemos que tenemos que factorizar una matriz de dispersión $|R_i(\theta)| \times |R_i(\theta)|$.

La aproximación de Laplace, simplificada como se explicó anteriormente, debe de calcularse para diferentes valores de x_i para encontrar la densidad, Para seleccionar estos puntos, usamos la media y la varianza de la aproximación gaussiana $\tilde{\pi}(x_i|\theta, y) = \mathcal{N}\{x_i; \mu_i(\theta), \sigma_i^2(\theta)\}$, y elegimos diferentes valores para la variable estandarizada

$$x_i^s = \frac{x_i - \mu_i(\theta)}{\sigma_i(\theta)}$$

según la elección correspondiente de abscisas dada por la cuadratura de Gauss-Hermite. Para representar la densidad $\tilde{\pi}_{LA}(x_i|\theta, y)$, usamos que

$$\tilde{\pi}_{LA}(x_i|\theta, y) \propto \mathcal{N}\{x_i; \mu_i(\theta), \sigma_i^2(\theta)\} \exp\{\text{spline cúbico}(x_i)\}.$$

El spline cúbico se ajusta a la diferencia de la log-densidad de $\tilde{\pi}_{LA}(x_i|\theta, y)$ y $\tilde{\pi}_G(x_i|\theta, y)$ a los puntos de abscisa seleccionados, y luego la densidad se normaliza utilizando la integración en cuadratura.

ENFERMEDADES SIN VECTOR DE TRANSMISIÓN

Según la Organización Mundial de la Salud (OMS) las enfermedades no transmisibles (ENT) son un grupo de afectaciones que no son principalmente originadas por infección aguda. Generalmente resultan de la combinación de factores de comportamiento y conducta, del entorno ambiental, genéticos y fisiológicos.

Los principales tipos de esta clase de enfermedades son el cáncer, las enfermedades respiratorias crónicas, la diabetes y las enfermedades cardiovasculares, siendo estas últimas la principal causa de muerte alrededor del mundo. Las ENT dan como resultado consecuencias para la salud y con frecuencia crean una necesidad de tratamiento y cuidados a largo plazo (OPS, 2018).

En términos epidemiológicos se puede decir que debido a la evolución que existe en las enfermedades no transmisibles hay una transición epidemiológica de enfermedades infecciosas a ENT. El perfil epidemiológico de la población se ha movido de las enfermedades infecciosas hacia padecimientos degenerativos y causados por el hombre como las principales causas de enfermedades, discapacidad y muerte.

Los cambios en el perfil de la salud y enfermedades de una población, sumados a la transición demográfica y económica que el mundo vive, impactan directamente a la demanda de servicios de salud y la manera en que estos deben transformarse y adecuarse para atender las nuevas necesidades de salud de la población.

Todas las regiones y países se ven afectados por las ENT. También, todos los grupos de edad, sin embargo, las cifras indican que de las muertes atribuidas a ENT 15 millones se producen entre los 30 y los 69 años de edad. Más del 85 % de estas muertes ocurren en países de bajos y medianos ingresos. Las ENT se ven propiciadas por modos de vida poco saludables, el envejecimiento de la población, causas de riesgo metabólico y por otros

factores como la urbanización rápida y no planificada (OMS, 2018).

Las ENT desde hace más de veinte años en México han tenido una gran relevancia en la mortalidad. Las muertes a causa de ENT han ido en aumento conforme pasan los años. Por ejemplo, en la década de los noventa del total de muertes 57% correspondió a este tipo de enfermedades, y para 2016 el 80% del total de las muertes de todas las edades correspondió a ENT (Global Burden Disease Collaborative Network, 2017).

Si bien los determinantes de cada una de las enfermedades no transmisibles son distintos, sí hay elementos que favorecen a la aparición de estas enfermedades. En la figura 3 se pueden observar los determinantes biológicos, de conducta, ambientales y de escala mundial que propician algunas ENT como: enfermedades cardiovasculares, diabetes, cáncer y enfermedades respiratorias.

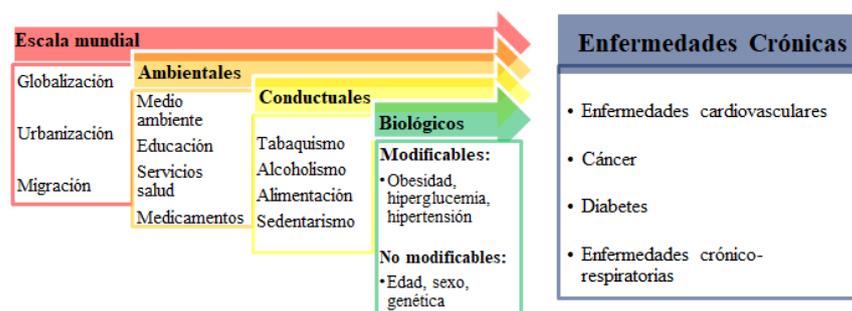


Figura 3: Determinantes de enfermedades no transmisibles (Tomado de la Secretaría de Salud, 2018).

Las enfermedades cardiovasculares son de principal interés para la presente tesis y dentro de los factores de riesgo para las ENT hay algunos que potencian el riesgo para desarrollar alguna enfermedad cardiovascular. Estos son: el tabaquismo, las dietas altas en grasas y sal, la obesidad, la inactividad física, el consumo nocivo de alcohol, la hipertensión arterial, la diabetes y la hiperlipidemia. En la figura 4 se muestran la prevalencia de los factores de riesgo para las enfermedades cardiovasculares en México en 2018 para personas adultas.

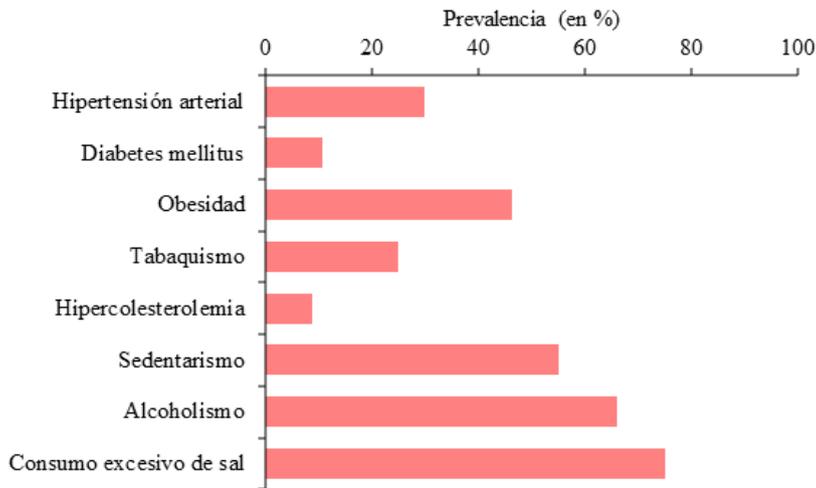


Figura 4: Prevalencia de factores de riesgo cardiovascular en México, 2018, para personas mayores de 20 años. (Secretaría de Salud, 2010).

Según el Sistema Nacional de Salud de México dos de los mayores problemas de salud pública son la diabetes mellitus y la cardiopatía isquémica. Y de acuerdo con el Instituto Mexicano del Seguro Social (IMSS) las enfermedades cardiovasculares son la principal causa de muerte en México, hecho respaldado por la Dirección General de Epidemiología que señala a las enfermedades cardiovasculares como la principal causa de muerte en la población mexicana (Instituto Nacional de Estadística, Geografía e Informática, 2011).

En la figura 5 se puede ver el desglose de las 10 principales razones de muerte a nivel mundial en 2016. Se puede ver que las enfermedades cardiovasculares, en específico la enfermedad isquémica del corazón, son la principal causa de muerte. El caso de México se puede observar en la figura 6, en donde al igual que en el resto del mundo las enfermedades cardiovasculares son la principal causa de defunción.

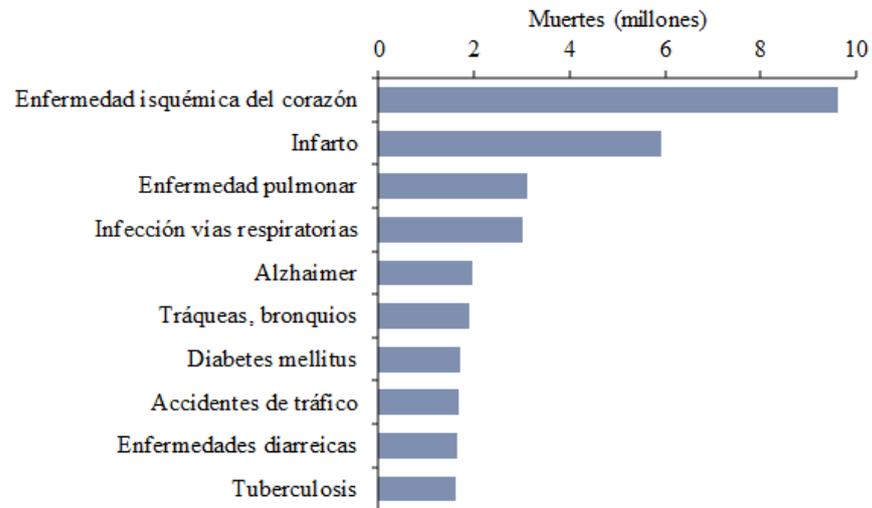


Figura 5: Diez principales causas de muerte en el mundo en 2016 (Cifras tomadas de la OMS, 2018).

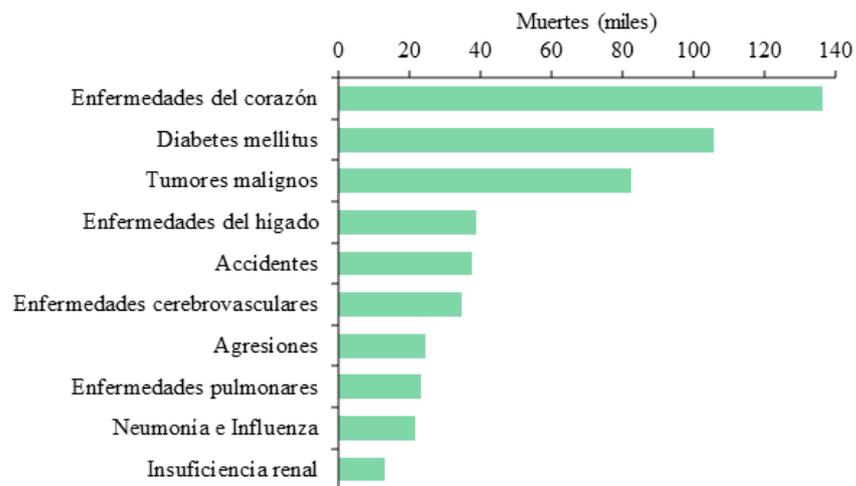


Figura 6: Diez principales causas de muerte en México en 2016 (Cifras tomadas de la Secretaría de Salud, 2018).

Uno de los indicadores más importantes para evaluar la eficacia de los sistemas de salud de los países es la cifra anual de muertes y su causa, estas estadísticas ayudan a las autoridades competentes a decidir sobre las actividades futuras en materia de salud pública. Por ejemplo, cuando la mortalidad por diabetes y enfermedades cardiovasculares se dispara en un periodo corto de tiempo, es muy importante poner en marcha un programa intenso para impulsar formas de vida que ayuden

a prevenir estas enfermedades, del mismo modo tomar la decisión a tiempo de aumentar el gasto en esta esfera con el objetivo de combatir e informar sobre dichas enfermedades (Secretaría de Salud, 2018).

Los países económicamente más desarrollados y de ingresos altos cuentan con sistemas que permiten recolectar información sobre las causas de muerte; no así los países con ingresos medianos y bajos. En estos países sin posibilidad de recoger datos, se necesita estimar el número de muertes por causa a partir de datos incompletos (OMS, 2018), tal y como se hace en este trabajo.

APLICACIÓN DEL MODELO PREDICTIVO PARA DATOS DE ÁREA

En México y en el mundo uno de los principales problemas de salud son las enfermedades cardiovasculares, ya que son la causa principal de muerte y un generador de diversos problemas sociales y económicos. Esto lleva a la necesidad de realizar diversas acciones y estudios que faciliten la toma de decisiones con respecto a la organización, diseño y valoración de proyectos sociales y de salud que permitan combatir el problema con base en técnicas y metodologías estadísticas. Se realizó este trabajo municipal ya que se busca contar con el mayor nivel de desagregación posible. En palabras del INEGI (2020)

«La demanda creciente de la sociedad para tener respuestas satisfactorias a sus necesidades de información por parte de las oficinas nacionales de estadística, se ha convertido en el transcurso del tiempo en una constante universal. Particularmente, los gobiernos locales requieren contar con información actualizada y desagregada para niveles geográficos pequeños».

Se busca estimar el riesgo relativo de padecer alguna enfermedad cardiovascular en todos los municipios de México mediante el modelo de Besag (1991) con aproximaciones de Laplace anidada integrada (INLA). También, se realizaron mapas de riesgo por sexo y grupo de edad. Se utilizaron datos de la Encuesta Nacional de Salud (ENSANUT) 2012 y el índice de desarrollo humano (IDH) por área tomado del censo de 2010 realizado por el Instituto Nacional de Estadística y Geografía a nivel municipal.

3.1 MODELO PREDICTIVO PARA LAS ENFERMEDADES CARDIOVASCULARES EN MÉXICO

La ENSANUT 2012 es una encuesta nacional probabilística con representatividad estatal, por estratos nacionales urbano y rural, cuyo esquema es polietápico y estratificado. La unidad primaria de muestreo son las áreas geoestadísticas básicas (AGEB), que son la unidad censal en México. En cada entidad se formaron siete estratos, seis que resultan de considerar las combinaciones de urbano, rural y metropolitano con mayor rezago y menor rezago, y un estrato para las localidades de nueva creación. En el caso de la Ciudad de México se formaron tres estratos geográficos dentro del estrato metropolitano, dado que todas las viviendas de esta ciudad fueron clasificadas como metropolitanas.

Para implementar la sobre muestra de hogares con mayores carencias se construyó un índice de rezago social para las AGEB. Las AGEB se ordenaron de acuerdo con el valor del índice en cada estado y se definió para cada uno el estrato de mayor rezago social que agrupaba a 20 % de la población. Esta sobre muestra fue considerada en la generación de los ponderadores para asignar el peso apropiado a las observaciones en la encuesta (ENSANUT, 2012).

En la figura 7 se ilustra el mapa de casos calibrados de personas con enfermedades cardiovasculares y se anexa la cantidad de casos que hay en cada municipio observado. Notemos que este no es un mapa de riesgo, sino que solo representa la cantidad de datos muestreados «positivos» con alguna enfermedad cardiovascular sin considerar el tamaño de población total, es decir, no es la proporción de personas con enfermedades cardiovasculares.

El 38 % de los encuestados fueron hombres, 57 % mujeres y del 5 % restante no se sabe su sexo. En la figura 8 se observa el número de casos de personas encuestadas con alguna enfermedad cardiovascular por grupo de edad y sexo en donde se puede ver que en la encuesta, para todas las agrupaciones, hay una mayor prevalencia para mujeres.

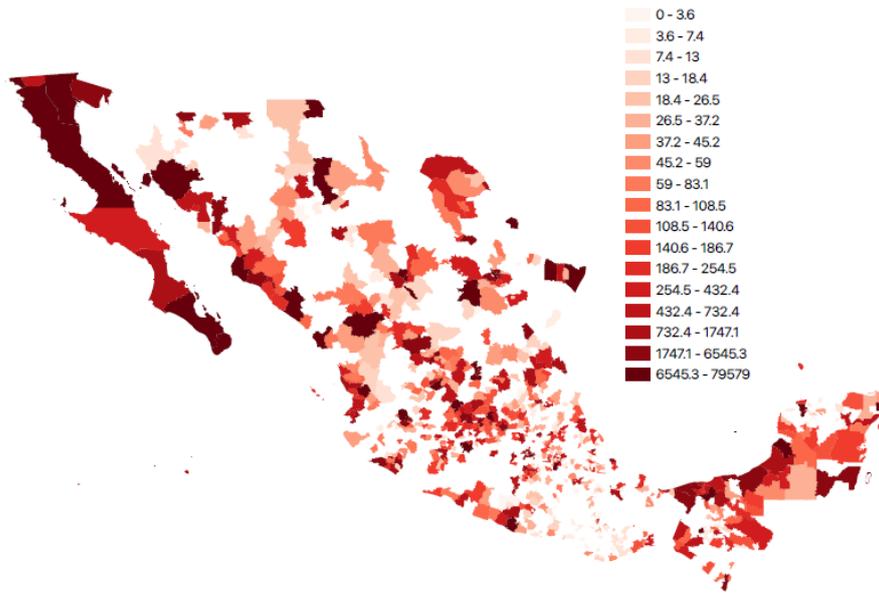


Figura 7: Mapa de México: casos calibrados de personas con enfermedades cardiovasculares, según la ENSANUT 2012.

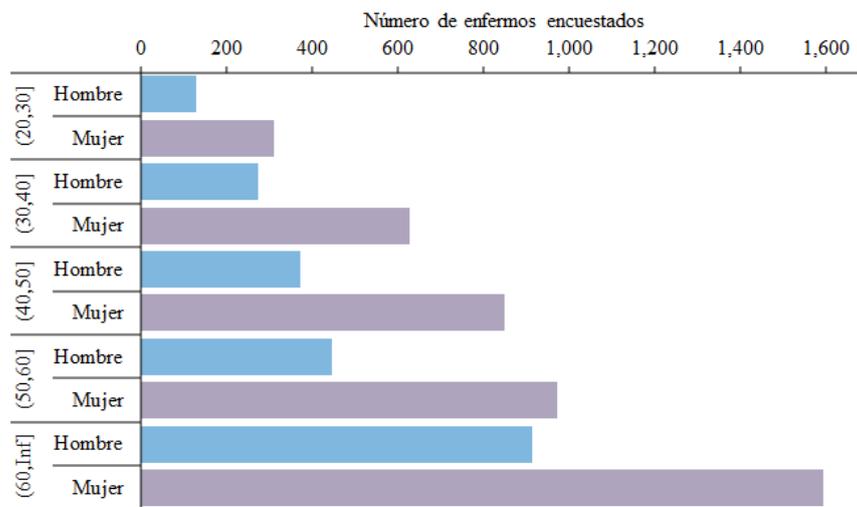


Figura 8: Número de enfermos encuestados observados, por sexo y grupo de edad. (Datos obtenidos de ENSANUT, 2012)

El número observado de casos de enfermedad cardiovascular se calibró utilizando el factor de expansión de cada observación y, como se tiene la intención de producir un análisis de área pequeña, se redujeron los datos disponibles multiplicándolos por

$$p_{\text{mun}} \frac{p_{\text{mun}}}{p_{\text{edo}}} (\text{pondef}),$$

donde:

- p_{mun} : población del área administrativa más pequeña de México, municipio.
- p_{edo} : población del estado.
- pondef : ponderador (ENSANUT, 2012).

La covariable por edad se discretizó en décadas, agrupando a las personas mayores de 60 años en un solo grupo y considerando solo a población adulta de 20 años o más.

Para estimar el riesgo empírico de padecer alguna enfermedad cardiovascular se utilizó el enfoque descrito en Besag et al. (1991).

Sea en el área i ,

- r_i : el riesgo relativo de padecer una enfermedad cardiovascular.
- y_i : el número observado de casos.
- c_i : el número esperado de casos.

$$c_i = \frac{a}{b} * (b * a)$$

- a : población municipal.
- b : población estatal.

Asumiendo el modelo estándar:

$$y_i | r_i \sim \text{Poisson}(c_i e^{r_i})$$

$$r_i = X\beta + u_i + v_i,$$

donde

- u_i : la variabilidad espacial.

- v_i : la variabilidad no estructurada.
- X : es la matriz de covariables, de n variables \times p observaciones, compuesta por la categoría de edad, el índice de desarrollo humano, el sexo y la cantidad de muertes por enfermedad cardiovascular en cada área.
- β : es el vector de coeficientes de la regresión.

La variabilidad espacial u_i y el ruido blanco v_i no son identificables en el modelo, sin embargo, existen y se expresan de dicha forma en la literatura. La distribución a priori se dio siguiendo lo propuesto por Besag et al. (1991).

$$\begin{aligned} \beta &\sim N(0, 100) \\ \pi(\mathbf{u}) &\propto \exp \left\{ - \sum_{i < j} w_{ij} \phi(u_i - u_j) \right\} \\ v &\sim N(0, 100), \end{aligned}$$

donde λ es una constante positiva, ϕ es una función par dada que aumenta con $|u_i - u_j|$ y w_{ij} son pesos positivos con $w_{ij} = 0$ si i, j no son aledaños y $w_{ij} = 1$ si i, j son vecinos.

El modelo de Besag (1991) se ajustó utilizando aproximaciones de Laplace anidadas integradas, INLA. Por mucho tiempo, la inferencia bayesiana se basó en los Métodos de Cadena de Markov Montecarlo para calcular la distribución posterior conjunta de los parámetros del modelo, lo cual no suele ser computacionalmente rápido (Gilks et al. 1996; Brooks et al. 2011).

Con INLA se propone un enfoque diferente ya que en lugar de estimar la distribución posterior conjunta de los parámetros sugiere centrarse en las marginales posteriores individuales de los parámetros del modelo (Havard et al. 2009). Luego, el enfoque se centra en modelos que pueden expresarse como GMRF lo cual proporciona ventajas computacionales que reducen el tiempo de cálculo del ajuste del modelo (ver Rue y Held 2005). Esto se hace aproximando las distribuciones marginales posteriores de los parámetros del modelo basándose en la aproximación de Laplace.

Sea $\mathbf{y} = (y_1, \dots, y_n)$ un vector de n observaciones, en donde algunas de ellas puede ser faltantes y tienen una probabilidad

asociada (no necesariamente de la familia exponencial). En general, la media μ_i de la observación y_i puede ser vinculada al predictor lineal η_i mediante una función conveniente. Las observaciones serán independientes dado su predictor lineal, es decir,

$$\eta_i = \alpha + \sum_{j=1}^{\eta_\beta} \beta_j z_{ji} + \sum_{k=1}^{\eta_f} f^{(k)}(u_{k_i}) + \epsilon_i; i = 1, \dots, n.$$

Aquí, α es el intercepto, β_j con $j = 1, \dots, n_\beta$, son coeficientes de algunas covariables $\{z_j\}_{j=1}^{n_\beta}$, las funciones $f^{(k)}$ definen n_f efectos aleatorios sobre algún vector de covariables $\{u_k\}_{k=1}^{n_f}$. Finalmente, ϵ_i es un término de error, que puede faltar dependiendo de la probabilidad.

En datos espaciales INLA es computacionalmente eficiente porque usa una ecuación de diferenciación parcial estocástica para estimar la autocorrelación espacial de los datos. Esto implica el uso de una “malla” de ubicaciones de muestreo discretas que se interpolan para estimar un proceso continuo en el espacio (Havard et al. 2009).

3.2 RESULTADOS

El cuadro 1 muestra las estadísticas de resumen para las estimaciones de parámetros del componente de regresión del modelo. Se puede deducir de los pequeños valores de la divergencia Kullback-Leibler (kld), entre los valores estimados y esperados, que las aproximaciones gaussianas proporcionan un buen ajuste para el modelo. Y observando las pruebas de hipótesis en el mismo cuadro y en la figura 9 vemos que las covariables de sexo y edad son significativas para el modelo.

Cuadro 1: Distribución posterior para los coeficientes de la regresión.

	media	ds	0.025	0.5	0.975	kld
Intercepto	-1.8200	0.4270	-2.6600	-1.8200	-0.9840	0.0000
IDH	-1.0400	0.6400	-2.3000	-1.0500	0.2120	0.0000
Sexo Mujer	0.4820	0.0020	0.4780	0.4820	0.4860	0.0000
Edad 30-40	0.3310	0.0040	0.3230	0.3310	0.3390	0.0000
Edad 40-50	0.7100	0.0037	0.7020	0.7100	0.7170	0.0000
Edad 50-60	0.8020	0.0037	0.7940	0.8020	0.8090	0.0000
Edad 60+	1.2400	0.0035	1.2400	1.2400	1.2500	0.0000
Defunciones	-0.0007	0.0002	-0.0012	-0.0007	0.0002	0.0000

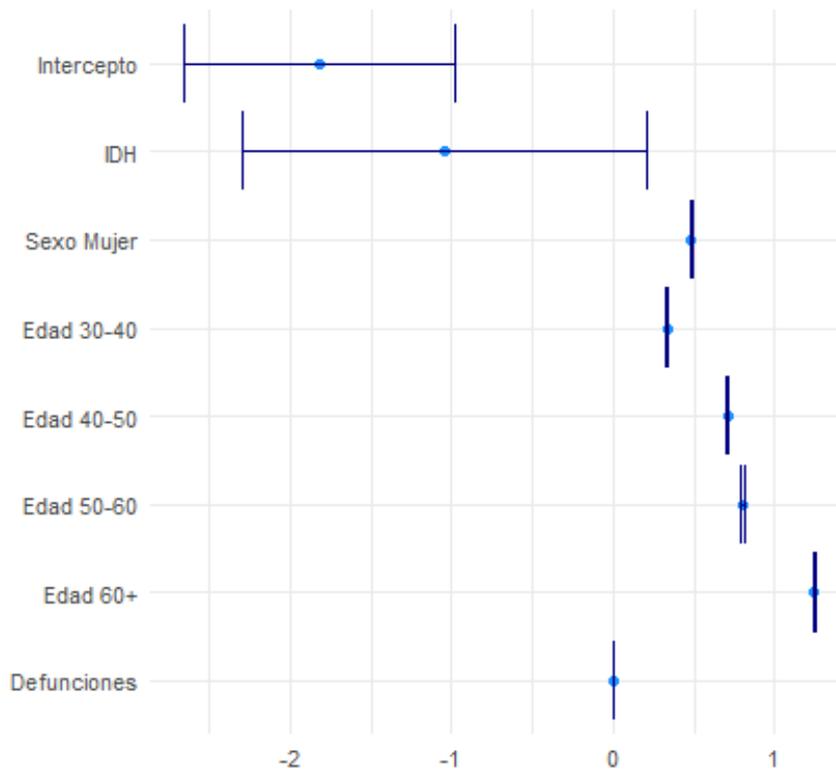


Figura 9: Cuantiles de 2.5 %, 97.5 % y estimación puntual media para las distribuciones posteriores de los parámetros de las covariables.

Además de estos resultados, se eligió el modelo de Besag sobre el modelo de Leroux y el IID debido a los mejores resultados en CPO y DIC. En el cuadro 2 se presenta el resumen de los criterios de selección del modelo.

Cuadro 2: Resumen de los criterios de selección del modelo.

Modelo	CPO	DIC
Besag	184,924.3	510,135.6
Leroux	184,929.5	510,137
IID	224,301.1	742,251.9

La interpretación de cada uno de los coeficientes del cuadro 1 varía dependiendo del tipo de la covariable. Por ejemplo, la covariable de sexo es dicotómica y la de edad es categórica. Entonces, para el IDH o las defunciones que no son de dicha forma: si el IDH o las defunciones aumentan en una unidad, en cualquier municipio, el riesgo de padecer alguna enfermedad cardiovascular aumenta o disminuye en un factor de $\exp\{\beta_{IDH}\}$. Es decir, si el IDH aumenta en una unidad el riesgo de padecer una enfermedad cardiovascular disminuye en un factor de $\exp\{-1.04\} = 0.35$.

Para la variable dicotómica sexo se tiene lo siguiente: ser mujer aumenta o disminuye el riesgo en un factor de $\exp\{\beta_{sexo}\}$ con respecto a ser hombre, es decir, con los datos recolectados y bajo el modelo realizado, en general para cualquier municipio ser mujer representa un aumento de riesgo en un factor de $\exp\{0.482\} = 1.6193$ con respecto a ser hombre.

En cuanto a la variable categórica edad el riesgo de pertenecer a cierto grupo de edad aumenta o disminuye en un factor de $\exp\{\beta_{edad}\}$ con respecto a pertenecer al grupo de las personas de 20 a 30 años. Esto es, el riesgo de padecer una enfermedad cardiovascular aumenta con respecto a pertenecer al grupo de las personas de 20 a 30 años en un factor de

- 30-40 años: $\exp\{0.3310\} = 1.3924$
- 40-50 años: $\exp\{0.7100\} = 2.0339$
- 50-60 años: $\exp\{0.8020\} = 2.2299$
- Mayores de 60 años: $\exp\{1.2400\} = 3.4556$

Ahora bien, si uno quisiera comparar el riesgo de una categoría respecto a alguna otra que no sea el grupo de 20 a 30 años ,

por ejemplo: ¿en cuánto aumenta el riesgo por pasar del grupo de 40 a 50 años al grupo de los mayores de 60? En este caso se deben tomar en cuenta los dos coeficientes de las categorías que se buscan comparar. Entonces, el riesgo de pertenecer al grupo de 60 ó más años con respecto al grupo de 40 a 50 aumenta en un factor de

$$\begin{aligned} \exp\{\beta_{60+} - \beta_{40-50}\} &= \exp\{1.2400 - 0.7100\} \\ &= 1.6989. \end{aligned}$$

Lo mismo puede hacerse para generar la comparación entre dos categorías de cualquier covariable, y así saber el aumento o disminución del riesgo de un grupo respecto a otro. Esta forma de interpretar los coeficientes de las covariables de la distribución posterior pueden ser consultadas en Colin y Trivedi (1998).

Mapas de riesgo

Dado que ya tenemos la interpretación de los coeficientes de distribución posterior, es interesante observar los resultados de forma gráfica y más intuitiva, por ejemplo en un mapa. Un mapa es una representación gráfica simplificada de un territorio con propiedades métricas sobre una superficie bidimensional, que en este caso es plana. Estos mapas nos van a permitir observar, además del efecto de los resultados vistos anteriormente, el riesgo relativo de padecer alguna enfermedad cardiovascular en todos los municipios de México para hombres y mujeres de cualquier grupo de edad.

Para lo anterior se realizaron diez mapas de riesgo de la República Mexicana (figuras 10 a 19) con su respectiva división política a nivel municipal en donde podemos ver el riesgo de padecer alguna enfermedad cardiovascular en cada municipio. Estos mapas están hechos para cada sexo y grupo de edad, es decir, hay cinco mapas para mujeres y 5 mapas para hombres, uno por cada grupo de edad. Esto nos permitirá, además de observar las zonas con mayor riesgo de padecer una enfermedad, hacer la comparación del mismo para cada sexo-edad.

Todos los mapas se construyeron en una escala del 0 (azul oscuro) al 20 (amarillo). Los municipios cuyo factor de aumento al riesgo es menor a 1 quiere decir que son lugares en

donde el riesgo estimado fue menor al número esperado de casos con enfermedad. Por lo tanto, los municipios que aumentan su riesgo relativo en un factor mayor a 1 son aquellos espacios en donde el riesgo es mayor al número esperado de casos. Todos los municipios de la República Mexicana tienen un riesgo asociado para cada grupo de edad y sexo. Lo esperado es que se puedan encontrar patrones en ciertas zonas que dicten que hay un riesgo similar entre municipios y estados vecinos. En los mapas es posible notar que sí existe correlación espacial en las zonas de México: noroeste, noreste, centro, occidente y sureste.

Noroeste

Esta región está constituida por los estados de Baja California, Baja California Sur, Chihuahua, Sinaloa y Sonora. Observemos que, en todos los mapas, los municipios de los estados de Baja California, Baja California Sur y Sinaloa tienen un riesgo relativo menor a 1 y que es muy similar entre sí. Estos estados, además, son de los que cuentan con municipios con menor riesgo para las dos covariables.

Por otro lado, los estados de Sonora y Chihuahua comparten tendencias en el riesgo de sus entidades municipales. Sin embargo, a diferencia de los otros estados de la zona son entidades de mayor riesgo, aunque su factor de aumento al riesgo relativo sigue estando por debajo de 1.

Noreste

Esta zona está formada por las entidades de Coahuila, Durango, Nuevo León, San Luis Potosí y Tamaulipas. Vemos que los municipios de todos estos estados muestran un riesgo similar en cada uno de los mapas para las dos covariables; en general un riesgo relativo menor a 1.

Hay municipios que no presentan la misma tendencia que el estado en general, pero sí lo hacen con respecto a los municipios vecinos más cercanos y/o los que los rodean, por lo cual sí es posible ver la correlación espacial entre ellos. En el caso de estos municipios hay algunos que tienen un riesgo relativo cercano a 1 ó incluso mayor a 1.

Occidente

Este espacio está conformado por Aguascalientes, Colima, Guanajuato, Jalisco, Michoacán, Nayarit, Querétaro y Zacatecas, y de igual forma se puede observar correlación espacial entre sus municipios. Estos estados tienen en general un factor de decrecimiento del riesgo relativo menor a 1 para todos los grupos de edad y ambos sexos, con excepción de algunos municipios.

Centro

Este es el territorio formado por la Ciudad de México, Estado de México, Guerrero, Hidalgo, Morelos, Puebla y Tlaxcala, también se puede ver que existe correlación espacial entre los municipios de estas entidades ya que el riesgo es parecido en ellos. El riesgo relativo de estas entidades es cercano a 1 para la mayoría de grupos de edad y sexos, con algunos municipios con un riesgo mayor a 1.

Algunos municipios del sureste del estado de Guerrero presentan mayor riesgo (factor mayor a 1) en general que la zona centro. Sin embargo, tienen un riesgo similar a los municipios del estado Oaxaca, el cual es un estado vecino. De igual forma, Puebla y Veracruz tienen una tendencia de mayor riesgo, y de la misma manera, son estados vecinos de Oaxaca y Guerrero.

Sureste

Esta zona constituida por los estados de Campeche, Chiapas, Oaxaca, Quintana Roo, Tabasco, Veracruz y Yucatán presenta patrones que indican correlación espacial. Para los estados pegados al océano Pacífico, Guerrero y Chiapas observamos que son las entidades con municipios de mayor riesgo en el país, ya que la mayoría de sus municipios tienen un riesgo relativo mayor a 1 para la mayoría de grupos de edad y ambos sexos.

Por otro lado, los municipios de los estados de Tabasco, Campeche y Quintana Roo, presentan un riesgo similar, que es bajo en comparación de la zona (factor menor 1 en casi todos los casos), mientras que los municipios del estado de Yucatán tienden a ser de mayor riesgo (cerca de 1), pero sin llegar a los niveles

mostrados por Guerrero y Chiapas.

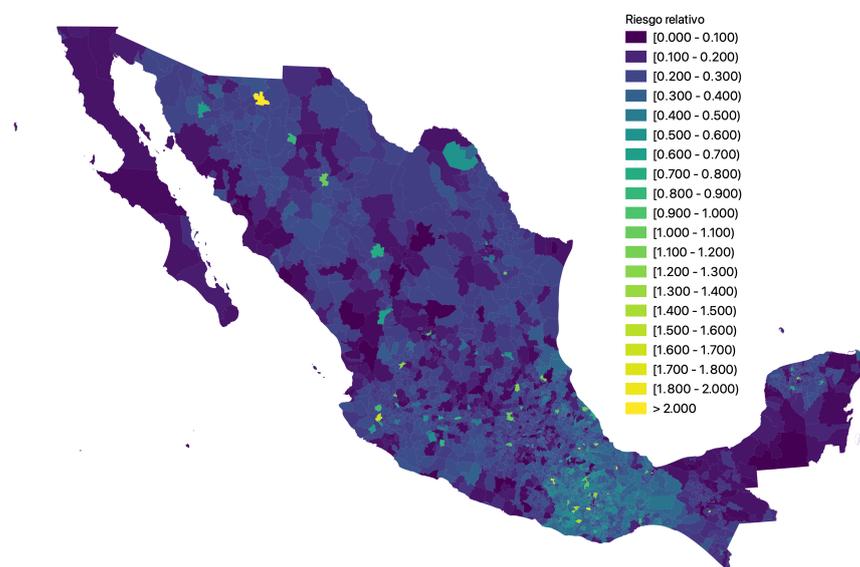


Figura 10: Mapa de México 1: riesgo relativo de enfermedades cardiovasculares por unidad de área, para mujeres de 20 a 30 años.

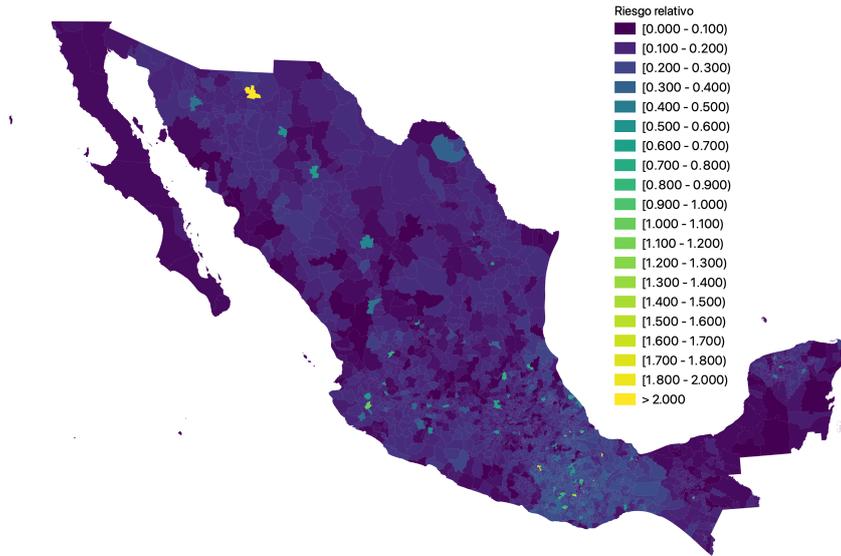


Figura 11: Mapa de México 2: riesgo relativo de enfermedades cardiovasculares por unidad de área, para hombres de 20 a 30 años.

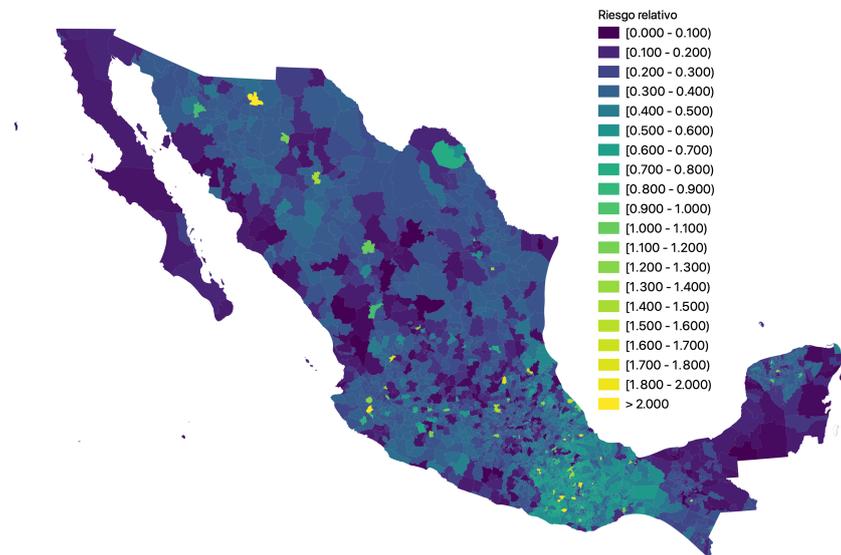


Figura 12: Mapa de México 3: riesgo relativo de enfermedades cardiovasculares por unidad de área, para mujeres de 31 a 40 años.

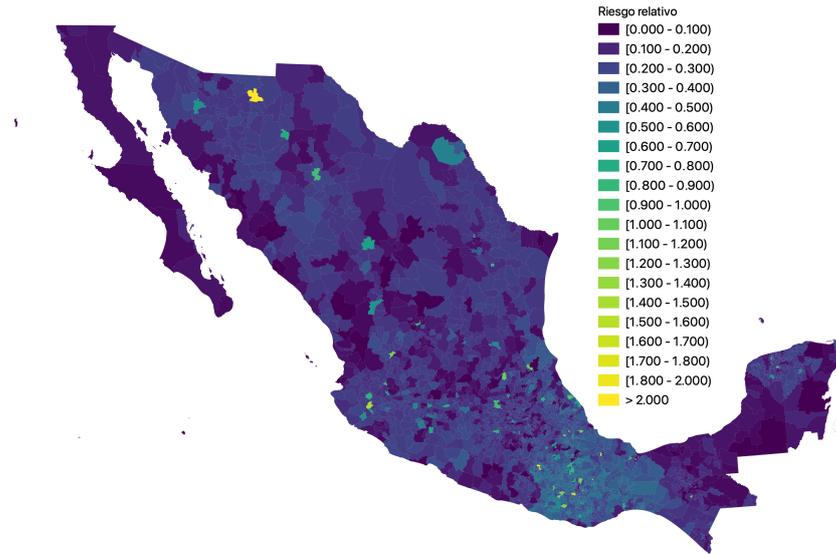


Figura 13: Mapa de México 4: riesgo relativo de enfermedades cardiovasculares por unidad de área, para hombres de 31 a 40 años.

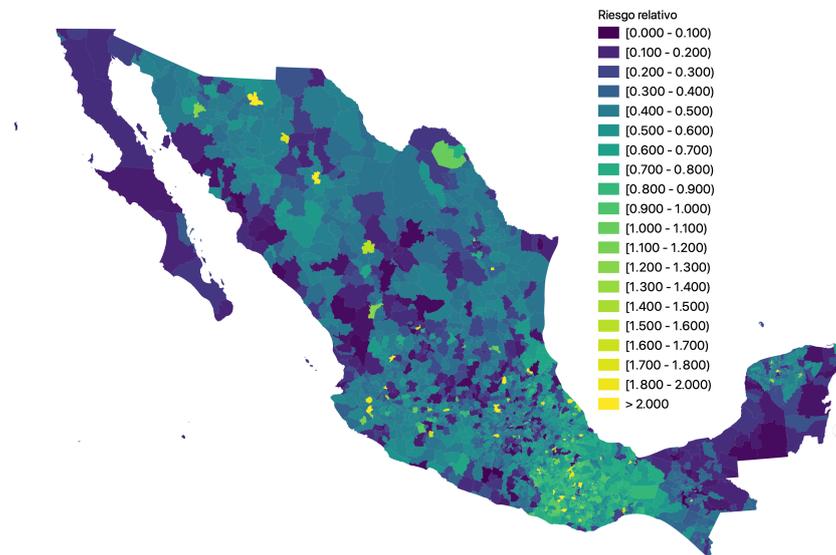


Figura 14: Mapa de México 5: riesgo relativo de enfermedades cardiovasculares por unidad de área, para mujeres de 41 a 50 años.

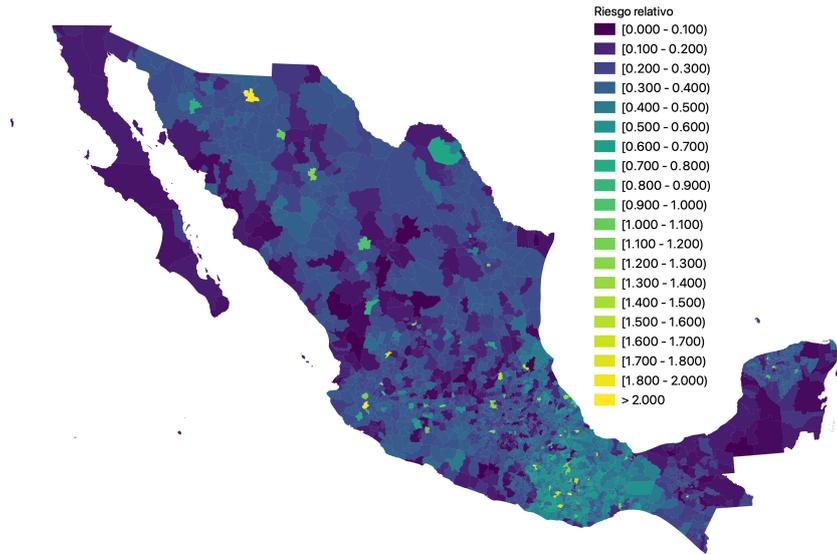


Figura 15: Mapa de México 6: riesgo relativo de enfermedades cardiovasculares por unidad de área, para hombres de 41 a 50 años.

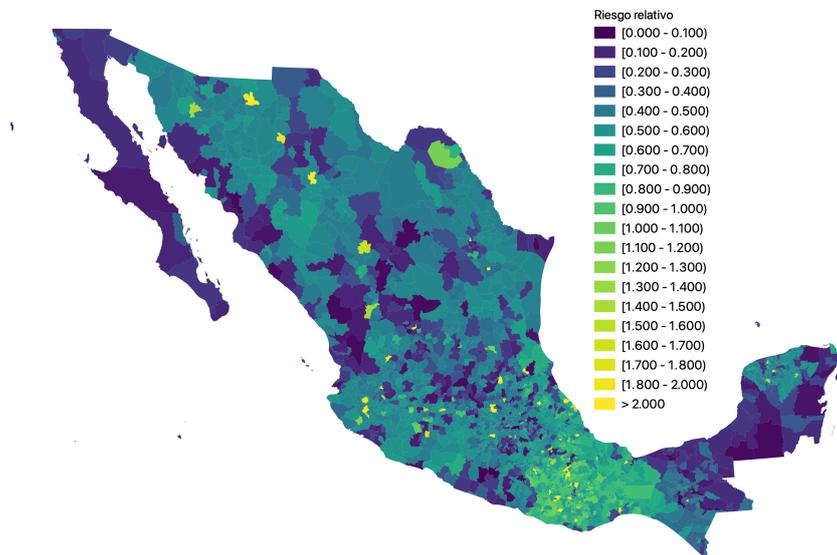


Figura 16: Mapa de México 7: riesgo relativo de enfermedades cardiovasculares por unidad de área, para mujeres de 51 a 60 años.

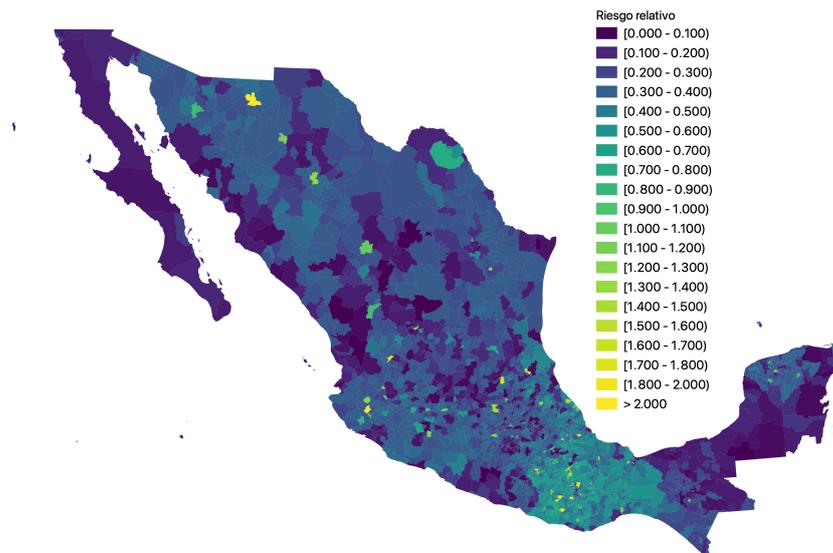


Figura 17: Mapa de México 8: riesgo relativo de enfermedades cardiovasculares por unidad de área, para hombres de 51 a 60 años.

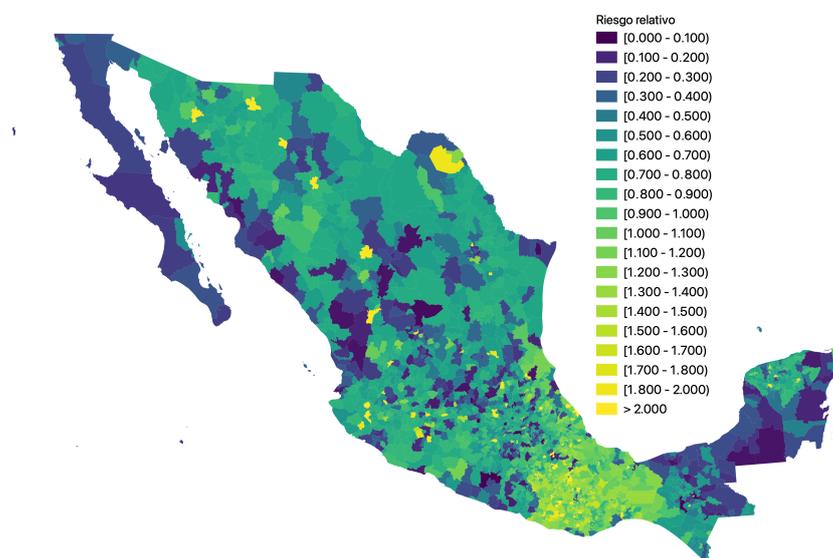


Figura 18: Mapa de México 9: riesgo relativo de enfermedades cardiovasculares por unidad de área, para mujeres mayores a 60 años.

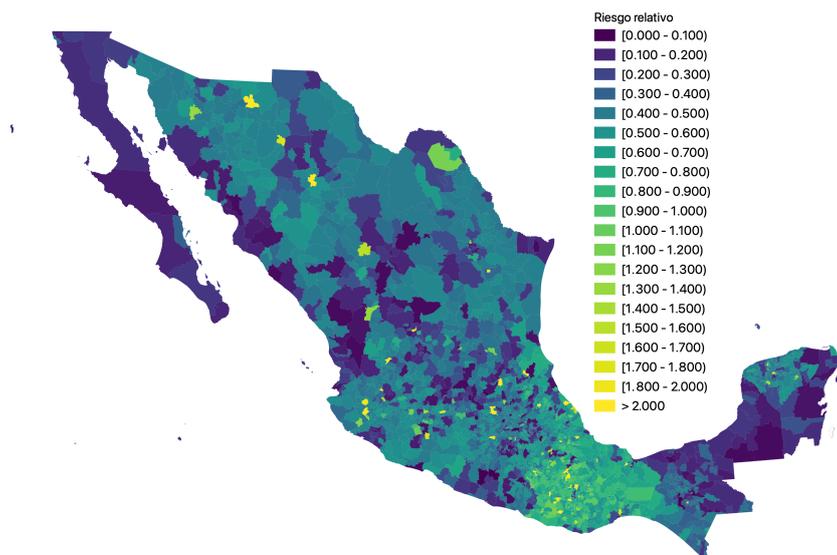


Figura 19: Mapa de México 10: riesgo relativo de enfermedades cardiovasculares por unidad de área, para hombres mayores a 60 años.

3.3 DISCUSIÓN

Existen estudios como el de Barquera y Tolentino (2005) en donde realizan un análisis de las principales características de la transición epidemiológica y nutricional en México (magnitud, distribución, determinantes, asociaciones y complicaciones), utilizando para ello encuestas nacionales y haciendo énfasis en las enfermedades crónicas asociadas a la nutrición.

En ese documento exponen que enfermedades como diabetes, hipertensión, enfermedad coronaria, dislipidemias, osteoporosis y cáncer se han convertido en los principales problemas de salud pública. En dicho estudio, de manera similar a lo hecho en esta tesis, se divide al territorio mexicano en áreas que permita ver similitudes entre los estados que se consideran en estas regiones. En él no se estima un riesgo sino que se observa la prevalencia de enfermedades en dichas regiones, lo cual lo hace buen parámetro para documentar que hay más estudios

que se interesan en la medición de la salud pública con base en métodos estadísticos.

La prevalencia es una medida que permite calcular la proporción de personas enfermas en un lugar y momento definido en el que se pretende medir el padecimiento. Por otro lado, el riesgo relativo es la probabilidad de que una persona enferme después de la exposición a una variable de riesgo en comparación con la probabilidad de que ocurra en un grupo de control o referencia (Pitas et al. 2004). En nuestro caso lo tomamos como la regresión lineal que existe entre las covariables y la variabilidad espacial:

$$r_i = X\beta + u_i + v_i,$$

En el presente trabajo, a diferencia del de Barquera y Tolentino (2005), se buscó estimar el riesgo relativo de padecer alguna enfermedad cardiovascular en México destacando en que esto puede ser medible en cada municipio a pesar de la falta de datos. Para esto se utilizó el modelo de Besag et al (1991) ajustado por medio de aproximaciones de Laplace. Una de la principales razones por las cuales se usó INLA fue por la cantidad de datos faltantes, ya que solo 850 de 2,457 municipios fueron encuestados. Citando al INEGI (2020):

«Lograr, mediante encuestas, estimaciones locales que permitan generar un análisis descriptivo y líneas base de indicadores, requiere de ampliar las muestras, con el respectivo aumento en los costos de los proyectos, situación que difícilmente puede ser subsanada por los organismos nacionales de estadística.»

Se obtuvieron resultados que son interesantes. Por ejemplo, se obtuvo que ser mujer representa un mayor riesgo con respecto a ser hombre. Esto es observable también en los mapas, ya que comparando a ambos sexos para cada grupo de edad los mapas de mujeres, los municipios tienden a tener un mayor riesgo asociado (cercaos o superiores a 1). Según el Texas Heart Institute (2020)

«La enfermedad cardiovascular ya no se considera una enfermedad que afecta sólo a los hombres. Anteriormente, las

mujeres típicamente recibían tratamiento menos intensivo para las enfermedades del corazón y se realizaban menos estudios diagnósticos. Por consiguiente, cuando muchas mujeres eran por fin diagnosticadas con enfermedades del corazón, generalmente se encontraban en un estado más avanzado de la enfermedad y su pronóstico era peor. Ahora sabemos que la enfermedad cardiovascular afecta a más mujeres que hombres».

Por otro lado, es posible notar que de igual forma los mapas son más oscuros (riesgo menor a 1) para los grupos de edad con personas más jóvenes, es decir, para el grupo que tiene entre veinte y treinta años. El grupo de personas mayores de 60 años es indudablemente el grupo más vulnerable de padecer una enfermedad cardiovascular, lo cual indica que existe un impacto de la edad sobre el riesgo cardiovascular. Según la Fundación Española del Corazón (2020)

«La persona de edad avanzada va a ser la principal víctima de la mayor parte de las cardiopatías. La prevalencia y la incidencia de insuficiencia cardiaca se duplica cada década a partir de los 40-45 años. Y más o menos cabe hacer consideraciones similares sobre procesos tan frecuentes e importantes como las enfermedades coronarias o la hipertensión arterial entre otros muchos.».

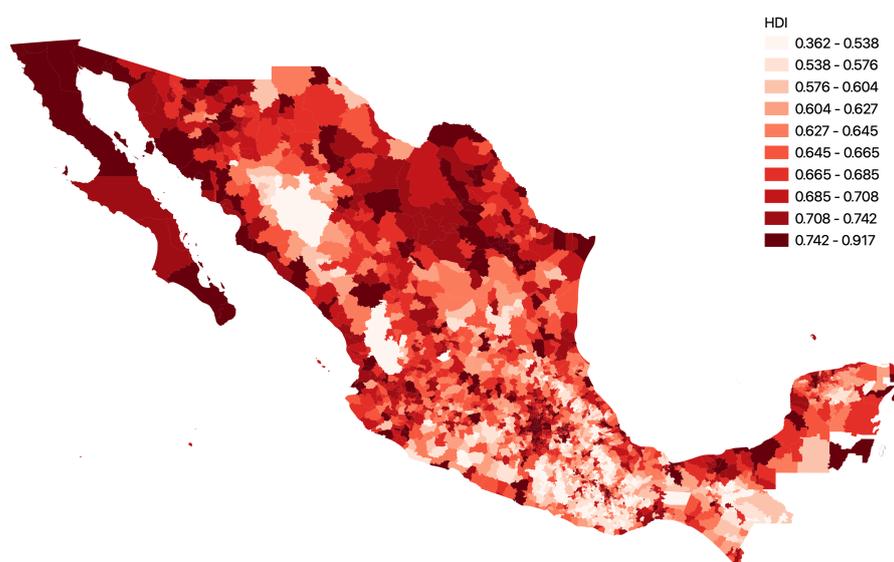


Figura 20: Mapa de México: Índice de desarrollo humano por municipio (datos obtenidos de INEGI, 2010).

En los mapas se puede ver que los estados de Chiapas, Oaxaca, algunos municipios Veracruz, de la península de Yucatán y de la sierra Tarahumara son los que tiene un mayor riesgo asociado (riesgo relativo mayor a 1). Los municipios de los estados de Baja California, Baja California Sur, algunos de la zona norte y de la costa del Pacífico entre Sonora y Nayarit, son en general los de menor riesgo asociado (riesgo relativo menor a 1). Estos ejemplos van de la mano con el Índice de Desarrollo Humano (IDH), en donde podemos ver que en general los municipios con menor riesgo coinciden con los municipios de mayor (IDH) observados en la figura 20. Esto puede estar asociado a algunos factores de riesgo como la falta de servicios de salud, medicamentos y educación.

CONCEPTOS DE PROBABILIDAD Y PROCESOS ESTOCÁSTICOS

Este apéndice se agregó con el objetivo de poder consultar algunos conceptos de probabilidad y procesos estocásticos que se hayan utilizado en los capítulos anteriores. Para esto se utilizaron las definiciones e ideas de Ross (1996).

CONCEPTOS BÁSICOS DE PROBABILIDAD

Una de las nociones básicas de la teoría de probabilidad es un experimento aleatorio, es decir, aquel en el cual no se puede determinar su resultado de antemano. Al conjunto de todos los posibles resultados de un experimento se le conoce como *espacio muestral S*.

Por otro lado, un evento es un subconjunto del espacio muestral, y se dice que este ocurre si el resultado del experimento es un elemento de ese subconjunto. Supongamos que para cada evento E del espacio muestral S un número $\mathbb{P}(E)$ es definido y satisface los siguientes axiomas:

1. $0 \leq \mathbb{P}(E) \leq 1$
2. $\mathbb{P}(S) = 1$
3. Para cualquier sucesión de eventos E_1, E_2, \dots , para los cuales $\mathbb{P}(E_i \cap E_j) = \emptyset$ cuando $i \neq j$, donde \emptyset es el conjunto vacío,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i).$$

Luego, llamamos a $\mathbb{P}(E)$ como la probabilidad del evento E.

Variable aleatoria

Consideremos un experimento aleatorio del espacio muestral S . Una *variable aleatoria* X es una función que asigna un valor real a cada resultado en S . Para cualquier conjunto de números reales A , la probabilidad de que X asuma un valor contenido en el conjunto A es igual a la probabilidad de que el resultado del experimento sea contenido en $X^{-1}(A)$. Es decir,

$$\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)),$$

en donde $X^{-1}(A)$ es el evento que consiste de todos los puntos $s \in S$ tal que $X(s) \in A$. La *función de distribución* F de la variable aleatoria X es definida para cada número real x por

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x]).$$

Luego, denotemos a $1 - F(x)$ como $\bar{F}(x)$, y por lo tanto

$$\bar{F}(x) = \mathbb{P}(X > x).$$

Se dice que la variable X es *discreta* si su conjunto de posibles valores es contable. Para variables aleatorias contables,

$$F(x) = \sum_{y \leq x} \mathbb{P}(X = y).$$

En cambio, se dice que una variable aleatoria X es *continua* si existe una función $f(x)$, llamada *función de densidad de probabilidad*, tal que

$$\mathbb{P}(X \in B) = \int_B f(x) dx$$

para todo conjunto B . Como $F(x) = \int_{-\infty}^x f(x) dx$, entonces

$$f(x) = \frac{d}{dx} F(x).$$

La *función de distribución conjunta* F de dos variables aleatorias X y Y es definida como

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

las funciones de distribución X y Y ,

$$F_X(x) = \mathbb{P}(X \leq x) \quad \text{y} \quad F_Y(y) = \mathbb{P}(Y \leq y),$$

pueden ser obtenidas de $F(x, y)$ haciendo uso de la propiedad de continuidad del operador de probabilidad. Sea y_n con $n \geq 1$, una sucesión creciente convergiendo a ∞ . Entonces, como los eventos $\{X \leq x, Y \leq y_n\}$ tal que $n \geq 1$, son crecientes y

$$\lim_{n \rightarrow \infty} \{X \leq x, Y \leq y_n\} = \bigcup_{n=1}^{\infty} \{X \leq x, Y \leq y_n\} = \{X \leq x\},$$

se sigue por la propiedad de continuidad que

$$\lim_{n \rightarrow \infty} \{\mathbb{P}(X \leq x, Y \leq y_n)\} = \mathbb{P}(X \leq x),$$

o equivalentemente,

$$F_X(x) = \lim_{y \rightarrow \infty} F(x, y) \quad \text{y} \quad F_Y(y) = \lim_{x \rightarrow \infty} F(x, y).$$

Se dice que las variables aleatorias X y Y son *independientes* si

$$F(x, y) = F_X(x)F_Y(y), \text{ para todo } x \text{ y } y.$$

Las variables aleatorias X y Y son *conjuntamente continuas* si existe una función $f(x, y)$, llamada *función de densidad de probabilidad conjunta*, tal que

$$\mathbb{P}(X \in A, Y \in B) = \int_A \int_B f(x, y) \, dy \, dx, \text{ para cualesquiera conjuntos } A \text{ y } B.$$

La distribución conjunta de cualquier colección X_1, X_2, \dots, X_n de variables aleatorias es definida como

$$F(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Además, se dice que las n variables aleatorias son independientes si

$$F(x_1, \dots, x_n) = F_{x_1}(x_1)F_{x_2}(x_2) \dots F_{x_n}(x_n),$$

en donde

$$F_{X_i}(x_i) = \lim_{x_i \rightarrow \infty} F(x_1, \dots, x_n).$$

Esperanza, varianza y covarianza

La *esperanza* o *media* de una variable aleatoria X , $\mathbb{E}[X]$, se define como

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} x \, dF(x) \\ &= \int_{-\infty}^{\infty} xf(x) \, dx \end{aligned}$$

Por otro lado, la *varianza* de una variable aleatoria X es definida como

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \end{aligned}$$

Por otra parte, se dice que dos variables aleatoriamente distribuidas conjuntamente X y Y son no correlacionadas si la covarianza, definida por

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E} [(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

es cero.

PROCESOS ESTOCÁSTICOS

Una vez que se tienen claras las ideas básicas de probabilidad se puede dar lugar a un concepto que pertenece a la misma teoría, pero en donde esta vez, los datos aleatorios varían con el tiempo; los procesos estocásticos. Este concepto será muy útil en el desarrollo de este documento, pues es muy utilizado dentro de la estadística espacial y es importante para comprender el concepto de campo aleatorio.

Un *proceso estocástico* $X_T = \{X(t), t \in T\}$ es una colección de variables aleatorias, tal que para cada t en el conjunto indexado T , $X(t)$ es una variable aleatoria. Es común, interpretar a t como tiempo y a $X(t)$ el estado del proceso al tiempo t . Si T es un conjunto contable, entonces a X_T lo llamamos proceso estocástico a tiempo discreto, y si T es continuo, lo llamamos proceso estocástico a tiempo continuo.

A cada iteración de X_T se le llama trayectoria del proceso. Por ejemplo, si los eventos ocurren aleatoriamente en el tiempo y $X(t)$ representa el número de eventos que ocurren en $[0, t]$, entonces la figura 21 muestra la trayectoria de X_T correspondiente al evento inicial en el tiempo 1, seguido del evento en tiempo 3 y el tercero en tiempo 4, y sin más eventos en otro tiempo.

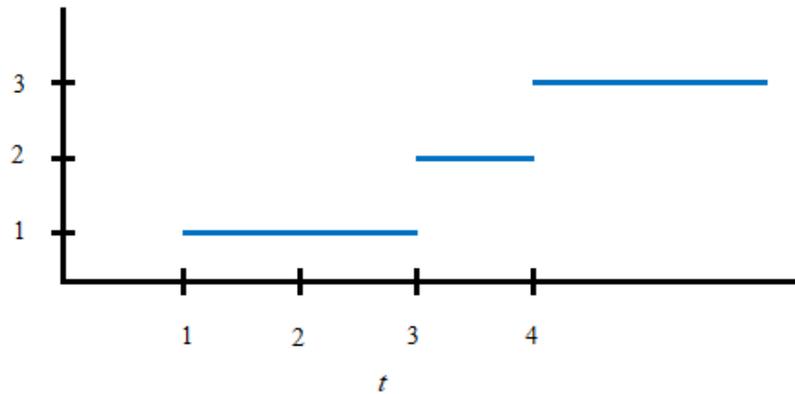


Figura 21: Trayectoria de $X(t) = \text{número de eventos en } [0, t]$ (tomada de Ross, 1996).

Por otra parte, se dice que un proceso a tiempo continuo $\{X(t), t \in T\}$ tiene *incrementos independientes* si para todo $t_0 < t_1 < t_2 < \dots < t_n$, las variables aleatorias

$$X(t_1) - X(t_0), X(t_2) - X(t_1), \dots, X(t_n) - X(t_{n-1})$$

son independientes. También, se dice que el proceso tiene *incrementos estacionarios* si $X(t+s) - X(t)$ tiene la misma distribución para todo t , es decir, posee incrementos independientes si los cambios en el valor de los procesos en intervalos de tiempo no superpuestos son independientes; y posee incrementos estacionarios si la distribución del cambio en valor entre dos puntos depende solo de la distancia entre esos puntos.

Existen diferentes tipos de procesos, dependiendo de las características de las trayectorias y de las relaciones de dependencia entre las variables aleatorias que conforman el proceso. A continuación se mencionan algunos:

Procesos gaussianos

Se dice que un proceso estocástico a tiempo continuo $\{X_t : t \geq 0\}$ es un proceso gaussiano si para cualquier colección finita de tiempos t_1, \dots, t_n , el vector $(X_{t_1}, \dots, X_{t_n})$ tiene distribución normal o gaussiana multivariada.

Procesos de Markov

Estos tipos de procesos son modelos en donde, suponiendo conocido el estado presente del sistema, los estados anteriores no tienen influencia en los estados futuros del sistema, esta condición es llamada *propiedad de Markov*, la cual podemos expresar de la siguiente forma:

Para cualesquiera estados x_0, x_1, \dots, x_{n-1} (pasado), x_n (presente), x_{n+1} (futuro), se cumple que

$$\mathbb{P}(X_{n+1} \in x_{n+1} | X_0 \in x_0, \dots, X_n \in x_n) = \mathbb{P}(X_{n+1} \in x_{n+1} | X_n \in x_n)$$

es decir, la probabilidad del evento futuro ($X_{n+1} \in x_{n+1}$) sólo depende del evento presente ($X_n \in x_n$).

Proceso Poisson

Decimos que un proceso estocástico $\{N(t), t \geq 0\}$ es un *proceso de conteo* si $N(t)$ representa al número total de eventos que han ocurrido hasta ahora. Por lo tanto, un proceso de conteo $N(t)$ debe satisfacer que

1. $N(t) \geq 0$.
2. $N(t)$ es un valor entero.
3. Si $s < t$, entonces $N(s) \leq N(t)$.
4. Para $s < t$, $N(t) - N(s)$ es igual al número de eventos ocurridos en el intervalo $(s, t]$.

Un proceso de conteo tiene *incrementos independientes* si el número de eventos que ocurren en intervalos de tiempo disjuntos son independientes. Esto significa que el número de eventos ocurridos en el tiempo t , $N(t)$, tienen que ser independientes del número de eventos ocurridos entre los tiempos t y $t + s$, $N(t + s) - N(t)$.

Por otro lado, un proceso de conteo tiene *incrementos estacionarios* si la distribución del número de eventos que ocurren en cualquier intervalo de tiempo dependen únicamente de la longitud del intervalo de tiempo. Es decir, el proceso tiene incrementos estacionarios si el número de eventos en el intervalo $(t_1 + s, t_2 + s]$, $N(t_2 + s) - N(t_1 + s)$, tiene la misma distribución que el número de eventos en el intervalo $(t_1, t_2]$, $N(t_2) - N(t_1)$, para todo $t_1 < t_2$ y $s > 0$.

Luego, se dice que el proceso de conteo $\{N(t), t \geq 0\}$ es un *proceso Poisson* con tasa λ , $\lambda \geq 0$ si

1. $N(0) = 0$.
2. El proceso tiene incrementos independientes.
3. El número de eventos en cualquier intervalo de longitud t se distribuye Poisson con media λt . Esto es, para toda $s, t \geq 0$,

$$\mathbb{P}(N(t+s) - N(s) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, \dots$$

Nótese que, siguiendo la tercera condición, un proceso Poisson tiene incrementos E y que

$$\mathbb{E}[N(t)] = \lambda t,$$

lo cual explica porqué a λ se le llama tasa del proceso.

Proceso Poisson no homogéneo

Un proceso de conteo $\{N(t), t \geq 0\}$ es no estacionario o no homogéneo con función de intensidad $\lambda(t), t \geq 0$ si

1. $N(0) = 0$
2. $N(t), t \geq 0$ tiene incrementos independientes
3. $\mathbb{P}(N(t+h) - N(t) \geq 2) = o(h)$
4. $\mathbb{P}(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$

Según Gelman et al. (2004) por análisis bayesiano de datos se refieren a métodos prácticos para hacer inferencias a partir de datos usando modelos probabilísticos. La característica esencial de los métodos Bayesianos es su uso explícito de la probabilidad para cuantificar la incertidumbre en inferencias basadas en el análisis de datos estadísticos. A diferencia de los métodos clásicos, los cuales se preguntan qué nos dicen los datos acerca del valor del parámetro (ignorando con ello toda evidencia externa a los datos), los métodos bayesianos se preguntan cómo cambia nuestro estado de información acerca del valor del parámetro (o de cualquier otra cantidad de interés) debido a los datos observados. Entonces, al aceptar información adicional, los métodos bayesianos reconocen que cada problema es distinto y promueven que el procedimiento de análisis se adapte al problema en cuestión (Gutiérrez, 2019).

Distribución a priori

La elección de un modelo $\pi(y|\theta)$ para describir el proceso que generó los datos dependerá del tipo de problema que se enfrente y deberá de ser consistente con la información que tengamos previamente. La asignación de una distribución inicial no es una tarea sencilla, sin embargo, es un punto importante para el enfoque bayesiano y aunque no tengamos información acerca del parámetro, debemos ser capaces de reflejarlo.

Supongamos que además de la distribución de la muestra $f(x|\theta)$, una distribución inicial de θ , $\pi(\theta)$, es asequible. Luego, la distribución conjunta de (θ, x) es

$$\varphi(\theta, x) = f(x|\theta)\pi(\theta),$$

mientras que la distribución marginal de x es

$$m(x) = \int_{\theta} \varphi(\theta, x) d\theta = \int_{\theta} f(x, |\theta) \pi(\theta) d\theta,$$

al no ser condicional sobre una observación previa (*prior*) y al ser la distribución de una cifra observable (*predictiva*), nosotros la llamaremos *distribución predictiva a priori*.

Distribución predictiva posterior

Siguiendo lo anterior, después de que x ha sido observada y siguiendo el teorema de Bayes,

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\theta} f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{f(x|\theta)\pi(\theta)}{m(x)}, \end{aligned}$$

a la cual llamamos la distribución posterior.

Distribución posterior como un arreglo entre datos e información a priori

El proceso de inferencia bayesiana implica pasar de una distribución a priori $\pi(\theta)$, a una distribución posterior $\pi(\theta|y)$, y es natural suponer que haya algunas relaciones generales entre estas dos distribuciones. Por ejemplo, se puede suponer que, dado que la distribución posterior incorpora la información del conjunto de datos, ésta será menos variable que la distribución a priori. Este concepto puede ser formalizado de la siguiente manera:

$$\mathbb{E} [\theta] = \mathbb{E} \left[\mathbb{E} [\theta|y] \right]$$

y

$$\text{Var} (\theta) = \mathbb{E} \left[\text{Var} (\theta|y) \right] + \text{Var} \left(\mathbb{E} [\theta|y] \right),$$

lo cual quiere decir que la media a priori de θ es el promedio de todas las posibles medias posteriores sobre la distribución de los datos, y que la varianza posterior es en promedio más pequeña que la varianza a priori.

Distribuciones a priori conjugadas

Definición 1 (Distribución a priori conjugada). Sea \mathcal{F} una clase de distribuciones muestrales $\pi(y|\theta)$, y sea \mathcal{P} una clase de distribuciones a priori para θ , se dice que la clase \mathcal{P} es conjugada por \mathcal{F} si

$$\pi(\theta|y) \in \mathcal{P}, \text{ para toda } \pi(\cdot|\theta) \in \mathcal{F} \text{ y } \pi(\cdot) \in \mathcal{P}.$$

Ejemplo

Supongamos que queremos obtener la distribución posterior de un modelo Gamma - Poisson, es decir,

$$x_1, x_2, \dots, x_n | \lambda \sim \text{Poisson}(\lambda)$$

$$\lambda \sim \text{Gamma}(\alpha, \beta).$$

Dada la distribución de λ , la densidad a priori $p(\lambda) \sim \text{Gamma}(\alpha, \beta)$:

$$p(\lambda) = \frac{(\beta)^\alpha}{\Gamma(\alpha)} e^{-\beta\lambda} \lambda^{\alpha-1}, \quad \alpha, \beta > 0,$$

entonces, la posterior debe de seguir una Gamma también.

Por otro lado, la verosimilitud de una Poisson es:

$$p(x_1, x_2, \dots, x_n | \lambda) = \frac{e^{-n\lambda} \lambda^{n\bar{x}}}{\prod_{i=1}^n x_i!} \propto e^{-n\lambda} \lambda^{n\bar{x}}$$

por lo que la posterior será:

$$p(x_1, x_2, \dots, x_n | \lambda) = \frac{p(\lambda)P((x_1, x_2, \dots, x_n | \lambda))}{\int_0^{\infty} p(\lambda)P((x_1, x_2, \dots, x_n | \lambda)) d\lambda}$$

$$\propto p(\lambda)P((x_1, x_2, \dots, x_n | \lambda)),$$

luego entonces, se tiene que

$$p(\lambda | x_1, x_2, \dots, x_n) \propto (e^{-\beta\lambda})(\lambda^{\alpha-1})(e^{-n\lambda})(\lambda^{n\bar{x}})$$

por lo que,

$$p(\lambda | x_1, x_2, \dots, x_n) \propto e^{-(\beta+n)\lambda} \lambda^{(\alpha+n\bar{x})-1}.$$

Por lo tanto,

$$\lambda | x_1, x_2, \dots, x_n \sim \text{Gamma}(\alpha + n\bar{x}, \beta + n).$$



CÓDIGO DE R

CÓDIGO DE R

```
#Se cargan librerías
pacman::p_load(tidyverse, magrittr, janitor, BBmisc, clipr,
.           readxl, dplyr)

#Se cargan los archivos shp
#Datos de las áreas geostatísticas municipales
#Esto es un dataframe espacial poligonal
mapa <- rgdal::readOGR(dsn = 'ruta', layer = 'datos',
.           use_iconv = TRUE, encoding = 'UTF-8')

#Se carga el dataframe de enfermedades cardiovasculares
cardio <- readRDS('base_enf_cardio.rds')
mapa@data <- mapa@data%>% as_tibble()%>%
.           janitor::clean_names()%>%
.           mutate(ent_mun = paste0(cve_ent, cve_mun))%>%
.           left_join(cardio%>%
.           select(ent_mun, ent_mun_2)%>% unique())

#Modelo Besag et al (1991)
formula <- round(yi) ~ 1 + idh + sexo + cat_edad + muertes
.           + f(ent_mun_2, model = 'bym',
.           graph = poly2nb(mapa), scale.model = TRUE)

inla_card <- inla(formula, family = "poisson", data = cardio,
.           control.predictor = list(compute=T, link=1),
.           E = ci, control.compute = list(dic = TRUE,
.           cpo = TRUE, waic = TRUE))

#Coeficientes de las covariables de la distribución posterior
inla_card$summary.fixed

#INLA CPO y DIC
inla_card$cpo$cpo
inla_card$dic$dic
```

```
#Índice de Moran
iid <- inla_card$summary.random%>%
.   as.data.frame()%>% slice(-(1:2458))%>%
.   as_tibble()%>% dplyr::select(ent_mun_2.mean)%>%
.   bind_cols(iid_besag%>%
.   rename(iid_besag = ent_mun_2.mean)%>%
.   dplyr::select(iid_besag))%>%
.   mutate(residuos = iid_besag - ent_mun_2.mean)

moran.test(iid$residuos, nb2listw(poly2nb(mapa)),
.   na.action = na.omit, zero.policy = T)
```

REFERENCIAS BIBLIOGRÁFICAS

- [1] Adler, R. y Taylor, E. (1993). Random fields and geometry. Montfort University.

- [2] Arab, A., Hooten, M. y Wikle, C. (2006). Hierarchical spatial models. University of Missouri-Columbia. 1, 1:16.

- [3] Barquera, S. y Tolentino, L. (2005). Geografía de las enfermedades asociadas con la nutrición en México: una perspectiva de transición epidemiológica. Instituto Nacional de Salud Pública de México.

- [4] Berrington, A., Hartge, P., Cerhan, J., Flint, A., Hannan, L., MacInnis, R., Moore, S., Tobias, G., Anton-Culver, H., Beane, L., Lawrence, W., Clipp, S., Folsom, A., Freedman, M., Giles, G., Hakansson, N., Henderson, K., Hoffman-Bolton, J., Hoppin, J., Koenig, K., Lee, I., Linet, M., Park, Y., Pocobelli, G., Schatzkin, A., Sesso, H., Weiderpass, E., Willcox, J., Wolk, A., Zeleniuch-Jacquotte, A., Willett, W. y Thun, M. (2010). Body-mass index and mortality among 1.46 million white adults. The New England Journal of Medicine. 363, 2211:2219.

- [5] Besag, J., York, J. y Mollié, A. (1991). Bayesian Image Restoration, With Two Applications In Spatial Statistics. University of Washington. 43(1), 1:20.

- [6] Casella, G. y Robert, C. (1999). Monte Carlo statistical methods. Springer.

- [7] Chilès, J. y Delfiner, P. (1999). Geostatistics modeling spatial uncertainty. Wiley - Interscience Publication.

- [8] Colin, A. y Trivedi, P. (1998). Regression analysis of count data. Cambridge University.

- [9] Cressie, N. (1993). *Statistics for spatial data*. Iowa State University.
- [10] Diggle, P., Fuentes, M., Gelfand, A. y Guttorp, P. (2010). *Handbook of spatial statistics*. Duke University.
- [11] Gelman, A., Carlin, J., Stern, H., Rubin, D. (2004). *Bayesian data analysis*. Chapman & Hall/CRC.
- [12] Global Burden Disease Collaborative Network. (2017). IH-ME Measuring what matters. Disponible en: [http:// www.healthdata.org/mexico](http://www.healthdata.org/mexico).
- [13] Goodchild, M. (1986). *Spatial autocorrelation*. University of Western Ontario.
- [14] Fundación Española del Corazón. (2020). *Prevención, Marcadores de Riesgo, Edad*. Disponible en: [https:// fundaciondelcorazon.com/prevencion/marcadores-de-riesgo/ edad.html](https://fundaciondelcorazon.com/prevencion/marcadores-de-riesgo/edad.html).
- [15] Gutiérrez, P. (2019). *El desarrollo de la estadística bayesiana*. RDU, UNAM.
- [16] Haining, R. (2003). *Spatial data analysis theory and practice*. Cambridge University.
- [17] Instituto Nacional de Estadística, Geografía e Informática. (2011). *Estadísticas del sector salud y seguridad social*. 19, 50:51.
- [18] Instituto Nacional de Salud Pública. (2012). *Encuesta Nacional de Salud*.
- [19] Leroux, B., X. Lei, and N. Breslow. (1999). Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence. In *Statistical Models in Epidemiology, the Environment and Clinical Trials*, 135–78.

- [20] Norris, J.R. (1997). Markov chains. Cambridge University.
- [21] Organización Mundial de la Salud. (2018). Las 10 principales causas de defunción. Disponible en: <https://www.who.int/es/newsroom/fact-sheets/detail/the-top-10-causes-of-death>.
- [22] Organización Panamericana de la Salud. (2018). Enfermedades no transmisibles. Disponible en: https://www.paho.org/hq/index.php?option=com_topics&view=article&id=345&Itemid=40933&lang=es.
- [23] Pita, P., Pértegas, S. y Valdéz, F. (2004). Medidas de frecuencia de enfermedad. Complejo Hospitalario-Universitario Juan Canalejo.
- [24] Ross, S. (1996). Stochastic processes. University of California, Berkeley.
- [25] Rue, H. y Held, L. (2005). Gaussian Markov random fields theory and applications. Chapman and Hall.
- [26] Rue, H. y Martino, S. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society*. 71, 319:392.
- [27] Secretaría de Salud. (2018). Panorama epidemiológico 2018 enfermedades no transmisibles. Observatorio Mexicano de Enfermedades No Transmisibles.
- [28] Wall, M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*. 121, 311:324.
- [29] Winkler, G. (1990). Image analysis, random fields and dynamic Monte Carlo methods. University of Southern California.