

*Modelación espacial de la mortalidad infantil, el período
postnatal de cribado y el volumen de atención
hospitalaria en Colombia*

MICHEL FELIPE CÓRDOBA PEROZO
ESTADÍSTICO
CÓDIGO: 832528



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
DICIEMBRE DE 2012

*Modelación espacial de la mortalidad infantil, el período
postnatal de cribado y el volumen de atención
hospitalaria en Colombia*

MICHEL FELIPE CÓRDOBA PEROZO

ESTADÍSTICO

CÓDIGO: 832528

DISERTACIÓN PRESENTADA PARA OPTAR AL TÍTULO DE
MAGISTER EN ESTADÍSTICA

DIRECTOR

EDILBERTO CEPEDA CUERVO, PH.D.

DOCTOR EN MATEMÁTICA



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
DICIEMBRE DE 2012

Título en español

Modelación espacial de la mortalidad infantil, el período postnatal de cribado y el volumen de atención hospitalaria en Colombia

Title in English

Spatial modeling of infant mortality, the postnatal screening period and the hospital care in Colombia

Resumen: En este trabajo se proponen modelos de sobredispersión generalizados con estructura econométrica espacial. Como punto de partida se consideran modelos clásicos de la econometría espacial, trabajados principalmente por Anselin (1988) cuya principal característica es el ajuste de un modelo de regresión lineal normal considerado bajo una estructura espacial en los residuales o en el mismo predictor lineal de la media. En este contexto, esta tesis extiende esa idea en datos de naturaleza discreta, específicamente en datos de conteo cuyos supuestos distribucionales pueden ser binomial o Poisson. No obstante, por la disposición geográfica y la naturaleza de la información, estas observaciones no cumplen con el supuesto de independencia, provocando un ajuste inadecuado de los modelos lineales generalizados usuales por la presencia de sobredispersión; así, en esta dirección el uso de modelos sobredispersos cobra sentido. Sin embargo, dado que existe una estructura de correlación espacial latente, los modelos propuestos que aquí se denominan modelos de sobredispersión autocovariantes, permiten cuantificar por medio de una llamada autocovariable el efecto que tiene en la variable respuesta el comportamiento de la misma en las unidades geográficas consideradas más próximas, soportado por el parámetro de sobredispersión, cuyo papel ahora es corregir la dispersión extra que el parámetro de la autocovariable no es capaz de controlar. El ajuste es hecho mediante el uso de métodos bayesianos con algoritmos MCMC. Estos se aplican al análisis de la tasa de mortalidad infantil, el período postnatal de cribado y el volumen de atención hospitalaria en Colombia por departamentos. Estos mismos se definen a partir de los sugeridos en Quintero, Cepeda-Cuervo and Núñez-Antón (2012). En las aplicaciones se analizan tres conjuntos de datos cuya fuente primaria es la encuesta nacional de salud en Colombia para el año 2007. Dos suministrados por Quintero, Cepeda-Cuervo and Núñez-Antón y un tercero generado directamente a partir de la encuesta.

Abstract: In this thesis we propose generalized overdispersion models with spatial econometric structure. As a base line, we consider classical spatial econometric models, worked mainly by Anselin (1998) which characteristic is to fit a normal linear regression model, taken under a spatial structure in the residuals or in the linear predictor of the mean. In this context, this thesis extends that idea in discrete nature data, specifically in data whose distributional assumption may be either binomial or Poisson. Nevertheless, due to the geographical nature, these observations do not meet the independence assumption, provoking an inadequate fit of the commonly generalized linear models, because of the presence of overdispersion. In this way, the use of overdispersed models makes sense. However, since there is a latent spatial correlation structure, the proposed models allow to quantify the effect over the response variable considering its behavior in the adjacent neighbors, supported by the overdispersion parameter, which role is to correct the additional dispersion that cannot be controlled by the spatial correlation. The fitting is done by the use of Bayesian methods with MCMC algorithms. These are

applied to estimate the infant mortality, the postnatal screening period and the hospital care in the departments of Colombia by departments. Those are defined, based on the ones suggested in the Quintero, Cepeda and Nuñez paper (2012). The primary source of the data is the National Health Survey (ENS) for the year 2007 in Colombia.

Palabras clave: Sobredispersión, heterocedasticidad, autocorrelación espacial, sobredispersión no constante, matriz de pesos espaciales, MCMC.

Keywords: Overdispersion, heteroskedasticity, spatial autocorrelation, no constant overdispersion, spatial weights matrix, MCMC (Markov Chain Monte Carlo methos), Bayesian estimation.

Índice general

Índice general	II
Introducción	V
1. Modelación del volumen de atención hospitalaria colombiana	1
1.1. Introducción	1
1.2. Heterogeneidad y dependencia o autocorrelación espacial	1
1.2.1. Matrices de contigüidad espacial	2
1.2.2. Operador de retardo espacial	5
1.2.3. Autocorrelación espacial	6
1.3. Modelos clásicos de la econometría espacial	7
1.3.1. Modelos Autorregresivos - AR	7
1.3.2. Modelos Condicionales Autorregresivos - CAR	7
1.3.3. Modelos Simultáneos Autoregresivos - SAR	9
1.4. Aplicación: Modelación del volumen de atención hospitalaria en Colombia	11
1.4.1. Análisis descriptivo y exploratorio	11
1.4.2. Modelo espacial del volumen de atención hospitalaria en Colombia	15
1.4.3. Otras matrices de pesos espaciales	21
1.4.3.1. Matriz de pesos espaciales de segundo orden	21
1.4.3.2. Matriz de pesos espaciales por regiones naturales	24
1.5. Conclusiones	28
2. Modelación de la mortalidad infantil y el período postnatal de cribado	29
2.1. Modelos lineales generalizados	30
2.2. Modelos usuales de sobredispersión	31
2.2.1. Modelos de conteo con sobredispersión	32

2.2.1.1.	Modelos asociados a la distribución Poisson	32
2.2.1.2.	Modelos asociados a la distribución Binomial	33
2.3.	Otros modelos de sobredispersión	35
2.3.1.	Modelos de cuasi - verosimilitud	36
2.3.2.	Familia doblemente exponencial	37
2.4.	Modelos de sobredispersión generalizados	37
2.4.1.	Modelos asociados a la distribución Poisson	37
2.4.2.	Modelos asociados a la distribución binomial	38
2.5.	Modelos de sobredispersión autocovariantes	38
2.5.1.	Modelos asociados a la distribución Poisson	39
2.5.2.	Modelos asociados a la distribución binomial	41
2.6.	Estimación Bayesiana	43
2.6.1.	MCMC - Markov Chain Monte Carlo	43
2.6.1.1.	Muestreador de Gibbs	44
2.6.1.2.	Algoritmo Metropolis - Hastings	45
2.6.2.	Inferencia Bayesiana	46
2.6.2.1.	Distribución a <i>posteriori</i> - Teorema de Bayes	47
2.6.2.2.	Distribución a <i>priori</i>	48
2.6.2.3.	Modelos de regresión	50
2.6.2.4.	Estimación en modelos de regresión - MLG	51
2.6.2.5.	Estimación en modelos autocovariantes sobredispersos	53
2.7.	Aplicación: Modelación espacial de la tasa de mortalidad infantil colombiana	53
2.7.1.	Modelo Poisson	54
2.7.2.	Modelo autocovariante Poisson	55
2.7.3.	Modelo de sobredispersión binomial negativo	55
2.7.4.	Modelo binomial negativo autocovariante	56
2.7.5.	Modelo Poisson Normal	57
2.7.6.	Modelo autocovariante Poisson normal	58
2.7.7.	Modelo Poisson normal generalizado	58
2.8.	Aplicación: Modelación espacial del período postnatal de cribado en Colombia	61
2.8.1.	Modelo logístico (Binomial)	62
2.8.2.	Modelo autocovariante logístico (Binomial)	63
2.8.3.	Modelo Beta - Binomial	63
2.8.4.	Modelo autocovariante Beta - Binomial	64
2.8.5.	Modelo Binomial Normal	65

2.8.6. Modelo autocovariante Binomial Normal	66
2.8.7. Modelo beta binomial generalizado	66
2.9. Aplicación: Otras matrices de pesos espaciales	68
2.9.1. Matriz de contigüidad física de segundo orden	68
2.9.1.1. Modelos con respuesta Poisson	68
2.9.1.2. Modelos con respuesta binomial	69
2.9.2. Matriz de contigüidad física por regiones naturales	70
2.9.2.1. Modelos con respuesta Poisson	70
2.9.2.2. Modelos con respuesta binomial	71
2.10. Aplicación: Volumen de atención hospitalaria en Colombia	72
2.11. Conclusiones	74
A. Algunos resultados para las estadísticas de correlación	75
B. Estimación por máxima verosimilitud e inferencia para el modelo SAR	79
C. Residuales Modelo CAR Residual	84
D. Códigos R y Winbugs	87
Bibliografía	90

Introducción

En el uso de los modelos lineales generalizados, una de las características más comunes y frecuentes en la estructura intrínseca de los datos, es la presencia de sobredispersión. Esto indica, que en muchos de los casos cuando se pretende modelar ciertos datos con las técnicas básicas ya desarrolladas, no es posible encontrar un ajuste adecuado, ni un modelo capaz de describir la disposición total de la información. De manera general, se supone que se tiene una muestra aleatoria de una variable Y_i que pertenece a la familia exponencial de distribuciones y que su media μ_i puede ser modelada por un conjunto de variables exógenas \mathbf{x}_i^T de forma que mediante el predictor lineal $g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, el vector de parámetros desconocido $\boldsymbol{\beta}$ cuantifica el grado de asociación entre estas y la transformación de la media $g(\mu_i)$; este conjunto de variables es denominado generalmente como covariables. La función $g(\cdot)$ debe ser monótona y diferenciable y se conoce como función de enlace.

La sobredispersión aparece cuando la varianza excede a la que es capaz de capturar el modelo escogido. Para el tratamiento de la información cuando hay presencia de este fenómeno, existe una variedad de técnicas desde diversos enfoques, que encierra desde la inclusión en el predictor lineal de un efecto aleatorio, hasta el modelamiento jerárquico, asumiendo distribuciones en los dos niveles. En cualquier caso, de manera general, se llega nuevamente a un modelo de la forma $g(\mu_i) = \eta_i$, donde algún parámetro que define la varianza puede ser modelado por alguna función de variables explicativas $h(\mathbf{Z}_i^T \boldsymbol{\gamma})$, con $\boldsymbol{\gamma}$ un vector de parámetros a ser estimado, y $h(\cdot)$ una función de enlace, monótona y diferenciable, posiblemente distinta de $g(\cdot)$.

La aparición del fenómeno de la sobredispersión se puede presentar cuando no se cumplen algunos de los supuestos básicos asumidos en la construcción de un modelo estándar; en particular la no independencia entre las observaciones y la varianza no constante. Si no hay independencia entre las observaciones, es posible pensar en algún tipo de correlación (y que si se tiene información de carácter espacial, puede haber influencia de los vecinos más próximos en el comportamiento de su respuesta).

En la construcción de un modelo que reúna la información de los vecinos más cercanos, se hace uso de la teoría econométrica espacial, en donde un modelo lineal está definido por la forma general

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{Y} \tag{1}$$

en donde $E(\mathbf{Y}) = \boldsymbol{\mu}$, con \mathbf{X} y $\boldsymbol{\beta}$ matriz diseño conocida de tamaño $n \times (p + 1)$ y vector de parámetros de tamaño $(p + 1) \times 1$ respectivamente, definidos como antes, ρ es llamado parámetro de autocorrelación espacial, \mathbf{W} es una matriz de pesos espaciales de dimensión

$n \times n$, cuya entrada w_{ij} refleja la intensidad de la interdependencia entre las regiones indexadas por i y j y el vector \mathbf{Y} es el vector de observaciones de la variable dependiente. En este escrito se revisa y usa un modelo que cumple con la unificación de los dos métodos mencionados, los modelos lineales generalizados y la econometría espacial, es decir, un modelo de forma general:

$$g(\mu_i) = \mathbf{X}_i\boldsymbol{\beta} + \rho\mathbf{W}_i\mathbf{Y}_i \quad (2)$$

con las características ya mencionadas. Esta tesis propone la especificación de este modelo para el caso Binomial y Poisson, en ambos considerándose la sobredispersión. Además de esto, se realiza el mismo procedimiento cuando la varianza de la variable aleatoria puede ser especificada en términos de otras variables explicativas, como ya se mencionó anteriormente. Así, un objetivo es determinar cómo mejora la inclusión en el predictor lineal de la variable rezagada. Además, se presentan tres modelos con efectos espaciales, uno con distribución de la respuesta Normal, otro con respuesta tipo Poisson y otro binomial. En efecto, para datos a nivel departamental en Colombia, la variable con respuesta Normal es el volumen de atenciones en total de las IPS en el año 2007, la variable con respuesta Poisson es el número de niños fallecidos entre el 2000 y 2005 y la variable con respuesta binomial es el número de madres que han pasado por un período postnatal de cribado entre 2000 y 2005. La estimación de los parámetros está enfocada a la perspectiva Bayesiana, con ayuda de los métodos de simulación MCMC (Markov Chain Monte Carlo), aunque se muestra una aplicación por el método de máxima verosimilitud para el modelo con respuesta normal.

Para la construcción los modelos mencionados anteriormente, se necesita recordar muchos aspectos de la teoría estadística. Esta tesis aborda en el capítulo uno, los modelos econométricos espaciales y sus características más importantes. En el capítulo dos se tratan los modelos de sobredispersión con respuesta Poisson y binomial; los más usuales y algunos propuestos recientemente, desarrollados por ejemplo por autores como Williams ([69], [70]), Breslow [10], Crowder [18] y Hinde [39] entre otros. A lo largo de este contenido, también se comenta la forma de estimación de los parámetros de regresión, se muestra un análisis descriptivo de las variables más relevantes que ayudan al modelamiento, y su aporte con sus respectivas correlaciones a la mejora de los modelos, se muestran los modelos resultantes, las conclusiones y las interpretaciones más importantes de las ecuaciones.

Modelación espacial del volumen de atención hospitalaria en Colombia por Modelos Econométricos Espaciales

En este capítulo se introducen conceptos relevantes de los modelos econométricos lineales espaciales. Se define la matriz de pesos espaciales y algunas técnicas descriptivas básicas de datos con estructura espacial.

1.1. Introducción

Los datos de los que se dispone para realizar el documento, pertenecen a información recolectada a nivel departamental en Colombia. Esto sugiere que posiblemente, hay una estructura de correlación intrínseca en cada una de las variables de interés, ya que las unidades sobre las que se mide la información presentan un orden en el espacio. Esta estructura debe ser definida a partir de los conceptos de dependencia y heterogeneidad espacial.

1.2. Heterogeneidad y dependencia o autocorrelación espacial

Anselin [4] establece que la heterogeneidad espacial puede ser dispuesta como la falta de uniformidad de los efectos espaciales de una variable. Comenta también que desde el punto de vista econométrico, la heterogeneidad espacial puede ser tratada considerando parámetros variables, coeficientes aleatorios o varias formas de cambio estructural tales como modelos de regresión “switching” (regresión con cambios bruscos). Además de esta falta de estabilidad estructural de los muchos fenómenos sobre el espacio, las unidades espaciales observacionales, están lejos de ser homogéneas. También señala que la heterogeneidad se puede reflejar en los errores de un eventual modelo de regresión, en donde seguramente habrá presencia de heterocedasticidad, una de las causas de sobredispersión.

En términos generales, la dependencia espacial puede ser considerada como la existencia de una relación funcional entre lo que pasa en un punto del espacio y lo que pasa en otro lugar [4]. Como ya se había mencionado, en datos que presentan características con esta estructura espacial, es muy probable que exista un problema de sobredispersión. La presencia de autocorrelación espacial en un conjunto de datos, es vista en muchos casos como un serio problema en el modelamiento de los mismos, ya que con esta estructura

no se mantienen algunas de las más básicas hipótesis que reposan sobre el modelamiento estándar, como la independencia y la distribución idéntica (para el caso de la heterocedasticidad) de los errores. De esta manera, el error de tipo I, se puede aumentar, ocasionalmente incluso invirtiendo la pendiente de alguna relación en algún estudio que no involucra efectos espaciales [24]. Para poder capturar en un modelo toda la correlación y los efectos espaciales intrínsecos en los datos, es necesario definir como punto de partida, la matriz de contigüidad espacial.

1.2.1. Matrices de contigüidad espacial

Una matriz de contigüidad espacial expresa de manera conjunta la conectividad entre dos regiones. Como es natural, existe la necesidad primero de describir o determinar cuáles regiones están influyendo de cierta manera el comportamiento de la variable en cada unidad espacial. En primer lugar, se asume que se dispone de un mapa sobre el cual es posible determinar las observaciones en cada unidad regional (en este caso el mapa de la división política de Colombia). A ese mapa se denomina sistema espacial y se denota por S . En total, ese sistema cuenta con n unidades espaciales, y un proceso Y medido en cada una de esas unidades espaciales o regiones; es decir, se tiene $Y(R_i)$ en donde $\{R_1, R_2, \dots, R_n\}$ forma lo que se conoce en la práctica como un *lattice* o grilla de S . Una vecindad para cada unidad espacial R_i es definida por la colección de unidades R_j para las cuales $Y(R_j)$ está contenida en la forma funcional de la probabilidad condicional de $Y(R_i)$, condicionada a $\mathbf{Y} = Y(S) = \{Y(R_1), Y(R_2), \dots, Y(R_n)\}$, el proceso en todas las otras unidades [4], [8], [36]). Formalmente, el conjunto de vecinos de i se denota por J mediante

$$p(Y(R_i)|\mathbf{Y}) = p(Y(R_i)|Y(R_J)) \quad (1.1)$$

donde $Y(R_J)$ es el proceso $Y(\cdot)$ observado en todas las regiones j tales que pertenecen al conjunto J , y $\mathbf{Y} = Y(S)$ es el proceso en todas las regiones del sistema definido conformado por el sistema S . Teniendo en cuenta la estructura espacial en la definición de vecindad Anselin [4] propone la siguiente definición del conjunto de vecinos para la unidad espacial R_i :

$$\{j : p(Y(R_i)) \neq p(Y(R_i)|Y(R_j)) \text{ para } d_{ij} < \epsilon_i\} \quad (1.2)$$

donde d_{ij} es una medida de distancia entre R_i y R_j en un espacio determinado, y ϵ_i es un punto crítico de corte para cada unidad espacial R_i , probablemente el mismo para todas las unidades espaciales. La distancia métrica denotada por d_{ij} de forma muy general puede no ser euclidiana. Al definir de esta manera el concepto de vecindad, se debe incluir una definición adicional con respecto a la estructura espacial; combinar la noción de dependencia estadística con la noción espacial. Por lo que esta definición no evita que las unidades espaciales que no cumplen el criterio de distancia dejen de influenciar sobre la probabilidad condicional de $Y(R_i)$, estos no son considerados los vecinos más cercanos. El resultante conjunto de vecinos para cada unidad puede ser representado en una estructura espacial, asociado a una matriz de conectividad [4].

Las primeras medidas de dependencia espacial fueron propuestas por Moran [53] y Geary [29], y estas fueron basadas en la noción de contigüidad binaria entre las unidades espaciales [60]. De acuerdo con esto, la estructura de vecindad se expresa por medio de valores 0 o 1. Si dos unidades espaciales tienen algún borde en común de longitud diferente de cero, ellas son consideradas contiguas, y entonces el valor de 1 es asignado [4]. Un buen uso de la introducción de estos coeficientes, es probar si existe correlación entre los

residuales de un modelo de regresión. Sin embargo, Cliff y Ord [14] generalizaron esas definiciones.

Esta definición de contigüidad presentada, asume por su estructura la existencia de un mapa, sobre el cual las fronteras de cada unidad espacial pueden ser observadas (un lattice). Para un arreglo irregular de las áreas de las unidades espaciales, como en el caso colombiano dividido por su convención política (departamentos), la construcción de la matriz de contigüidad espacial resulta relativamente sencilla; este tipo de matriz, es una de las consideradas para el trabajo posterior en la aplicación. En el caso en que las unidades espaciales se refieren a un mapa regular (como una cuadrícula), o a un arreglo de puntos irregulares, la determinación de contigüidad no es única [4].

Se considera un mapa regular y se asocia una letra a cada área, indicando una unidad espacial, como en la figura 1.1.

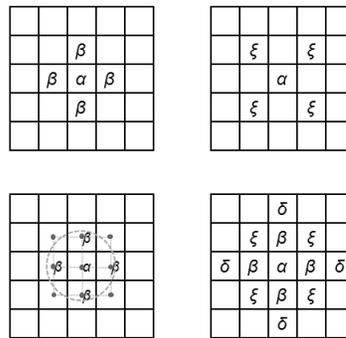


FIGURA 1.1. Contigüidad en un mapa regular

La noción de vecindad en una grilla regular, puede ser establecida de muchas maneras. Anselin [4] explica algunas de ellas como sigue: Un borde común entre la unidad α y sus vecinos circundantes puede ser considerado de diferentes maneras. Por ejemplo, pueden considerarse contiguos cuando hay un borde en común entre ellos, como el caso de la primera, en donde α es vecino de los β 's. También, puede considerarse contiguo cuando se comparte algún vértice, como el caso de la segunda ilustración, en donde los ξ son contiguos a la α , o también, una combinación de ambos criterios de contigüidad puede ser usada; estas situaciones reciben el nombre de torre, alfil y reina, en analogía del juego de ajedrez.

Cuando las unidades espaciales constan de puntos, como las ciudades en una jerarquía urbana, que son regular o irregularmente espaciados sobre el sistema, el significado de contigüidad puede ser derivado de la noción de camino más corto sobre una red formada por conectar puntos, tales como la red formada por la línea que une todos los nodos en la ilustración 3 de la gráfica 1.1 (este es un enfoque econométrico espacial, sin embargo, en dicha situación, debe tenerse en cuenta otros tipos de técnicas, como las geo estadísticas, véase [23]. Los puntos sobre la red son considerados como vecinos siempre y cuando estén dentro de una distancia máxima el uno del otro. Por ejemplo, los nodos llamados β caen dentro de una distancia demarcada por un radio de magnitud d desde el centroide α , y pueden ser considerados como contiguos por esta distancia crítica. La noción de vecindad, también es definida en [71], mostrando las mismas características dadas en [4].

De manera similar, otro orden de contigüidad puede ser considerado. Esto se logra de manera recursiva, definiendo una contigüidad de orden k cuando las unidades espaciales son contiguas de primer orden a una unidad espacial contigua de orden $k - 1$. Por ejemplo,

las celdas llamadas por ξ y δ en la figura 1.1 son contiguas de segundo orden a α de acuerdo al criterio torre, ya que ellos son contiguos de primer orden a β .

Desde el enfoque matemático y para hacer uso de la correlación existente en una vecindad particular teniendo en cuenta que la autocorrelación espacial depende de la interdependencia entre dos o más regiones cuyo valor en el proceso de interés tenga algún tipo de asociación, se usa la matriz de pesos espaciales que tiene la siguiente forma general:

$$\mathbf{W} = \begin{bmatrix} 0 & w_{12} & \dots & w_{1n} \\ w_{21} & 0 & \dots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & \dots & 0 \end{bmatrix} \quad (1.3)$$

en donde cada una de las entradas de la matriz, define, por ejemplo en el caso de w_{ij} la fuerza de la dependencia entre las unidades espaciales indexadas por i y j . La construcción de la matriz \mathbf{W} no es única y no siempre es simétrica. En el caso de la propuesta de Moran y Geray [53] y [29], $w_{ij} = 1$ si las regiones i y j son contiguas, y es cero en caso contrario, según el criterio de contigüidad escogido. Esta matriz de pesos espaciales \mathbf{W} presenta características importantes que deben ser cuidadosamente consideradas, según se defina el criterio de contigüidad. Por ejemplo, en el caso de una contigüidad física, la matriz de pesos espaciales no tendrá la capacidad de captar correlaciones con unidades espaciales que estén aisladas y que por la naturaleza de la variable medida, la correlación presente entre ellas, refleja un alto grado de dependencia.

El concepto de contigüidad binaria fue generalizado por Cliff y Ord [15] como se mencionó anteriormente. Esta propuesta incluye una medida general de la interacción potencial entre dos unidades espaciales [4]; consiste en usar una combinación de la medida de distancia y el tamaño relativo de la frontera entre dos unidades espaciales. Los valores resultantes serán obviamente asimétricos, a menos que ambas unidades espaciales tengan el mismo tamaño y proporción del respectivo límite. La propuesta formalmente es:

$$w_{ij} = d_{ij}^{-a} \beta_{ij}^b \quad (1.4)$$

con d_{ij} la distancia definida entre la unidad espacial i y j , β_{ij} es la proporción del límite de la unidad i el cual está en contacto con la unidad j , y a y b son dos parámetros a estimar. De manera similar, Dacey [19] sugirió formar los pesos que también tomaran en cuenta el área relativa de las unidades espaciales:

$$w_{ij} = d_{ij} \alpha_j \beta_{ij} \quad (1.5)$$

con d_{ij} el factor binario de contigüidad, α_j es la parte de la unidad i en el área total de todas las unidades espaciales presentes en el sistema, y β_{ij} como se usó en el anterior criterio.

Varios autores han sugerido de otra forma la construcción de los pesos, como es el caso de Dacey [19], Anselin [2] y Bodson y Peeters [9]; estos últimos proponen construir los pesos relacionados directamente con el fenómeno que tratan.

1.2.2. Operador de retardo espacial

En el contexto de series de tiempo, un operador de retardo es usado para desplazarse o rezagar la variable por uno o más períodos de tiempo. Es el caso del operador L , por ejemplo $y_{t-k} = L^k y_t$, en donde y_j es la observación de la variable en el tiempo j . En el caso espacial, este enfoque tan simple no puede ser usado, por las múltiples direcciones a donde puede rezagarse o desplazarse la variable. Si se considera por ejemplo una malla regular, el proceso $Y(\cdot)$ observado en el lugar (i, j) puede ser rezagado o desplazado a las siguientes direcciones, usando un criterio de contigüidad físico de primer orden:

- Usando un criterio de contigüidad *torre* o *rook*, a

$$y_{i-1,j}, y_{i,j-1}, y_{i+1,j}, y_{i,j+1}$$

- Usando el criterio del *alfil* o *bishop*, a

$$y_{i-1,j-1}, y_{i+1,j-1}, y_{i+1,j+1}, y_{i-1,j+1}$$

- Para un criterio de contigüidad *reina* o *queen*, el número total de posibles direcciones se incrementa a ocho, según la siguiente ilustración

$i - 1, j + 1$	$i, j + 1$	$i + 1, j + 1$
$i - 1, j$	i, j	$i + 1, j$
$i - 1, j - 1$	$i, j - 1$	$i + 1, j - 1$

En el caso de mapas o sistemas regulares, el número de vecinos resultante depende del criterio de contigüidad que se utilice. Cuando el sistema es irregular, se puede encontrar una gran cantidad de posibles direcciones. Por esto el número de parámetros asociados con las diferentes vecindades puede llevar a un análisis algo complicado. Además, si se tiene una gran cantidad de datos y un sistema o estructura regular, podría suceder que los grados de libertad restantes no serán suficientes para permitir una estimación eficiente de estos parámetros. Este problema se soluciona si se considera la suma de los pesos de todos los valores pertenecientes a una clase de contigüidad dada, en vez de tomar cada una de ellas individualmente. Los términos de la suma son obtenidos multiplicando el proceso observado en la vecindad i , con los pesos asociados de la matriz de pesos espaciales. De manera formal:

$$L^k Y(R_i) = \sum_j w_{ij} Y(R_j) \text{ para todo } j \in J_i \tag{1.6}$$

en donde L^k es el operador de rezago asociado con la clase de contigüidad k , j es el índice de las observaciones pertenecientes a la vecindad de la unidad R_i de la clase de contigüidad k (J_i es el conjunto de índices de los vecinos de la región R_i), y w_{ij} son los pesos espaciales. En notación matricial, se tiene

$$L^k \mathbf{Y} = \mathbf{W} \mathbf{Y} \tag{1.7}$$

con $\mathbf{Y} = Y(S) = \{Y(R_1), Y(R_2), \dots, Y(R_n)\}$ como antes, obteniendo que \mathbf{W} es el operador de retardo espacial. La matriz de pesos \mathbf{W} usada en el retardo espacial es frecuentemente estandarizada por filas, es decir, para cada R_i , $\sum_j w_{ij} = 1$. Por lo tanto, el retardo espacial puede ser interpretado como un promedio de pesos del proceso aleatorio $Y(\cdot)$ en las vecindades de R_i .

1.2.3. Autocorrelación espacial

La dependencia espacial puede ser considerada como la existencia de una relación funcional entre lo que pasa en un punto del espacio y lo que pasa en los demás [4]. En términos generales, la autocorrelación presente en una variable puede conducir en el caso de la proposición de un modelo al no cumplimiento de uno de los supuestos básicos, como la no correlación entre los residuales. Si esto ocurre, el modelo que se asume de la varianza podría resultar insuficiente, y en este caso, la variabilidad captada por el modelo es menor que la que está presente en los datos, produciendo el fenómeno de la sobredispersión, explicado brevemente en el capítulo siguiente. De acuerdo con esto, el objetivo ahora es presentar unas pruebas de no correlación entre las observaciones. Para un criterio de contigüidad dado, en este caso binario, Moran [53] y Geray [29] propusieron los siguientes coeficientes de autocorrelación:

$$I = \frac{n \sum_{i,j} w_{ij} (Y(R_i) - \bar{Y})(Y(R_j) - \bar{Y})}{\left(\sum_{i,j} w_{ij} \right) \sum_i (Y(R_i) - \bar{Y})^2} \quad (1.8)$$

$$C = \frac{(n-1) \sum_{i,j} w_{ij} (Y(R_i) - Y(R_j))^2}{\left(2 \sum_{i,j} w_{ij} \right) \sum_{i,j} (Y(R_i) - Y(R_j))^2} \quad (1.9)$$

respectivamente, donde $Y(R_i)$, $i = 1, 2, \dots, n$ son las observaciones del proceso $Y(\cdot)$ relacionadas con cada una de las n regiones o unidades espaciales, y, para este caso, $w_{ij} = 1$ si $i \neq j$ y si R_i y R_j son contiguos, y $w_{ij} = 0$ en otro caso. Estas estadísticas, en este caso son mostradas con el objetivo de probar si $\rho = 0$ en procesos que tienen la forma funcional

$$Y(R_i) = \beta_0 + \sum_{j=1}^p x_{ij} \beta_j + \rho \sum_{j=1}^n w_{ij} Y(R_j) + \epsilon_i \quad (1.10)$$

para un modelo lineal (o modelos con estructura más general, con $g(\cdot)$ una función de enlace que relaciona la media del proceso de interés con las covariables), β_i con $i = 0, 1, \dots, p$ parámetros a estimar, x_{ij} la entrada i, j de la matriz diseño conocida (en esta notación, se hace referencia al proceso j -ésimo que hace de covariable $X_j(\cdot)$ en cada una de las áreas o unidades espaciales del sistema S , es decir, $x_{ij} = X_j(R_i)$). En general, esos coeficientes han sido pensados como puntos de partida para la prueba $\rho = 0$ en modelos como el siguiente:

$$\mathbf{Y} = \boldsymbol{\mu} + \rho \mathbf{W}(\mathbf{Y} - \boldsymbol{\mu}) + \boldsymbol{\epsilon} \quad (1.11)$$

mencionado con más detalle adelante. Una demostración de normalidad asintótica para este par de estadísticas fue trabajada por Cliff y Ord [14] y por Sen [62]. La conclusión de estos estudios, es que asumiendo normalidad, no se tiene una buena aproximación para regiones conectadas de una manera escasa [60]. El apéndice A muestra en detalle apartes de inferencia para las estadísticas en mención.

1.3. Modelos clásicos de la econometría espacial

Esta sección se ocupa de reseñar de manera general los aspectos más importantes de los modelos clásicos usados en datos con estructura espacial en donde se tiene el supuesto de normalidad en la variable respuesta. Se hace mención de los modelos autocovariantes o autorregresivos y los modelos econométricos espaciales basados en modelos lineales. En el apéndice se relaciona la estimación vía máxima verosimilitud de los parámetros del modelo que se conoce como modelo SAR. Todos los modelos revisados en este capítulo y en este escrito incorporan autocorrelación espacial usando matrices de vecindad o de contigüidad espacial vistas en la sección anterior.

1.3.1. Modelos Autorregresivos - AR

Los modelos autorregresivos espaciales pueden escribirse de manera general por la siguiente expresión:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{Y} + \boldsymbol{\epsilon} \quad (1.12)$$

con $\mathbf{Y} = \{Y(R_1), Y(R_2), \dots, Y(R_n)\}$, $\mathbf{X} = [\mathbf{1}, \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p]$ con $\mathbf{X}_j^T = \{X_j(R_1), X_j(R_2), \dots, X_j(R_n)\}$, una matriz diseño conocida ($\mathbf{X} = [X_j(R_i)]$ con $i = 1, 2, \dots, n$ y $j = 0, 1, 2, \dots, p$, otros procesos subyacentes, observados sobre las regiones R_i que conforman S ; a menudo $X_0(R_i) = 1$ para todo R_i), $\boldsymbol{\beta}$ un vector de parámetros desconocidos, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. En este caso, la matriz de covarianzas del modelo es

$$\mathbf{V} = \sigma^2[(\mathbf{I} - \rho\mathbf{W})^T(\mathbf{I} - \rho\mathbf{W})]^{-1}. \quad (1.13)$$

En este tipo de modelos, el término autorregresivo definido por $\mathbf{A} = \mathbf{W}\mathbf{Y}$ es independiente de los predictores y así es apropiadamente aplicado cuando una alta correlación es el resultado de procesos endógenos que no están relacionados con las variables predictivas [43]. Los siguientes modelos mencionados son los CAR (Modelos Condicionales Autorregresivos) y SAR (Modelos Simultáneos Autorregresivos). Una distinción fundamental con un impacto significativo sobre las estimaciones (desde el enfoque clásico) y las estrategias de inferencia es la diferencia entre dichos procesos. El modelo condicional es basado en la especificación de una probabilidad condicional, mientras que el modelo simultáneo es expresado como una probabilidad conjunta. Ambos modelos son presentados más a menudo en una forma autorregresiva, es decir, donde el valor de una variable en un punto es relacionada a sus valores en el resto del sistema espacial [4].

Los modelos espaciales autocovariantes con estructura gaussiana toman una de estas dos posibles formas, dependiendo de cómo la estructura de error correlacionada espacialmente es especificada [47]. De hecho, por la forma de construcción de los modelos, si la estructura de correlación es considerada en la variable respuesta se está en el caso de un modelo CAR, y si la correlación también se considera en los residuales se está en el caso de un modelo SAR.

1.3.2. Modelos Condicionales Autorregresivos - CAR

El modelo CAR fue propuesto por Besag [8] para obtener una alternativa para el método de estimación no lineal propuesto por Whittle [68] en los modelo SAR mencionados en la siguiente sección. El modelo consiste en una relación lineal entre una esperanza

condicional de la variable dependiente y sus valores en el resto del sistema S , es decir

$$E(Y(R_i)|Y(R_{(-i)})) = \rho \mathbf{W}_i \mathbf{Y} \quad (1.14)$$

en donde $Y(R_{(-i)}) = \{Y(R_j) : j \neq i\}$. En particular, Besag [8] define el modelo CAR como

$$E(Y(R_i)|Y(R_{(-i)})) = \mu_i + \rho \sum_{k=1}^n w_{ik}(Y(R_k) - \mu_k) \quad (1.15)$$

en donde $\mu_i = \sum_{j=1}^p \beta_j x_{ij}$. En tal caso, el modelo CAR (*Conditional Autoregressive Models*) puede ser escrito como

$$Y(R_i)|Y(R_{(-i)}) \sim N \left(\mu_i + \rho \sum_{k=1}^n w_{ik}(Y(R_k) - \mu_k), \sigma_i^2 \right)$$

si $\{Y(R_i) : R_i \in S\}$ es un proceso aleatorio normal. Si $\sigma_i^2 = \sigma^2$ para todas las unidades espaciales R_i , la matriz de covarianzas para \mathbf{Y} es

$$\mathbf{V}_C = \sigma^2(\mathbf{I} - \rho \mathbf{W})^{-1} \quad (1.16)$$

en donde \mathbf{W} es simétrica. Si se tiene que $\text{Var}(Y(R_i)|Y(R_{(-i)})) = \sigma_i^2$, la matriz de varianzas del vector Y incondicionado es $(\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{\Sigma}$ con $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$. Se puede encontrar la distribución incondicionada del proceso aleatorio \mathbf{Y} , es decir, se puede demostrar que $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I} - \rho \mathbf{W})^{-1})$ en el caso particular de igualdad de varianzas. Definiendo el vector $\boldsymbol{\epsilon}$ como

$$\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \rho \mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

se encuentra fácilmente que $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2(\mathbf{I} - \rho \mathbf{W})$ y como resultado importante $\text{Cov}(\boldsymbol{\epsilon}, \mathbf{Y}) = \sigma^2 \mathbf{I}$. Para revisar estos resultados más detenidamente, se sugiere ir a [60]. Nótese que para esta estructura, \mathbf{V}_C determina \mathbf{W} y además se establece que hay independencia entre los residuales y el vector \mathbf{Y} . Dada esta definición de $\boldsymbol{\epsilon}$, y con la estructura encontrada, finalmente se puede definir el modelo CAR de la siguiente manera [43]:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \rho \mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\epsilon} \quad (1.17)$$

con todas las características encontradas para los respectivos componentes del modelo. Dada la estructura de la varianza del proceso \mathbf{Y} , se observa que la matriz \mathbf{W} que define la estructura de vecindad y de rezago espacial, debe ser simétrica. Consecuentemente, los modelos CAR no son adecuados cuando los procesos trabajan con distancias no Euclidianas, resultando así en matrices de covarianza asimétricas. En tales situaciones, los modelos relacionados SAR (*simultaneous autoregressive models*) son una mejor opción, ya que su matriz \mathbf{W} no necesita ser simétrica [24]. En general, los modelos CAR solo consideran efectos de vecindad de primer orden, mientras que los modelos SAR permiten por su recursividad, efectos de vecindad de mayor orden [43].

1.3.3. Modelos Simultáneos Autoregresivos - SAR

Los modelos SAR datan del trabajo de Whittle [68], en el cual, campos aleatorios sobre una malla regular, fueron expresados en la forma de una ecuación en diferencias estocástica sobre las coordenadas del lattice. Besag [8] define un proceso simultáneo autorregresivo como el conjunto de n ecuaciones autorregresivas

$$Y(R_i) = \mu_i + \rho \sum_{k=1}^n w_{ik}(Y(R_k) - \mu_k) + \epsilon_i$$

para $i = 1, 2, \dots, n$, donde $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ son variables aleatorias normales independientes con media cero y varianza σ^2 . Whittle [68] describe el proceso como una autorregresión lineal en el plano bajo la hipótesis que se dispone de una grilla regular en la forma

$$L(T_s, T_t)Y_{s,t} = \epsilon_{s,t} \tag{1.18}$$

donde T_s y T_t son operadores de translación definidos por

$$T_s Y_{s,t} = Y_{s+1,t}, \quad T_t Y_{s,t} = Y_{s,t+1}$$

y

$$L(T_s, T_t)Y_{s,t} = \sum_j \sum_k \alpha_{jk} T_s^j T_t^k$$

con $j, k = \dots, -1, 0, 1, \dots$, como índices espaciales, Y como el proceso bajo consideración, α_{jk} como parámetros que definen la estructura, y los ϵ como variables aleatorias independientes normalmente distribuidas. En forma matricial, esta ecuación puede escribirse como:

$$\mathbf{Y} = \rho \mathbf{WY} + \boldsymbol{\epsilon} \tag{1.19}$$

con \mathbf{Y} y $\boldsymbol{\epsilon}$ vectores del proceso y términos de errores respectivamente, y \mathbf{W} la matriz correspondiente de pesos espaciales [4]. Los modelos SAR pueden tomar tres formas diferentes, dependiendo sobre en donde el proceso autorregresivo se asume que ocurre. El primero asume que el proceso autorregresivo ocurre en el proceso de interés $Y(\cdot)$. El modelo toma la forma

$$\mathbf{Y} = \rho \mathbf{WY} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1.20}$$

que es equivalente a la expresión:

$$\mathbf{Y} = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\epsilon} \tag{1.21}$$

con ρ el parámetro autorregresivo, \mathbf{W} la matriz de pesos conectividad espacial y $\boldsymbol{\beta}$ un vector representando las pendientes asociadas con los predictores de las variables de la matriz diseño \mathbf{X} . El segundo, asume que la correlación espacial afecta ambos términos, el proceso aleatorio $Y(\cdot)$ y los términos de las variables explicativas. En dicho caso, otro término $\mathbf{WX}\boldsymbol{\gamma}$ aparece en el modelo, definiendo el rezago y la autocorrelación de las variables definidas en \mathbf{X} . Este modelo se escribe como

$$\mathbf{Y} = \rho \mathbf{WY} + \mathbf{X}\boldsymbol{\beta} + \mathbf{WX}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{1.22}$$

con $\boldsymbol{\gamma}$ un vector de parámetros a estimar. El tercer modelo, asume que el proceso autorregresivo ocurre solo en el término de error. En este caso, el modelo de regresión usual, es

complementado por el término $\lambda \mathbf{W}\boldsymbol{\xi}$, de la forma:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \lambda \mathbf{W}\boldsymbol{\xi} + \boldsymbol{\epsilon} \quad (1.23)$$

El término $\lambda \mathbf{W}$ representa la estructura espacial del término de error dependiente $\boldsymbol{\xi}$. [24]. De manera general, el modelo SAR se puede expresar mediante la ecuación

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \rho \mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\epsilon} \quad (1.24)$$

con $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. Se verifica fácilmente que la matriz de covarianzas del proceso aleatorio $Y(\cdot)$, es

$$\text{Var}(Y) = \mathbf{V}_S = \sigma^2 [(\mathbf{I} - \rho \mathbf{W})^T (\mathbf{I} - \rho \mathbf{W})]^{-1} \quad (1.25)$$

con \mathbf{W} una matriz de contigüidad espacial, posiblemente no simétrica [43]. Un resultado importante es $\text{Cov}(\boldsymbol{\epsilon}, \mathbf{Y}) = \sigma^2 [(\mathbf{I} - \rho \mathbf{W})^T]^{-1}$ [60]. Nótese que para un proceso SAR, la elección de \mathbf{W} para obtener \mathbf{V}_S puede no ser única, en contraste con el modelo CAR. También se puede ver que existe una correlación entre el proceso de interés Y y los errores, determinado por los pesos espaciales escogidos, a diferencia del modelo CAR. Para una comparación más extensa entre los dos modelos SAR y CAR, se sugiere ver [34], [35], [36] y [37].

Anselin [4] hace una mención más notoria a los modelos SAR. De manera general, propone un modelo econométrico espacial extendiendo la idea desde las series de tiempo, por ejemplo una adaptación de un proceso de promedios móviles llevado a la estructura

$$\mathbf{Y} = \boldsymbol{\epsilon} + \rho \mathbf{W}\boldsymbol{\epsilon} \quad (1.26)$$

con \mathbf{Y} como antes, $\boldsymbol{\epsilon}$ es un vector de variables aleatorias independientes y \mathbf{W} es la matriz usual de contigüidad espacial. Para una propuesta más general, escribe que el punto de partida es dado por

$$\mathbf{Y} = \rho \mathbf{W}_1 \mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (1.27)$$

$$\boldsymbol{\epsilon} = \lambda \mathbf{W}_2 \boldsymbol{\epsilon} + \boldsymbol{\xi} \quad (1.28)$$

con $\boldsymbol{\xi} \sim N(\mathbf{0}, \boldsymbol{\Omega})$, y los elementos de la diagonal de la matriz de covarianzas de los errores $\boldsymbol{\Omega}$ son

$$\boldsymbol{\Omega}_{ii} = h_i(\mathbf{z}\boldsymbol{\alpha}) \quad \text{con} \quad h_i > 0$$

en donde $\boldsymbol{\beta}$ es un vector de parámetros asociado a los procesos independientes $X_j(\cdot)$ no rezagados, \mathbf{X} es una matriz considerada diseño como antes, ρ y λ los coeficientes para el proceso rezagado y para el término de error rezagado respectivamente. Se observa que los elementos de la diagonal pueden ser modelados por un grupo de variables exógenas \mathbf{z} . Nótese además que las matrices de contigüidad \mathbf{W}_1 y \mathbf{W}_2 pueden no ser iguales, permitiendo una definición diferente de cada proceso espacial en el proceso de interés y en el término de ruido. Si el tamaño del vector $\boldsymbol{\beta}$ es $p + 1$, y el del vector $\boldsymbol{\alpha}$ es $k + 1$, en general, se tienen $3 + k + p$ parámetros que estimar. Siguiendo esta estructura, se pueden definir cuatro modelos de regresión espacial, según los valores de cero de algunos de estos parámetros:

- Para $\rho = 0$, $\lambda = 0$, $\boldsymbol{\alpha} = \mathbf{0}$ ($p + 2$ restricciones), se obtiene el modelo clásico de regresión lineal, sin efectos espaciales.

- Para $\lambda = 0$ y $\boldsymbol{\alpha} = \mathbf{0}$ ($p + 1$ restricciones) se obtiene el modelo

$$\mathbf{Y} = \rho \mathbf{W}_1 \mathbf{Y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (1.29)$$

modelo espacial autorregresivo, que es el mismo modelo mostrado en la ecuación (1.20). Como un caso más general de este, se puede asociar el modelo anteriormente denominado mixto, en donde participaba la matriz \mathbf{W} rezagando los procesos subyacentes $X_j(\cdot)$.

- Para $\rho = 0$ y $\boldsymbol{\alpha} = \mathbf{0}$ ($p + 1$ restricciones), se tiene el modelo que coincide con el tercer modelo mostrado anteriormente (expresión (1.23) también conocido como SAR residual) cuya expresión Anselin [4] la muestra como

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + (\mathbf{I} - \lambda \mathbf{W}_2)^{-1} \boldsymbol{\xi} \quad (1.30)$$

es decir, el modelo de regresión lineal con efectos espaciales autorregresivos en el término del error.

- Para $\boldsymbol{\alpha} = \mathbf{0}$, (p restricciones) se tiene:

$$\mathbf{Y} = \rho \mathbf{W}_1 \mathbf{Y} + \mathbf{X} \boldsymbol{\beta} + (\mathbf{I} - \lambda \mathbf{W}_2)^{-1} \boldsymbol{\xi} \quad (1.31)$$

modelo espacial autorregresivo mixto con una estructura espacial en el término del error. Nótese que si $\boldsymbol{\alpha} = \mathbf{0}$ en cualquier caso, se está en la situación de $h = \sigma^2$, lo que implica homocedasticidad. (Para esto, ver [4] y [24]). Cuatro más especificaciones pueden ser obtenidas permitiendo heterocedasticidad de una forma específica (es decir, un valor específico de $h(\mathbf{z}\boldsymbol{\alpha})$) en cada uno de los modelos anteriores. Los diferentes modelos son construidos imponiendo restricciones sobre la forma general (1.27) y (1.28), como en los casos anteriores [4]. En el apéndice B se muestra brevemente el procedimiento para la estimación por máxima verosimilitud de los parámetros que influyen en el modelo que Anselin [4] llamó modelo de regresión lineal espacial.

En adelante, para mantener la notación seguida por los principales autores, se hace $Y(R_i) =: Y_i$, para $R_i \in S$ asimismo para cualquier proceso o covariable $X(\cdot)$. Los siguientes capítulos muestran en resumen los principales resultados de la sobredispersión en los modelos lineales generalizados y la estadística Bayesiana, método por el cual estimamos eventualmente los parámetros de los modelos considerados.

1.4. Aplicación: Modelación espacial del volumen de atención hospitalaria en Colombia

1.4.1. Análisis descriptivo y exploratorio

Además de considerar las estadísticas mencionadas en las secciones anteriores, se introducen algunas metodologías gráficas o exploratorias de correlación univariadas. Por análisis descriptivo o exploratorio de datos, comúnmente conocido como AED, se entiende el conjunto de herramientas gráficas y descriptivas usadas para el descubrimiento de patrones de comportamiento en los datos y el establecimiento posible con la menor estructura posible [66].

En Anselin [5] se explica que una de las tres formas en que la estadística espacial contribuye a las ciencias sociales, es la aplicación de técnicas descriptivas conocidas como análisis exploratorio de datos espaciales (AEDE) y la visualización de los mismos en un enfoque inductivo para descubrir patrones, provocando hipótesis y sugiriendo asociaciones. Considera asimismo que AEDE es un subconjunto de todas las herramientas AED.

Como primera medida se realiza un análisis descriptivo no espacial de la variable respuesta considerada en este estudio. Con base en la información disponible en la Encuesta Nacional de Salud del 2007 para Colombia, se presenta una variable que mide el volumen de atención en las IPS (Institución Prestadora de Servicios; es decir, todos los centros, clínicas y hospitales donde se prestan los servicios médicos, bien sea de urgencia, de consulta o de hospitalización, ver [http : //www.elcolombiano.com](http://www.elcolombiano.com)), de manera que se agrupa una estimación para el total de estas mismas, ya sean públicas, privadas o mixtas, llamada en el escrito variable VOLAT (volumen de atención). Un evento de consulta externa se define como cualquier tipo de servicio recibido en la IPS de manera ambulatoria (consulta por médico general o especialista, diagnósticos preventivos y promocionales que incluyen hasta cirugía ambulatoria). Están considerados como eventos en urgencias todas las consultas definidas como tales y en hospitalización se consideraron como eventos los egresos hospitalarios (Encuesta Nacional de Salud, 2007 - Resultados Nacionales). Mediante una razón de la variable VOLAT y el total de habitantes por departamento (datos demográficos proyectados para el 2007), se determina la variable respuesta para nuestro modelo, la cual contiene información del número de eventos en que se hace uso de la IPS mensualmente por cada mil habitantes (variable IDVOLAT).

La proporción estimada de hogares en el país en el que residen personas que han sufrido desplazamiento provenientes de otro departamento desde enero del año 2003 variable Displc. También se cuenta con una variable que mide el porcentaje de hogares en cada departamento que han sido afectados por riesgos naturales entre 2005 y 2007 (variable RNtr).

Se dispone también de información sobre el porcentaje de la población que tiene necesidades básicas insatisfechas para el año 2004 (variable NBI), sobre los recursos otorgados para la educación y atención integral para niños pequeños (en miles) por familias, brindados por el gobierno central a las diferentes administraciones departamentales o municipales en el año 2005 (variable Rec) y sobre el porcentaje de niños menores de un año que han recibido la tercera dosis de la vacuna contra el polio aplicada en el año 2004 (variable Vac). Por último, por parte de la Asociación Colombiana de Clínicas y Hospitales, se cuenta con una variable que mide el número total de instituciones prestadoras de servicios de salud, bien sea naturaleza privada, mixta o pública, para el año 2010 variable TotIPS. Con base en esta última, se construye una variable indicadora que muestra el comportamiento de las IPS en cada departamento, según su naturaleza pública o privada, manteniendo las mixtas como privadas (variable IndIPS).

Cada una de estas variables está medida en los 33 departamentos, incluyendo Bogotá. La figura 1.2 muestra los gráficos de densidad empírica y cajas de la variable IDVOLAT.

La figura 1.2 muestra que no hay departamentos cuyo índice de volumen de atención sea mayor que el promedio (candidatos a valores extremos). El diagrama de cajas y el gráfico de densidad muestran un leve sesgo negativo, indicando que hay algunos departamentos con en valor del índice mucho más bajo que el promedio. Cabe mencionar que el número de eventos de atención por cada 1000 habitantes en Colombia es en promedio 214 y su desviación estándar es de 4.81 indicando una dispersión relativa de 47.21 %. Esta variabi-

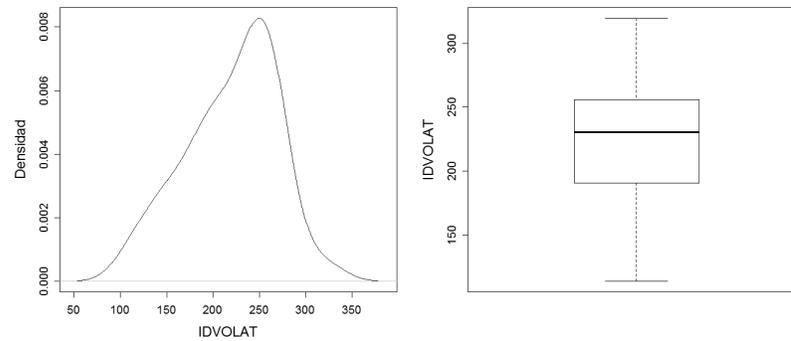


FIGURA 1.2. Gráfico de densidad y gráfico de cajas para la variable Volumen de Atención IDVOLAT

lidad en la información sugiere que la media de esta variable puede estar influenciada por el comportamiento de algunas otras. Los coeficientes de correlación lineal entre la variable de interés y las consideradas covariables son -0.43 , 0.20 , -0.59 , 0.49 , -0.39 , 0.17 y -0.43 para Displc, TotIPS, el índice mencionado IndIPS (razón entre el total de IPS públicas y el total de IPS privadas por departamento), Vac, NBI, Rec y RNtr respectivamente. Los gráficos de dispersión más destacados son mostrados en la figura 1.3

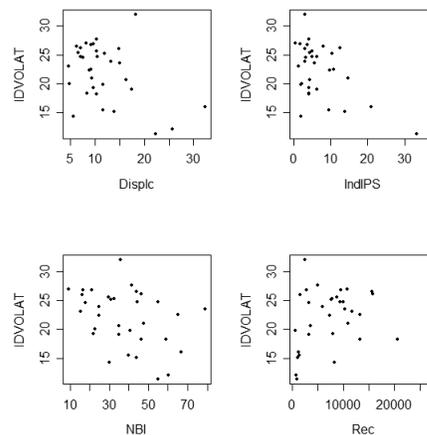


FIGURA 1.3. Gráficos de dispersión de la variable IDVOLAT contra las variables Displc, IndIPS, NBI y Rec

Según la figura 1.3 las variables Displc, IndIPS y NBI, muestran una relación lineal negativa con la variable IDVOLAT. Sin embargo la variable Rec, muestra comportamiento no lineal con respecto a la variable respuesta. En tal caso, habría que pensar en una eventual transformación. Continuamos ahora con en análisis exploratorio revisando la estructura espacial de esta variable.

Análisis descriptivo y exploratorio espacial

Como se había comentado en la introducción y al inicio del presente capítulo, los datos tienen un orden específico en el espacio. Cada una de las unidades muestrales pertenece a un espacio definido y limitado, lo que convierte los puntos muestrales en unidades espaciales. El primer gráfico que se muestra (figura 1.4), es un mapa de Colombia con su respectiva división política y administrativa, en donde cada departamento se ve represen-

tado por un tono de color distinto, que indica el rango en el que la variable de interés *IDVOLAT* toma valores.

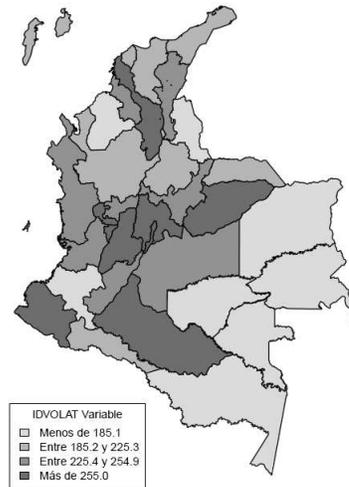


FIGURA 1.4. Distribución Colombiana de la variable *IDVOLAT*

El mapa de la figura 1.4 muestra el comportamiento de la variable a través del territorio colombiano. Se aprecia que hay indicios de correlación espacial, ya que al oriente y al sur del país se encuentran las tasas más bajas de volumen de atención en salud por habitantes.

Autocorrelación espacial - I de Moran

En secciones anteriores se definieron las estadísticas de Moran [53] y Geray [29] y en el apéndice A se muestra la de Getis [30] (Autocorrelación Local) para verificar posible presencia de correlación espacial en la información. Para la variable *IDVOLAT* que se comporta como en el mapa de la figura 1.4, se tienen los respectivos valores de las estadísticas 0.2686, 0.6850 y 0.03141. Las pruebas de hipótesis que contrastan la hipótesis nula de que no hay correlación espacial de variable de interés, es decir, I, C y G igual a cero para cada una, versus la alternativa que son diferentes de cero, usando la distribución normal como resultado asintótico, muestran un respectivo p - valor de 0.0048, 0.0034 y 0.0552. Sin embargo, se realiza la prueba por el método de aleatorización únicamente para el caso de Moran (100000 permutaciones), obteniendo un p valor de 0.0089.

La figura 1.5 muestra la densidad de la estadística I obtenida y la línea vertical marca el lugar en donde se encuentra el estadístico de este caso. También se puede ver en esa figura el diagrama de dispersión de Moran univariado. Este consiste en realizar un gráfico de dispersión simple, entre la variable Y (en este caso *IDVOLAT*) y la variable rezagada WY , donde W es la matriz de pesos espaciales; la matriz de contigüidad de primer orden totalizada por filas (el objetivo es revisar gráficamente si existe alguna dependencia lineal entre la variable de interés y el valor promedio de los vecinos más próximos).

El valor de la estadística de Moran en el diagrama se visualiza en la pendiente de la recta que atraviesa la nube de puntos. En ocasiones se prefiere estandarizar las variables para identificar con mayor facilidad si la autocorrelación espacial es debida a un pequeño conjunto de datos (posiblemente autocorrelación local) o por algún punto atípico. Aquí no se muestra este resultado ya que guarda el mismo comportamiento de la figura 1.5

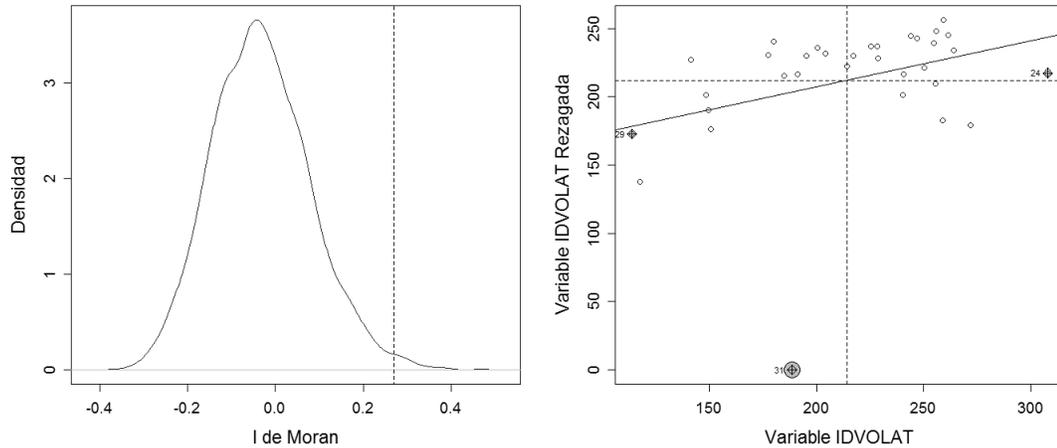


FIGURA 1.5. Densidad para la estadística I y Diagrama de Moran univariado para la variable $IDVOLAT$

Con la información que suministran los gráficos de la figura 1.5, se concluye que en esta variable hay presencia de autocorrelación espacial al menos a un 99 % de nivel de confiabilidad. El punto más influyente según el gráfico, que no pertenece al patrón que establecen los integrantes al sistema de vecindad establecido, es San Andrés, ya que no tiene vecinos de contigüidad física como los trabajados en este caso, y el valor de la variable rezagada es cero.

El software **GeoDa** es un paquete especialmente diseñado para trabajar con datos económicos espaciales; fue creado por Luc Anselin y se basa específicamente en análisis descriptivo y en herramientas AEDE, aunque incluye módulos para algunas regresiones clásicas y estimación de parámetros vía máxima verosimilitud. Algunos de los resultados obtenidos por nosotros en R son contrastados paralelamente con el software en mención, que puede ser descargado gratuitamente de la página <http://geodacenter.asu.edu/>.

1.4.2. Modelo espacial del volumen de atención hospitalaria en Colombia

En esta sección se dan a conocer los resultados obtenidos al aplicar el modelo de regresión espacial con las tres tipologías SAR dispuestas en (1.20), (1.22) y (1.23). Se considera la variable respuesta como el indicador del volumen de atención en cada departamento por cada 1000 habitantes (número de eventos por cada mil habitantes), y las variables explicativas como la proporción de desplazados $Ddisplc$, $TotIPS$, el llamado índice $IndIPS$ (razón entre el total de IPS públicas y el total de IPS privadas por departamento), Vac , NBI , Rec y $RNtr$. Vale la pena discutir sobre el impacto que genera cada una de estas variables en el volumen de atención hospitalaria en Colombia, y determinar entre ellas cuáles son los factores que más influyen en la determinación de su comportamiento. Entre los resultados obtenidos anteriormente, se determinan los coeficientes de correlación lineal entre cada una de estas variables y la respuesta. Sin embargo, se obtienen bajos coeficientes de correlación con respecto a las variables $TotalIPS$ y Rec , pero se puede observar que la relación, con la variable Rec no es lineal. La figura 1.6 muestra que dos relaciones viables entre esta y la variable respuesta es $y = \log(x)$ y $y = 1/x$, de forma que una transformación adecuada en cada uno de los casos es el logaritmo natural o el recíproco.

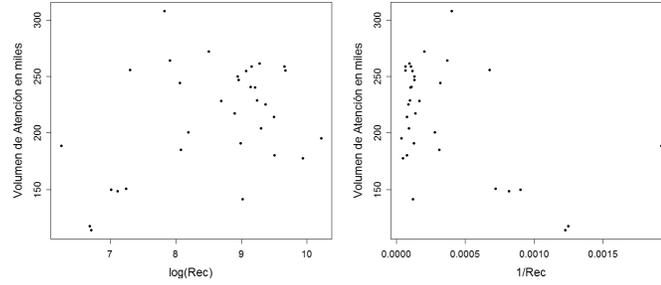


FIGURA 1.6. Diagrama de dispersión de $\log(Rec)$ y $1/(Rec)$ con la variable $IDVOLAT$

De acuerdo a esta nueva transformación, el coeficiente de correlación respectivamente de la variable transformada con respecto a la variable respuesta aumenta a 0.4016 y -0.5014 , y gráficamente la relación se puede observar lineal positiva y lineal negativa.

El primer modelo presentado es el modelo lineal normal, en donde se asume que los residuales y la variable respuesta siguen una distribución normal con media cero y μ respectivamente, y varianza común σ^2 . El modelo de la media es:

$$\mu = \beta_0 + \beta_1 \text{Displc} + \beta_2 \text{IndIPS} + \beta_3 \text{Vac} + \beta_4 \text{NBI} + \beta_5 1/(\text{Rec}) + \beta_6 \text{RNtr} \quad (1.32)$$

Las estimaciones de los parámetros están consignadas en la tabla 1.1

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	231.510	0.517	-1.676	0.500	0.361	-33.183	-1.230
<i>d.e</i>	54.317	1.567	1.288	0.560	0.652	20.739	0.550

TABLA 1.1. Estimación y desviación estándar de los parámetros del modelo (1.32)

En este caso la estimación para la varianza del modelo σ^2 es de 1316.7. La tabla 1.1 muestra que la mayoría de los parámetros que entran en juego no son significativos (solamente dos lo son). Sin embargo, el objetivo inmediato es determinar si existe algún tipo de correlación espacial. Mediante el análisis descriptivo concluimos que puede haber indicios de que la variable respuesta presente niveles de correlación; el objetivo ahora es revisar si los residuales de este modelo también lo están.

Modelo espacial SAR con rezago en el residual

Anteriormente se vio que la forma funcional que adquiere el modelo en este caso es (1.23). Entonces, además de tener la el modelo para la media como (1.32), la estructura del error es de la forma

$$\xi = \lambda \mathbf{W} \xi + \epsilon \quad (1.33)$$

Las estimaciones para este nuevo modelo, son mostrados en la tabla 1.2:

En este caso la estimación para la varianza de los errores es de 819.11. El parámetro de interés λ tiene una estimación de -0.7949 y una desviación estándar de 0.2493. Mediante el uso de las pruebas apropiadas (véase apéndice B), probamos la hipótesis nula de que $\lambda = 0$. En efecto, el valor del test de razón de verosimilitud es 3.6565, cuyo p valor es

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	157.956	1.177	-1.101	1.254	-0.352	-25.028	-0.757
<i>d.e</i>	40.150	1.057	0.931	0.411	0.488	15.253	0.397

TABLA 1.2. Estimación y desviación estándar de los parámetros del modelo con estructura de error (1.33)

de 0.055 (distribuido χ_1^2 bajo la hipótesis nula). Otra estadística de uso frecuente para probar esta hipótesis es la de Wald. El valor que toma es de 10.162 y el p valor respectivo es 0.0014 (al igual que el caso anterior distribuido χ_1^2 bajo la hipótesis nula). Estos resultados muestran que el parámetro que define la autocorrelación en los residuales es significativamente distinto de cero. Revisando las estimaciones de los parámetros de la tabla 1.2, se encuentra que en comparación a la tabla 1.1 hay más variables significativas, que desde el análisis descriptivo mostraban correlación con la variable. También se nota que las diferencias de las estimaciones de σ^2 en ambos modelos, en donde para el modelo SAR disminuye, mostrando así la reducción de la variabilidad de los residuales y un eventual mejor ajuste. El AIC en el primer modelo (Modelo Lineal Normal) toma el valor de 338.81, mientras que en el modelo SAR con rezago en el término del error es de 337.160, mostrando una pequeña baja tal vez despreciable.

Modelo espacial SAR con rezago en la variable respuesta

El siguiente modelo en consideración corresponde al especificado en la sección (1.3.3), más específicamente el de la ecuación (1.20), y tiene la siguiente estructura en el predictor lineal de la media

$$\mu = \beta_0 + \beta_1 \text{Displc} + \beta_2 \text{IndIPS} + \beta_3 \text{Vac} + \beta_4 \text{NBI} + \beta_5 1/(\text{Rec}) + \beta_6 \text{RNtr} + \rho A \quad (1.34)$$

con $A_i = \mathbf{W}_i \mathbf{Y} = \mathbf{W}_i \text{IDVOLAT}$. Los respectivos estimadores de los parámetros están en la tabla 1.3.

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	295.817	1.603	-1.170	0.733	-0.018	-67.456	-1.263
<i>d.e</i>	67.182	1.450	1.112	0.475	0.561	27.843	0.480

TABLA 1.3. Estimación y desviación estándar de los parámetros del modelo (1.34)

Como se aprecia, según esta estructura, las variables que resultan siendo significativas son $1/\text{Rec}$ y RNtr , al menos a un 90% de significancia. El valor estimado de σ^2 es 950.59, algo más bajo que el modelo lineal simple, pero más alto que el SAR residual (debe tenerse cuidado con la comparación directa con los anteriores, ya que la varianza para este modelo cambia por la estimación del ρ que es diferente de cero). La estimación del parámetro ρ es -0.3463 y su desviación estándar de 0.2006. Para probar la significancia del parámetro ρ se usa la estadística de razón de verosimilitud, la cual toma la forma de (B.26) y la estadística de Wald que se reduce a la forma (B.18). Respectivamente los valores que toman las estadísticas son 2.096 y 2.987 con p valores de 0.1476 y 0.0843. En ambos casos la conclusión con respecto a la decisión sobre la hipótesis cambia; para el caso del test de razón de verosimilitud no se rechaza y para el test de Wald se rechaza a un 90% de confiabilidad (son estadísticas distribuidas χ^2 con un grado de libertad), concluyendo que el parámetro que define la autocorrelación es significativo. El AIC para este caso es de

338.72 y se rechaza también la hipótesis de presencia de autocorrelación en los residuos. Nótese que el AIC aumenta con respecto al SAR residual, pero esta diferencia puede ser despreciable.

Modelo espacial SAR mixto

Se realiza también el modelo SAR mixto, en donde el predictor toma la estructura (1.22) explicado en la misma sección que los dos anteriores. En este sentido, el modelo permite rezagar además de la variable respuesta, el conjunto de las variables explicativas. Así, se estiman 15 parámetros, 6 más que en los casos anteriores. El modelo tiene forma estructural dada por

$$\begin{aligned} \mathbf{Y} &= \rho \mathbf{WY} + \mathbf{X}\boldsymbol{\beta} + \mathbf{WX}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad \text{con;} \\ \mathbf{X}\boldsymbol{\beta} &= \beta_0 + \beta_1 \text{Displc} + \beta_2 \text{IndIPS} + \beta_3 \text{Vac} + \beta_4 \text{NBI} + \beta_5 1/(\text{Rec}) + \beta_6 \text{RNtr} \\ \mathbf{WX}\boldsymbol{\gamma} &= \mathbf{W}[\gamma_1 \text{Displc} + \gamma_2 \text{IndIPS} + \gamma_3 \text{Vac} + \gamma_4 \text{NBI} + \gamma_5 1/(\text{Rec}) + \gamma_6 \text{RNtr}] \end{aligned}$$

Nótese que la última igualdad, hace referencia a las covariables rezagadas por la matriz de pesos. La tabla 1.4 muestra las estimaciones de cada uno de los parámetros para el predictor lineal sin las covariables rezagadas

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	218.873	2.017	0.364	1.072	-0.157	-42.206	-1.054
<i>d.e</i>	64.623	1.223	1.172	0.373	0.452	29.296	0.414

TABLA 1.4. Estimación y desviación estándar de los parámetros del modelo SAR mixto, variables explicativas

La tabla 1.5 contiene las estimaciones y las desviaciones de las variables rezagadas.

Parámetro	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
<i>Estim</i>	4.664	2.580	2.806	-0.965	-143.322	-1.538
<i>d.e</i>	2.434	3.077	0.623	0.984	65.901	0.906

TABLA 1.5. Estimación y desviación estándar de los parámetros del modelo SAR mixto, variables explicativas rezagadas

De ambas tablas, la 1.4 y 1.5, se obtiene que las variables Displc, Vac, RNtr, Displc rezagada, Vac rezagada, 1/Rec rezagada y RNtr rezagada son significativas. Sin embargo, el AIC no cambia mucho con respecto a los demás, y para este nuevo caso asume un valor de 336.47. El valor de ρ cambia como era de suponer ya que la estructura no es la misma, pasa a tomar un valor de -1.0578 con una desviación estándar de 0.2073, y con este, los respectivos valores de las estadísticas de pruebas de razón de verosimilitud y Wald, son 11.696 y 26.038 con p valores es de 0.0006 y 3.347×10^{-7} . El valor de la varianza de los errores, es dada en este caso por 502.37, muy baja en comparación a los demás modelos.

La tabla 1.6 muestra en resumen los valores de AIC, estimaciones de σ^2 , ρ y λ según corresponda, para los modelos considerados.

El cuadro 1.6 muestra que el modelo más adecuado es el SAR mixto, en donde la varianza de los residuales y el AIC son más bajos en comparación con los demás. Debe

Modelo	AIC	σ^2	ρ	λ
Normal Lineal	338.81	1316.7		
SAR Residual	337.16	819.11		-0.795
SAR variable respuesta	338.72	950.59	-0.346	
SAR mixto	336.47	502.37	-1.057	

TABLA 1.6. AIC , σ^2 , ρ y λ para los modelos considerados

tenerse en cuenta que el coeficiente estimado ρ es diferente de cero significativamente. No debe compararse directamente el valor de σ^2 en los modelos SAR variable respuesta y SAR mixto con los demás, ya que en ese caso, la estructura de la varianza del término del error depende de la matriz \mathbf{W} y de ρ ; y admitimos una comparación para los casos en que $\rho \approx 0$ o simplemente lo es, como en SAR Residual.

Para trabajar con un modelo CAR, como lo presentamos en la sección 3, se necesitan matrices de pesos simétricas; hasta aquí \mathbf{W} no lo es, ya que está totalizada por filas.

Modelo espacial CAR

Para este caso, se toma la matriz de contigüidad física de primer orden que es simétrica, como matriz de pesos. Así, la matriz no produce en la autocovariable un promedio de los vecinos más cercanos, sino una suma de ellos. Una manera de visualizar esto, es que en el modelo (1.17) la matriz de varianzas del vector de ruidos ϵ depende de ella. Por definición, una matriz de varianzas debe ser simétrica, lo que lleva a la restricción de simetría de la matriz $(\mathbf{I} - \rho\mathbf{W})^{-1}$ y a su vez de la matriz \mathbf{W} . En realidad, los modelos CAR y SAR están relacionados, pero los términos $\rho\mathbf{W}$, por razones técnicas ya mencionadas, no son idénticos. Sea el término $\rho\mathbf{W}$ de los modelos CAR denotado por \mathbf{K} y el término $\rho\mathbf{W}$ de los modelos SAR denotado por \mathbf{S} ; entonces, cualquier proceso SAR es un proceso CAR con $\mathbf{K} = \mathbf{S} + \mathbf{S}^T - \mathbf{S}^T\mathbf{S}$ (ver [38] y [24]). La estructura del modelo entonces, es tomada de la expresión (1.17), y puede verse como un modelo CAR de los errores correlacionados espacialmente, muy similar a un modelo SAR residual haciendo

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\nu} \quad (1.35)$$

y

$$\boldsymbol{\nu} = \lambda\mathbf{W}\boldsymbol{\nu} + \boldsymbol{\epsilon} \quad (1.36)$$

Las estimaciones de los parámetros son dadas en el cuadro 1.7

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	184.201	0.865	-1.262	1.016	-0.204	-25.825	-0.892
<i>d.e</i>	42.742	1.141	1.001	0.443	0.520	15.920	0.426

TABLA 1.7. Estimación y desviación estándar de los parámetros del modelo (1.35)

El valor estimado de λ que define la estructura (1.36) es -0.254 y su desviación estándar es 0.0993 . La hipótesis de interés en este caso es dada por $H_0 : \lambda = 0$ versus $H_0 : \lambda \neq 0$. Aquí, solo se usa el test de razón de verosimilitud, cuya estadística toma el valor de 2.990 y se contrasta con una χ_1^2 , lo que indica un p -valor de 0.083 , mostrando que se rechaza la hipótesis nula de que λ es cero (hay correlación espacial en los residuales). El cuadro 1.7

muestra que con esta nueva estructura de \mathbf{W} las variables Vac y RNtr son significativas a un 95 % de confiabilidad. La estimación de σ^2 es 821.05 y el AIC es de 337.83, resultados que sugieren un ajuste muy similar al SAR residual.

Otra manera de hacer comparaciones entre los modelos anteriormente trabajados, puede ser la generalización del coeficiente de determinación R^2 definido para una modelo lineal usual. El R^2 es un valor en un modelo lineal que resume la proporción de variabilidad resumida por el modelo. Es deseable estandarizar la definición del R^2 a modelos más generales, para los cuales el concepto de varianza residual no puede ser fácilmente definido, y la máxima verosimilitud es el criterio de ajuste [54]. El pseudo R^2 propuesto por Cox y Snell [17] y Magee [49] es dado por

$$R^2 = 1 - \left\{ \frac{L(0)}{L(\hat{\beta})} \right\}^{2/n} \tag{1.37}$$

donde $L(0)$ y $L(\hat{\beta})$ denota la verosimilitud del modelo nulo y bajo consideración, respectivamente, β el vector de parámetros que se consideran en el modelo. Esta definición de pseudocoefficiente de determinación tiene las siguientes propiedades:

- Es consistente con el R^2 clásico, es decir, en caso de una regresión lineal, coinciden.
- Es consistente con estimación de máxima verosimilitud como un método de estimación, es decir, las estimaciones de máxima verosimilitud de los parámetros del modelo maximizan R^2
- Es asintóticamente independiente del tamaño de la muestra.
- Se puede interpretar como la proporción de variación explicada.

[54]. Para el último caso mencionado, Nagelkerke en su escrito, explica las restricciones con un buen ejemplo. También presenta su propuesta de R^2 , para el caso en que se trabaje alguna eventual regresión logística, como una generalización del coeficiente de Cox y Snell. El pseudo R^2 de Nagelkerke es definido por la expresión

$$\bar{R}^2 = \frac{R^2}{\text{máx}(R^2)} \tag{1.38}$$

en donde $\text{máx}(R^2) = 1 - L(0)^{2/n}$ y R^2 definido como en la ecuación (1.37).

Sin embargo, en cada uno de los casos del presente escrito, ambos coeficientes, el de Nagelkerke y el de Cox y Snell coinciden y cada uno toma los valores mostrados en la tabla 1.8

Modelo	$-2 \log L$	R^2 de Nagelkerke
Normal Lineal	322.816	0.534
SAR Residual	319.160	0.563
SAR variable respuesta	320.720	0.583
SAR mixto	306.475	0.716
CAR error	319.826	0.574

TABLA 1.8. pseudo R^2 de Nagelkerke para los modelos considerados

La tabla 1.8 muestra que el mejor modelo o el que mejor ajusta los datos, según este criterio, es el SAR mixto, con un 71.63% de variabilidad explicada. Sin embargo, cabe mencionar que a diferencia de los demás modelos, este tiene seis parámetros de más. A pesar de esto, retirando del modelo las variables IndIPS y NBI, se reduce en 4 el número de parámetros a estimar, el pseudo R^2 se mantiene cerca, toma valores de 70.52%, y el AIC disminuye considerablemente con el mostrado en el cuadro 1.6, ya que toma un valor de 329.74.

1.4.3. Otras matrices de pesos espaciales

Ya en secciones anteriores se construyeron modelos autorregresivos definidos y presentados cada uno en la sección (1.3). Para ello se trabaja con la matriz de pesos espaciales \mathbf{W} encargada de recoger la información de los vecinos más cercanos; se usa una matriz de contigüidad física de primer orden totalizada por filas. En esta nueva sección se utilizan otras definiciones de matrices, esto es, se establece una nueva vecindad para cada uno de los departamentos. La primera opción es trabajar con una contigüidad física de segundo orden, y la otra, es establecer cada una de las regiones naturales del país (división política y cultural) como una clase de vecindad.

1.4.3.1. Matriz de pesos espaciales con contigüidad física de segundo orden

Se considera primero la matriz de pesos espaciales de contigüidad física de segundo orden. En este caso, se suponen vecinos, además de los ya establecidos por la contigüidad física de primer orden, los vecinos de primer orden de los vecinos ya considerados. Para esta nueva definición de matriz de pesos espaciales, los valores de la I de Moran, C de Geray y G de Getis, son 0.1057, 0.8227 y 0.03111 respectivamente. Los respectivos p valores de cada una las pruebas de interés (si es cero estadísticamente el valor de la estadística) son 0.0208, 0.0040 y 0.09868. En las tres se rechaza la hipótesis nula, al menos a un nivel de confiabilidad del 90%. Como es de suponer, la estructura de la ecuación que define cada modelo no cambia con respecto a los considerados en la sección anterior; únicamente, las estimaciones de los parámetros.

Modelo espacial SAR con rezago en el residual

La ecuación para la media del modelo SAR con rezago en el residual, es la misma (1.32) y para los términos de error (1.33)

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	228.824	0.555	-1.955	0.478	0.195	-34.446	-0.958
<i>d.e</i>	49.503	1.302	1.157	0.517	0.562	18.040	0.410

TABLA 1.9. Estimación y desviación estándar de los parámetros del modelo con estructura de error (1.33) - Matriz de contigüidad física de segundo orden

La tabla 1.9 muestra las estimaciones respectivas del caso. Haciendo una comparación con la tabla 1.2, se observa un cambio importante en cada uno de los parámetros estimados. Cabe mencionar también que las variables que son significativas son IndIPS, 1/Rec y RNtr, a un 90% de confiabilidad. También se tiene que la estimación del parámetro de interés

λ que define la estructura de correlación espacial para este es -0.665 con una desviación estándar de 0.541 . Nuevamente mediante las pruebas mostradas en el apéndice, se prueba la hipótesis de que $\lambda = 0$ (juegos de hipótesis (B.14)). Así, el valor del test de razón de verosimilitud (B.26) es 0.8321 cuyo p valor es de 0.3616 . La otra estadística de uso frecuente que se menciona para probar esta hipótesis es la de Wald ((B.19) en esta situación). El valor que toma es de 1.5091 y el p valor respectivo es 0.220 . Los resultados anteriores muestran que no se rechaza la hipótesis nula, es decir, λ es estadísticamente igual a cero (no hay correlación significativa en el modelo). La estimación de la varianza del residual que define el modelo como σ^2 es 982.95 , mientras que el AIC es 339.98 .

Modelo espacial SAR con rezago en la variable respuesta

La ecuación de la media es dada al igual que el caso anterior por la misma estructura del modelo correspondiente en la sección anterior, es decir, por la ecuación (1.34). Las estimaciones de cada parámetro están consignadas en la tabla 1.10

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	345.608	1.574	-0.789	0.490	0.136	-81.198	-1.397
<i>d.e</i>	78.733	1.438	1.218	0.471	0.558	31.717	0.482

TABLA 1.10. Estimación y desviación estándar de los parámetros del modelo (1.34) - Matriz de contigüidad física de segundo orden

De la tabla 1.10 se puede observar que algunas de las estimaciones de los parámetros cambian un poco en comparación a los mostrados en la tabla 1.3. Sin embargo, al igual que cuando se usó la matriz de contigüidad física de primer orden, las únicas variables que salen significativas son $1/Rec$ y $RNtr$, a un 95 % de confiabilidad. La estimación de σ^2 es 936.6 y el AIC es de 337.95 . Además la estimación de ρ es de -0.474 con una desviación estándar de 0.265 y las estadísticas de razón de verosimilitud y Wald son 2.870 y 3.209 , cada una respectivamente con un p valor de 0.0902 y 0.0732 ; indicando que la correlación espacial incluida en la media es significativa al menos a un 90 %.

Modelo espacial SAR mixto

La ecuación que describe este modelo es la misma usada en la especificación del modelo SAR mixto cuando se trabaja la matriz de contigüidad física de primer orden. Asimismo se muestra en la tabla 1.11 las estimaciones de cada uno de los parámetros para el predictor lineal sin las covariables rezagadas.

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	335.761	2.946	-0.519	0.494	0.483	-81.187	-1.736
<i>d.e</i>	93.374	1.568	2.311	0.488	0.758	37.143	0.540

TABLA 1.11. Estimación y desviación estándar de los parámetros del modelo SAR mixto - Variables explicativas - Matriz de contigüidad física de segundo orden

La tabla 1.12 contiene las desviaciones y estimaciones de las variables rezagadas.

Cada una de las tablas mencionadas anteriormente muestran que las variables $Displc$, $1/Rec$ y $RNtr$ son significativas. Sin embargo y en contraste con el mismo modelo pero

Parámetro	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
<i>Estim</i>	14.430	-0.870	0.999	-0.409	350.877	-0.813
<i>d.e</i>	9.360	11.285	1.965	4.131	276.563	2.570

TABLA 1.12. Estimación y desviación estándar de los parámetros del modelo SAR mixto - Variables explicativas rezagadas - Matriz de contigüidad física de segundo orden

con matriz de contigüidad física de primer orden, ninguna de las variables rezagadas con esta estructura de vecindad resulta significativa. Esto sugiere que la definición de una estructura de vecindad así en este caso puede no ser oficiosa. El valor de la estimación del parámetro de interés ρ es -0.968 con una desviación de 0.531 y los valores de las estadísticas de razón de verosimilitud y Wald son 2.136 y 3.323 respectivamente, con p valores de 0.1438 y 0.0682 , mostrando que en uno de los dos casos es posible no rechazar la hipótesis que el parámetro es cero. La estimación de la varianza residual es 746.02 , subiendo en comparación al mismo modelo con diferente estructura de vecindad. El AIC que muestra este modelo es 343.84 .

Modelo espacial CAR

Al igual que los casos anteriores no se especifica la estructura del modelo ya que coincide con las mostradas para el mismo con estructura de vecindad diferente. Así, el cuadro 1.13 muestra las estimaciones y desviaciones para cada uno de los parámetros considerados en este.

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	235.815	0.455	-1.911	0.422	0.265	-31.851	-1.038
<i>d.e</i>	49.264	1.314	1.152	0.515	0.558	17.863	0.427

TABLA 1.13. Estimación y desviación estándar de los parámetros del modelo (1.35) - Matriz de contigüidad física de segundo orden

La estimación de λ es -0.101 con una desviación de 0.109 . El valor de la estadística de razón de verosimilitud para probar la hipótesis si $\lambda = 0$ es de 0.8803 y su respectivo p valor es 0.3481 . Esto indica que el modelo no presenta correlación espacial y la hipótesis nula podría ser rechazada. Para este modelo las variables que pueden ser consideradas significativas son IndIPS, $1/Rec$ y RNtr, al menos a un 90% de confiabilidad. La estimación de la varianza del residual es de 964.98 y el AIC es de 339.94 . Nótese que con respecto al modelo CAR con la primera estructura de vecindad considerada, todos estos valores cambian un poco mostrando un ajuste menos adecuado. Se muestra un cuadro que resume cada uno de los valores de interés de cada modelo con la nueva estructura de vecindad considerada. La tabla 1.14 resume los valores de AIC, $\hat{\sigma}^2$, $\hat{\rho}$ y $\hat{\lambda}$ según corresponda.

En cuanto al cuadro que relaciona el criterio AIC, los estimadores de la varianza residual y los parámetros de autocorrelación según corresponda, se nota que aunque la diferencia de los valores del criterio AIC son despreciables entre algunos modelos; el modelo que menor criterio muestra es el SAR variable respuesta. Los respectivos valores del R cuadrado de Nagelkerke, son mostrados en el cuadro 1.15

Según el cuadro 1.15, el modelo con mayor R^2 es SAR mixto. Debe tenerse en cuenta el número de parámetros que se usa en el ajuste del modelo. Sin embargo, sacando del

Modelo	AIC	σ^2	ρ	λ
Normal Lineal	338.81	1316.7		
SAR Residual	339.98	982.95		-0.665
SAR variable respuesta	337.95	936.36	-0.474	
SAR mixto	343.84	746.02	-0.968	
CAR	339.94	964.98		-0.101

TABLA 1.14. AIC , σ^2 , ρ y λ para los modelos considerados

Modelo	$-2 \log L$	R^2 de Nagelkerke
Normal Lineal	322.816	0.534
SAR Residual	321.984	0.546
SAR variable respuesta	319.946	0.573
SAR mixto	313.838	0.645
CAR error	321.936	0.547

TABLA 1.15. pseudo R^2 de Nagelkerke para los modelos considerados - Matriz de contigüidad de segundo orden

modelo SAR mixto las variables que no son significativas, solo dejando Displc, 1/Rec, RNtr y sus respectivas cantidades rezagadas, el valor de $-2 \log L(\beta)$ es de 316.32 y el R cuadrado se transforma en 0.6176, considerando así 9 parámetros que estimar en vez de 15.

1.4.3.2. Matriz de pesos espaciales por regiones naturales

En esta parte del escrito se construye la vecindad de acuerdo a las regiones naturales. La figura 1.7 muestra el mapa de las regiones naturales de Colombia por departamento.



FIGURA 1.7. Regiones naturales por departamento en Colombia

Para esta vecindad los valores de las estadísticas de correlación espacial son respectivamente 0.1080, 0.8650 y 0.0314 para la I de Moran, C de Geary y estadística G . Para cada uno los p valores de la prueba de hipótesis de que son estadísticamente igual a cero son 0.0935, 0.0762 y 0.0762 respectivamente, en donde tampoco se puede rechazar

la hipótesis de que hay correlación espacial. Al igual que en los dos casos anteriores en que consideramos dos estructuras de vecindad diferentes, mostramos los resultados más relevantes cuando consideramos esta nueva vecindad.

Modelo espacial SAR con rezago en el residual

La ecuación corresponde a la mostrada en la expresión (1.32) para el predictor lineal y para los términos de error (1.33). La tabla 1.16 muestra los resultados de las estimaciones.

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	148.877	1.270	-2.408	1.126	-0.371	-22.725	-0.117
<i>d.e</i>	39.925	0.556	0.921	0.487	0.305	15.941	0.365

TABLA 1.16. Estimación y desviación estándar de los parámetros del modelo con estructura de error (1.33) - Matriz de vecindades por regiones naturales

La tabla 1.16 muestra que las únicas variables que resultan significativas son Displc, IndIPS y Vac, al menos a un 90 % de confiabilidad. En este caso el valor de $\hat{\lambda}$ es de -1.777 con una desviación estándar de 0.355 y los p valores de las pruebas de razón de verosimilitud y Wald son 0.081 y 5.441×10^{-07} respectivamente, en donde cada una de las estadísticas toman un valor de 3.035 y 25.101. En ambos casos podemos rechazar la hipótesis nula de ausencia de correlación espacial mediante esta nueva definición de vecindad. El valor del AIC para este es de 337.78 y la estimación del término σ^2 es 638.76.

Modelo espacial SAR con rezago en la variable respuesta

La configuración de la media del modelo se muestra en la ecuación (1.34). Las estimaciones para esta propuesta son mostradas en la tabla 1.17.

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	237.061	0.528	-1.705	0.495	0.350	-33.512	-1.218
<i>d.e</i>	74.377	1.397	1.164	0.506	0.579	18.905	0.488

TABLA 1.17. Estimación y desviación estándar de los parámetros del modelo con estructura (1.34) - Matriz de vecindades por regiones naturales

En la tabla 1.17 se puede apreciar que las variables significativas son 1/Rec y RNtr a un 90 % de confiabilidad. La estimación del parámetro ρ que define la autocovariable es -0.0237 con desviación estándar de 0.210 y así, los valores de cada uno de los respectivos estadísticos razón de verosimilitud y Wald son 0.012 y 0.013, con p valores de 0.912 y 0.910, indicando que no se rechaza la hipótesis que $\rho = 0$; luego, se concluye que estadísticamente no hay correlación de acuerdo a la estructura de vecindad expuesta. El valor del AIC es de 340.80 y el valor estimado de σ^2 es 1036.9.

Modelo espacial SAR mixto

Anteriormente se mostraron las estimaciones para los modelos espaciales SAR con estructura de correlación en el residual y en la variable respuesta. Recordemos que la especificación del modelo es la misma que cuando consideramos este mismo con una matriz

de contigüidad física de primer orden. Al igual que los casos anteriores, primero se consig-
nan las estimaciones de los parámetros del modelo sin rezago; estos resultados pueden ser
visualizados en la tabla 1.18.

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	432.582	3.067	-0.522	0.336	-0.501	-77.616	-0.650
<i>d.e</i>	334.416	1.635	3.217	0.771	0.713	53.704	0.526

TABLA 1.18. Estimación y desviación estándar de los parámetros del modelo SAR mixto - Varia-
bles explicativas - Matriz de vecindades por regiones naturales

Se puede notar que según la tabla 1.18 no hay ninguna variable significativa con esta
estructura de vecindad. El valor estimado de ρ es -1.395 con una desviación estándar de
 0.379 . Ahora se observan las estimaciones de los parámetros que definen las covariables
rezagadas; la tabla 1.19 muestra esta información

Parámetro	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
<i>Estim</i>	3.780	6.740	1.071	-3.712	-295.529	3.284
<i>d.e</i>	4.718	19.407	3.520	3.165	367.341	2.280

TABLA 1.19. Estimación y desviación estándar de los parámetros del modelo SAR mixto - Varia-
bles explicativas rezagadas - Matriz de vecindades por regiones naturales

Como se aprecia, ninguna de las variables rezagadas de acuerdo a esta estructura de
vecindad resulta siendo significativa. El AIC es 336.27 y el valor de la estimación para σ^2
es 490.66. A pesar de esto, cabe mencionar que con esta estructura se rechaza la hipótesis
de que los residuales no están espacialmente correlacionados. Como se puede entonces
apreciar, este modelo que en casos anteriores presentaba una mejoría con respecto a los
demás aun quitándole parámetros en la estimación, no muestra un adecuado ajuste.

Modelo espacial CAR

La estructura del modelo se muestra en las ecuaciones (1.35) y (1.36). Se recuerda
que en este caso la matriz \mathbf{W} debe ser simétrica. De esta manera, se escoge la misma
matriz que relaciona la vecindad, solo que no es totalizada por filas. Las estimaciones son
presentadas en el cuadro 1.20.

Parámetro	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Estim</i>	136.283	1.481	-2.645	1.206	-0.439	-22.142	0.048
<i>d.e</i>	37.708	0.195	0.927	0.486	0.182	16.809	0.326

TABLA 1.20. Estimación y desviación estándar de los parámetros del modelo (1.35) y (1.36) -
Matriz de vecindades por regiones naturales

Se puede entonces concluir con respecto a esta nueva tabla que las únicas variables
que no son significativas, son 1/Rec y RNtr. En contraste, las otras lo son a un 95%
de confiabilidad. El resultado de este modelo es bastante bueno en comparación a los
modelos CAR presentados con anterioridad. El estimado de λ es cercano a -6.66 con
una desviación estándar muy cercana a cero, y la estadística de razón de verosimilitud
que prueba la hipótesis que $\lambda = 0$ toma un valor de 86.534, lo que conduce a un p valor
menor que 2.22×10^{-16} . En este mismo caso, el valor estimado de la varianza residual

σ^2 es 2.664×10^{-5} y el AIC disminuye considerablemente hasta 254.28. Sin embargo, al revisar los supuestos básicos del modelo, encontramos que los residuales no se comportan adecuadamente. Por esto, se trata de ajustar el modelo, y se encuentra uno que lo hace mejor. Al quitar del predictor lineal las variables $1/Rec$ y $RNtr$, se obtiene la forma funcional

$$y = \mathbf{X}\beta + \nu \tag{1.39}$$

y

$$\nu = \lambda \mathbf{W}\nu + \epsilon \tag{1.40}$$

con

$$\mathbf{X}\beta = \beta_0 + \beta_1 \text{Displc} + \beta_2 \text{IndIPS} + \beta_3 \text{Vac} + \beta_4 \text{NBI} \tag{1.41}$$

La tabla 1.21 muestra las estimaciones de este modelo

Parámetro	β_0	β_1	β_2	β_3	β_4
<i>Estim</i>	145.900	1.205	-3.821	1.093	-0.250
<i>d.e</i>	0.011	1.622×10^{-4}	2.423×10^{-4}	1.279×10^{-4}	7.057×10^{-5}

TABLA 1.21. Estimación y desviación estándar de los parámetros del modelo (1.40) - Matriz de vecindades por regiones naturales

Para el modelo presentado en el cuadro 1.21 se tiene que el AIC alcanza un valor de 251.41, un poco más bajo que el anterior. Además, la estimación de la varianza es de 2.7568×10^{-5} . El valor estimado de λ sigue siendo muy cercano a -6.66 con una desviación estándar de casi cero, y la estadística de razón de verosimilitud que prueba la hipótesis $\lambda = 0$ toma un valor de 92.963, tomando un p valor muy cercano a cero. Además de esto, se puede apreciar de la tabla en mención que todas las variables son significativamente distintas de cero. Al incluir un efecto de vecindad como el trabajado en esta pequeña sección, se concluye que este es el mejor modelo CAR en todas las tres estructuras espaciales consideradas. Como se hizo para cada uno de los casos anteriores, se presenta una tabla que resume el comportamiento en general de cada uno de los modelos considerados con esa estructura de vecindad, las regiones naturales.

Modelo	AIC	σ^2	ρ	λ
Normal Lineal	338.81	1316.7		
SAR Residual	337.38	638.76		-1.777
SAR variable respuesta	340.8	1036.9	-0.023	
SAR mixto	336.27	490.66	-1.395	
CAR	251.41	2.756×10^{-5}		-6.666

TABLA 1.22. AIC , σ^2 , ρ y λ para los modelos considerados

La tabla 1.22 muestra los valores más relevantes de cada modelo considerado. Claramente, se puede establecer que para esta nueva propuesta, el modelo más adecuado que describe mejor la información, es el modelo CAR, pues el AIC es mucho más bajo que todos los demás. Si consideramos el R cuadrado de Nagelkerke mostrado para cada uno en el cuadro 1.23, podemos apreciar un importantísimo comportamiento del modelo CAR.

Como se ha mencionado anteriormente, no es viable compararlo directamente, si no se considera el número de parámetros estimados como es el caso del modelo SAR mixto

Modelo	$-2 \log L$	R^2 de Nagelkerke
Normal Lineal	322.816	0.534
SAR Residual	319.781	0.575
SAR variable respuesta	322.804	0.535
SAR mixto	306.270	0.718
CAR error	237.415	0.965

TABLA 1.23. pseudo R^2 de Nagelkerke para los modelos considerados - Matriz de contigüidad por regiones naturales

en donde se estiman 15 parámetros en comparación de los 9 de los demás salvo los 7 del modelo CAR. Recordemos que el R^2 se puede inflar arbitrariamente en dado caso cuando el número de parámetros aumenta.

1.5. Conclusiones

Como se puede apreciar a lo largo del capítulo, existe una evidente correlación espacial en las variables medidas en los departamentos de Colombia. Nótese que el modelo SAR mixto se comporta mejor que los otros modelos SAR considerados, comparándolo con estas estructuras usando la misma definición de matriz de contigüidad. En efecto, el modelo con mejor comportamiento es el modelo CAR residual, en donde se considera una estructura de autocorrelación en los residuales. Observamos también que el índice de desplazamiento está positivamente correlacionado con el índice de volumen de atención hospitalaria, al igual que la variable VAC. Esto sugiere que entre más proporción de hogares con personas desplazadas, mayor es el volumen de atención hospitalaria como era de esperarse. Asimismo, entre mayor es el porcentaje de niños menores de un año que han recibido la tercera dosis de la vacuna contra el polio, mayor es el volumen de atención hospitalaria, como era de esperarse. También, se puede concluir que en los departamentos en donde hay más IPS públicas en comparación con las IPS privadas, se presenta menos volumen de atención hospitalaria, al igual que en los departamentos en donde hay mayores índices de NBI. Los resultados mencionados, son los resultados esperados en el país. Además de todos lo mencionado, se observa que el modelo mejora sustancialmente, cuando en el término del error, le incluimos una variable que resuma la estructura espacial intrínseca en los datos, agrupando vecindades mediante el criterio de las regiones naturales. El modelo CAR residual no se comporta de manera similar al SAR residual, ya que en el caso del CAR, los valores que toma la matriz \mathbf{W} , son mucho más grandes que los SAR.

Un enfoque Bayesiano en la modelación espacial de la mortalidad infantil colombiana y período postnatal de cribado

En Colombia durante muchos años se ha sufrido el conflicto armado que ha dejado a los menos favorecidos en circunstancias de pobreza, sobre todo en las regiones más apartadas en donde los grupos al margen de la ley cometen actos delictivos, atacando a su vez a la población infantil. Junto a esto, la corrupción latente en este país ha generado durante todo este tiempo bajo nivel educativo que afecta el desarrollo con el paso del tiempo. Uno de los principales retos que tiene el actual gobierno es reducir la tasa de muertes infantiles por causas como la pobreza y el conflicto armado, entre otras.

De la mano del gobierno, la UNICEF ha actuado y ha hecho pública su preocupación en los últimos años sobre los acontecimientos de violencia en contra de la niñez en Colombia, a pesar de que las cifras de muerte y violencia se han reducido. Sin embargo, las políticas colombianas deberían estar enfocadas en la protección infantil, y en general, en la protección de los derechos humanos. Se conoce que para el año 2005 la tasa de mortalidad infantil para niños menores de 5 años fue de 19 por cada mil infantes que nacieron vivos, indicando una relevante diferencia comparándolo con el mismo indicativo de los países desarrollados, que fue de 6 por cada mil. Además de la violencia que azota el territorio colombiano hay que señalar que las políticas de salud pública no son las mejores, y de esta manera, las causas de muerte en esta población, en específico, se multiplican.

Para determinar y caracterizar posibles factores que inciden en la muerte de niños menores de cinco años en Colombia se hace uso de la información de la Encuesta Nacional Demográfica y de Salud para el año 2005, la cual contiene información por departamentos sobre el nivel familiar al cual pertenece el infante. Ya que se dispone de datos ordenados en el espacio, es importante determinar si la tasa de muertes infantiles en Colombia también depende de la zona o territorio, por el factor ya mencionado que incide drásticamente en el comportamiento de la sociedad colombiana: la violencia.

Es importante señalar que en Colombia, existen algunas regiones más sacudidas por la violencia que otras, y por consiguiente, son regiones con mayor índice de pobreza. Algunos de los departamentos con más índices de violencia, son el Chocó, Putumayo, Cauca, Amazonas, Nariño y Cauca, donde algunos de ellos son los límites con Perú y Ecuador. También las selvas colombianas y el llano, departamentos como Arauca, que limitan con Venezuela.

En este estudio se analiza la mortalidad infantil y sus posibles causas. El número de niños menores de cinco años que murieron durante el período de 2000 al 2005 (variable NF) fue estimado para cada uno de los 32 departamentos, y también en las cuatro ciudades de Bogotá, Cartagena, Barranquilla y Santa Marta. Además se consideró el número de madres (en miles) que había en cada una de esas 36 áreas (variable Nmadres), así como el porcentaje de mujeres mayores de 18 años que han sufrido algún tipo de abuso físico de sus actuales familias (variable Viol). Además, para cada una de las áreas de estudio, hay también información sobre el porcentaje de la población que tiene necesidades básicas insatisfechas para el año 2004 (variable NBI), sobre los recursos otorgados para la educación y atención integral para niños pequeños (en miles) por familias, brindados por el gobierno central a las diferentes administraciones departamentales o municipales en el año 2005 (variable REC), sobre el porcentaje de personas jóvenes (entre los 18 y los 23 años) quienes han accedido a educación superior en 2005 (variable ES), sobre el porcentaje de niños menores de un año que han recibido la tercera dosis de la vacuna contra el polio aplicada en el año 2004 (variable Vac), sobre el número de madres en la i -ésima área que pasaron por un período postnatal de cribado (NScree), sobre el número de madres (en miles) que han tenido su último hijo después de 1999 (NMadres99) y sobre el porcentaje de madres que han pagado los costos totales del período postnatal de cribado (variable Pay).

En las siguientes secciones se revisan brevemente los conceptos de sobredispersión y del enfoque o estimación bayesiana de los parámetros.

2.1. Modelos lineales generalizados

Sea Y_i una variable aleatoria que pertenece a la familia exponencial. La función de densidad de probabilidad de Y_i puede escribirse de la forma

$$f(y_i; \theta_i, \phi) = \exp[\phi\{y_i\theta_i - b(\theta_i)\} + c(y, \phi)] \quad (2.1)$$

donde $E(Y_i) = \mu_i = b'(\theta_i)$, $\text{Var}(Y_i) = \phi^{-1}V_i$, $V = \partial\mu/\partial\theta$ es la función de varianza y $\phi^{-1} > 0$ es el parámetro de dispersión [55]. Como ejemplos de distribuciones que pertenecen a la familia exponencial, se debe mencionar a la Normal, a la Gamma, a la Poisson y a la binomial entre muchas otras. En los dos primeros casos el coeficiente ϕ respectivamente toma los valores de σ^{-2} y CV^{-2} , con CV el coeficiente de variación. Esto permite cierta flexibilidad en el manejo de la varianza de forma que la variabilidad que presentan los datos puede resumirse sin ningún problema de acuerdo al valor que dispone el parámetro. Sin embargo, en los otros dos casos, $\phi = 1/m$ para el caso de la binomial y $\phi = 1$ en la situación Poisson, indicando que la estructura de varianza asociada a estas distribuciones, está sujeta únicamente a los cambios que puedan ofrecer las funciones de varianzas respectivas o el tamaño de muestra como el caso de la binomial. En el caso de la distribución Poisson la función de varianza es $V(\mu) = \mu$ y en el caso de la binomial, $V(\mu) = \mu(1 - \mu)$.

Desde los modelos lineales generalizados, se permite establecer una relación entre la media de las observaciones, μ_i , $i = 1, \dots, n$, con las variables explicativas dispuestas previamente en la recolección de la información, de la forma

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}, \quad \text{para } i = 1, \dots, n \quad (2.2)$$

donde η_i es llamado el predictor lineal para la i -ésima observación, \mathbf{x}_i es el vector de tamaño $(p+1) \times 1$ de variables explicativas para la i -ésima observación, β es un vector de tamaño $(p+1) \times 1$ de parámetros desconocidos, y $g(\cdot)$ es una función de enlace, monótona y diferenciable. Usualmente, la función de enlace canónica $g(\mu_i) = \theta_i$ con θ_i como en (2.1) se elige, garantizándose la existencia o convergencia de la estimación por máxima verosimilitud para el vector β y a su vez una relativamente fácil interpretación de los coeficientes del modelo.

Para el caso de la distribución binomial, la variable que pertenece a la familia exponencial es Y_i^* , la proporción de éxitos obtenidos en muestras de tamaño m_i , $i = 1, \dots, n$. En tal caso, el parámetro de dispersión es $\phi_i = 1/m_i$ como habíamos mencionado anteriormente. Además, si π_i representa la probabilidad de éxitos en la i -ésima observación binomial, el valor esperado de la variable respuesta es $E(Y_i^*) = \pi_i$ y la función de varianza es dada por $V(\pi_i) = \pi_i(1 - \pi_i)$, así que $\text{Var}(Y_i^*) = \pi_i(1 - \pi_i)/m_i$. Es importante resaltar que en este caso se está modelando la proporción de éxitos. Sin embargo, las propiedades asociadas a la variable binomial que cuenta el número de éxitos, puede ser directamente obtenida después de definir la variable $Y_i = m_i Y_i^*$, de tal forma que su media y varianza están dadas por $m_i \pi_i$ y $m_i \pi_i(1 - \pi_i)$, respectivamente [55].

Para el caso de la distribución Poisson, donde Y_i es la variable aleatoria que cuenta el número de éxitos que ocurre en un tiempo dado, distancia o área específica, el parámetro de dispersión es $\phi_i = 1$. Si λ_i es la tasa media de ocurrencia de un evento dado, y la función de varianza está dada por $V(\lambda_i) = \lambda_i$, entonces la media y la varianza de la variable aleatoria Y_i son ambas dadas por λ_i . La función de enlace canónica, para el caso binomial es la logística, dada por la expresión $g(\pi_i) = \log[\pi_i/(1 - \pi_i)] = \eta_i$, y para el caso de la variable Poisson, es la logarítmica, dada por la expresión $g(\lambda_i) = \log(\lambda_i) = \eta_i$, en donde $\log(\cdot)$ hace referencia al logaritmo natural.

Así, una vez establecidas las principales características de un modelo lineal generalizado, recordamos que el fenómeno de la sobredispersión ocurre cuando la varianza de los datos excede a la que puede capturar la distribución. Para el caso de la binomial, hay presencia de sobredispersión si $\text{Var}(Y_i) > m_i \pi_i(1 - \pi_i)$, y para el caso de la Poisson, si $\text{Var}(Y_i) > \lambda_i$. Cuando no se tiene en cuenta este fenómeno en la construcción del modelo, se puede cometer subestimación de los errores estándar de las estimaciones, lo que implica una inferencia errónea para los estimadores de los parámetros de regresión [39].

2.2. Modelos usuales de sobredispersión

En esta sección se aborda de manera general los modelos más usados para tratar el fenómeno de sobredispersión en los casos cuando la variable respuesta tiene distribución Poisson y binomial. Hay muchos modelos que capturan la sobredispersión, los cuales surgen desde posibles mecanismos alternativos para el proceso subyacente [39]. Los modelos considerados en este escrito, son aquellos producto de asumir alguna distribución al parámetro que define la distribución (el caso de los modelos jerárquicos) y el de incluir un efecto aleatorio en el predictor lineal. Estos dos métodos permiten extender la varianza de la variable respuesta, como se puede ver a continuación.

2.2.1. Modelos de conteo con sobredispersión

Uno de los problemas más comunes en los modelos estadísticos es la presencia de *sobredispersión*. Por este término entendemos que la varianza de la variable respuesta Y excede la varianza nominal [51], implícita en la relación media - varianza, aplicando los modelos lineales generalizados[39]. A lo largo de esta sección se asume que la variable respuesta de interés Y sigue una distribución discreta Poisson de media μ , o, binomial de parámetros m y π , con m el número de ensayos y π la probabilidad de éxito. De esta manera, si se establece que $\phi = 1$ en ambos casos, se puede presentar que $\text{Var}(Y) = \sigma^2\mu$ para el caso de la Poisson, y $\text{Var}(Y) = \sigma^2m\pi(1 - \pi)$ para el caso de la binomial, con $\sigma^2 > 1$ [55].

La sobredispersión puede surgir de muchas formas, y diferentes causas deben ser evaluadas o tenidas en cuenta para reducir el impacto que produce formular un modelo incapaz de capturar toda la variabilidad con la que cuenta la información de la que se dispone. Algunas de ellas representan una aparente sobredispersión, como el caso de presencia de algunos puntos extraños o mediciones atípicas que pueden ser simplemente eliminados y así reducir las evidencias de la presencia del fenómeno en cuestión. Otras causas aparentes son la ausencia de algún término en la parte sistemática del modelo y correlación entre las respuestas de un mismo grupo. Si no se tiene en cuenta el exceso de variabilidad, se pueden cometer serios errores de subestimación de errores estándar y por supuesto inferencias erróneas sobre los parámetros de regresión[39]. Con base en todo esto, se definen en las secciones siguientes cada uno de los principales modelos usados para capturar el exceso de variabilidad de la variable respuesta.

2.2.1.1. Modelos asociados a la distribución Poisson

Sea Y_1, Y_2, \dots, Y_n una muestra aleatoria con distribución Poisson de media μ . Se supone que la dispersión de los datos es más grande que la capturada por el modelo Poisson, es decir, $\text{Var}(Y) > \text{E}(Y)$ [51], o como se mencionó anteriormente, si $\text{Var}(Y) = \sigma^2\mu$ con $\sigma^2 > 1$ [55]. Una causa probable de este fenómeno es la heterogeneidad de las unidades muestrales que puede ser debido a la variabilidad inter unidades experimentales[55]. Una forma de contrarrestar este fenómeno y encontrar la manera de capturar la variabilidad extra en este modelo, es asumiendo que la media de la Poisson es una cantidad estocástica, teniendo esta su propia distribución; por conveniencia, Gamma [69].

De esta manera, se supone que $Y|Z = z \sim \text{Pois}(z)$, en donde $Z \sim G(\mu, \phi)$ en que ϕ no depende de μ . Con esta parametrización y notación (Paula, G., 2010), se tiene entonces $\text{E}(Z) = \mu$ y $\text{Var}(Z) = \mu^2/\phi$, donde la distribución de Y incondicionada tiene los dos primeros cumulantes

$$\text{E}(Y) = \mu \quad \text{y} \quad \text{Var}(Y) = \mu + \mu^2/\phi \quad (2.3)$$

con distribución binomial negativa [50]. Nótese que la varianza es más grande que la media, obteniendo además que la nueva función de varianza para este caso es la cuadrática en vez de la lineal [51]. Margolin et al. [50] y G. Paula [55] muestran que la función de densidad de probabilidad resultante es

$$\text{Pr}(Y = y) = \binom{y + \phi - 1}{y} (1 - \pi)^\phi \pi^y \quad (2.4)$$

con $\pi = \mu/(\mu + \phi)$. Se supone ahora que Y_1, Y_2, \dots, Y_n son variables aleatorias independientes de manera que $Y_i \sim BN(\mu_i, \phi)$ en que $E(Y_i) = \mu_i$ y $\text{Var}(Y_i) = \mu_i + \mu_i^2/\phi$, se modela la parte sistemática

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

con $g(\cdot)$ una función de enlace que cumpla todos los requisitos de los modelos lineales generalizados, \mathbf{x}_i es el vector columna de todas las covariables medidas en el i -ésimo individuo y $\boldsymbol{\beta}$ es el vector de parámetros a ser estimado. Por lo general, $g(\mu_i) = \log(\mu_i)$ [55]. Estos ajustes dan la posibilidad de considerar distintos modelos de este tipo asumiendo que los parámetros en la distribución asumida para Z varían.

El siguiente modelo considerado para tratar la variación excesiva, es el Poisson normal (conocido también como modelo Poisson log-lineal mixto para el caso en que su función de enlace es la logarítmica). Aquí se considera la inclusión de un efecto aleatorio en el predictor lineal [39]. De manera general, para una función de enlace $g(\cdot)$, se supone ahora que dado un vector de efectos aleatorios $\boldsymbol{\alpha}$, los conteos Y_1, Y_2, \dots, Y_n , son condicionalmente independientes tales que $Y_i | \boldsymbol{\alpha} \sim \text{Pois}(\mu_i)$ donde

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\alpha} \quad (2.5)$$

con \mathbf{x}_i y \mathbf{z}_i vectores fila de las matrices \mathbf{X} y \mathbf{Z} respectivamente, ambas conocidas. Este es un caso especial de los modelos lineales mixtos generalizados, en el cuál la familia exponencial (condicionada) es Poisson. El parámetro de dispersión en este caso es nuevamente $\phi = 1$ [41]. En particular para este caso, la función de enlace es $g(\mu) = \log(\mu)$ (función de enlace canónica), y para tratar sobredispersión se hace $\mathbf{Z} = \mathbf{I}_n$, con \mathbf{I}_n una matriz identidad de tamaño $n \times n$ y así, $\boldsymbol{\alpha}$ es un vector de tamaño $n \times 1$, lo que indica un efecto aleatorio para cada uno de los individuos. La ecuación (2.3) se convierte para este caso en específico

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i \quad (2.6)$$

con α_i la entrada i -ésima del vector $\boldsymbol{\alpha}$. La función de varianza para el modelo (2.4) es de la forma $\text{Var}(Y_i) = \mu_i + k' \mu_i^2$, que coincide con la de la binomial negativa. En efecto, con esta función de enlace canónica (logarítmica) y un efecto aleatorio aditivo en el predictor lineal denotado por α_i para cada individuo, siempre se obtiene una función de varianza similar a esta forma, para cualquier efecto aleatorio [39]. La suposición distribucional a menudo que se usa para el vector aleatorio $\boldsymbol{\alpha}$ es normal multivariado con media $\mathbf{0}$ y matriz de covarianza \mathbf{G} que puede depender de un vector $\boldsymbol{\zeta}$ de componentes de varianza. Con la estructura general de los modelos lineales mixtos generalizados GLMM (por sus siglas en inglés) se obtiene una opción para modelar cuando las respuestas son correlacionadas. Por tal razón, GLMM es a menudo usado para tratar variables categóricas o discretas cuando las respuestas son correlacionadas [41]. De esta manera, se obtiene una forma de tratar la sobredispersión, cuando la causa de esta es el no cumplimiento de independencia entre las observaciones de la respuesta Y .

2.2.1.2. Modelos asociados a la distribución Binomial

Uno de los modelos considerados por Hinde para datos con exceso de variación y de respuesta binomial, es el beta binomial, asumiendo sobredispersión no constante. Williams en 1975 [69], propone este modelo para un experimento toxicológico, diseñado para investigar los efectos teratógenos o fetotóxicos de cierto químico en un laboratorio de animales.

Se consideran los dos niveles de respuesta de una variable como éxito o fracaso. Se supone que la población bajo estudio, se agrupa bajo ciertas características formando lo que se conoce como Clúster. El ensayo presentado en el escrito de D.A. Williams [69], es un ejemplo de esto. Ahora se supone que el número de clústeres es fijo, denotado por k . En el i -ésimo clúster, el número de unidades medidas con respuesta positiva o éxito, Y_i , es asumida con distribución binomial de parámetros m_i, π_i [51]. De esta manera, $E(Y_i) = m_i\pi_i$ y $\text{Var}(Y_i) = m_i\pi_i(1 - \pi_i)$: Nótese que la variabilidad cambia de clúster a clúster. Para permitir variación binomial extra, se introducen variables continuas inobservables Π_i distribuidas independientemente sobre $(0, 1)$ con $E(\Pi_i) = \pi_i$, $\text{Var}(\Pi_i) = \delta\pi_i(1 - \pi_i)$, y se asume que, condicionado sobre $\Pi_i = \pi_i$, la variable Y_i tiene distribución binomial de parámetros (m_i, π_i) [70]. De esta forma, $Y_i|\pi_i \sim \text{Bin}(m_i, \pi_i)$, entonces $E(Y_i|\pi_i) = m_i\pi_i$ y $\text{Var}(Y_i|\pi_i) = m_i\pi_i(1 - \pi_i)$. De ahí se obtiene

$$E(Y_i) = E\{E(Y_i|\pi_i)\} = m_i\pi_i \quad (2.7)$$

y

$$\text{Var}(Y_i) = E\{\text{Var}(Y_i|\pi_i)\} + \text{Var}\{E(Y_i|\pi_i)\} \quad (2.8)$$

$$= \sigma_i^2 m_i \pi_i (1 - \pi_i) \quad (2.9)$$

en donde $\sigma_i^2 = 1 + (m_i - 1)\delta$ [55]. El caso particular que interesa en este punto es $\delta > 0$, ya que se necesita que $\sigma_i^2 > 1$. El modelo beta - binomial asume que Π_i sigue una distribución beta con función de densidad

$$\text{Pr}(\Pi_i = \pi_i) = \pi_i^{a-1}(1 - \pi_i)^{b-1}/B(a, b) \quad (2.10)$$

con $0 < \pi_i < 1$, $a > 0$, $b > 0$ [69]. Luego, la distribución de Y_i incondicionada tiene función de probabilidad

$$\text{Pr}(Y_i = y_i) = \frac{\binom{m_i}{y_i} B(a + y_i, m_i + b - y_i)}{B(a, b)} \quad (2.11)$$

para $y_i = 0, 1, 2, \dots, m_i$. Esta es la distribución beta binomial y su forma límite es una binomial negativa, análogo al límite Poisson de la binomial [32]. Es preferible la parametrización

$$\pi_i = \frac{a_i}{a_i + b} \quad \text{y} \quad \tau_i = \frac{1}{a_i + b}$$

luego, la varianza de Y_i es $m_i\pi_i(1 - \pi_i)(1 + m_i\tau_i)/(1 + \tau_i)$ y así, por definición $\tau_i > 0$ y $\pi_i > 0$. Esta reparametrización es conveniente especialmente en casos donde se considera apropiado el modelo binomial puro ($\tau_i = 0$), y por lo tanto a y b son ambos infinitos, lo cual podría causar dificultades en estimación e hipótesis de interés [32]. En más detalle, si la variable aleatoria Π_i tiene una distribución degenerada con probabilidad 1 a un único punto (o de otra manera, escrito anteriormente como $a_i \implies \infty$ y $b_i \implies \infty$), entonces $\text{Var}(\Pi_i) = 0$ y Y_i resulta siendo una binomial, ya que $\delta = \text{Var}(\Pi_i)$ y así, $\delta = 1/(a_i + b + 1)$ [1]. Sea \mathbf{X} una matriz diseño conocida, como se menciona anteriormente. Al igual que en el caso Poisson, suponiendo ahora que Y_1, Y_2, \dots, Y_n es una muestra aleatoria de manera que $Y_i \sim \text{BBD}(\pi_i, \tau_i)$ (Beta Binomial Distribution), en que $E(Y_i) = \pi_i$, se modela la media de Y_i por medio de la parte sistemática

$$g(\pi_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

con $g(\cdot)$ una función de enlace y todo lo demás como en el caso Poisson - Gamma de las secciones anteriores. Por lo general, $g(\pi) = \log(\pi/(1 - \pi))$, que es la función de enlace logística, canónica para una distribución binomial perteneciente a la familia exponencial. Si bien, el modelo beta binomial asume que las variables Π_i tienen distribución beta, otra posibilidad es asumir que el predictor lineal tiene alguna distribución continua [39]. Al igual que en la sección (2.2.1) para el caso de la Poisson, supongamos que dado un vector de efectos aleatorios $\boldsymbol{\alpha}$, las variables de conteo Y_1, Y_2, \dots, Y_n son condicionalmente independientes tales que $Y_i | \boldsymbol{\alpha} \sim Bn(m_i, \pi_i)$, en donde $\mu_i = m_i \pi_i$ y así la parte sistemática tiene la forma (2.3) en la sección (2.2.1). Si la función de enlace escogida es la logística (función de enlace canónica), uno obtiene

$$\log\left(\frac{\mu_i}{m_i - \mu_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\alpha} \quad (2.12)$$

con \mathbf{x}_i y \mathbf{z}_i definidas con en la sección (2.2.1). Este es un caso especial de los GLMM, en el cual la familia exponencial condicionada es Binomial. Al igual que en el caso simple, el parámetro de dispersión ϕ es uno [41]. Para tratar la sobredispersión, se escoge $\mathbf{Z} = \mathbf{I}_n$, lo que indica un efecto aleatorio para cada uno de los individuos o clúster. La ecuación (2.9) se convierte en

$$\log\left(\frac{\mu_i}{m_i - \mu_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i \quad (2.13)$$

con α_i la entrada i -ésima entrada del vector $\boldsymbol{\alpha}$. Una ventaja de este modelo sobre el anterior, es que este puede ser extendido más fácilmente a modelar datos en los cuales hay varios niveles de variación extra - binomial [70]. Si σ es la desviación estándar del efecto aleatorio del predictor lineal α_i , y si σ^2 es pequeño, La función de varianza asociada para este caso está relacionada con una expresión de la forma $\mu(1 - \mu)$ con ϕ el parámetro de dispersión dependiendo de σ^2 y de la misma expresión $\mu(1 - \mu)$. En este caso, la varianza de la variable respuesta es dada por

$$\text{Var}(Y_i) = m_i \pi_i (1 - \pi_i) [1 + \sigma^2 (m_i - 1) \pi_i (1 - \pi_i)] \quad (2.14)$$

(ver [70] y [39]). Este modelo recibe el nombre de modelo logístico mixto. Si $\boldsymbol{\alpha} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, y como es usual $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$, el modelo recibe el nombre de logístico normal o binomial normal si $g(\cdot)$ es genérico. Sin embargo, a lo largo de este escrito se hace referencia al modelo logístico normal y al binomial normal indistintamente.

2.3. Otros modelos de sobredispersión

Entre los modelos alternativos de sobredispersión más mencionados en la literatura se encuentran los modelos de cuasi-verosimilitud, la familia doblemente exponencial y los modelos lineales generalizados con media aleatoria, brevemente mencionados en la sección anterior como una generalización del modelo Poisson Normal o el modelo Binomial Normal. Mencionamos brevemente el modelamiento por el enfoque de la cuasi-verosimilitud y el de la familia doblemente exponencial.

2.3.1. Modelos de cuasi - verosimilitud

En 1974, Wedderburn publica su artículo proponiendo un nuevo tipo de modelamiento para datos con respuesta diferente a la normal. En tal caso, se debe conocer de antemano la relación funcional existente entre la media y la varianza de la variable respuesta, es decir se debe conocer $V(\mu)$ con $E(Y) = \mu$. Supone también que cada μ_i puede ser modelada mediante un conjunto de variables exógenas por medio de un vector de parámetros β como es habitual. En este sentido, los modelos de cuasi - verosimilitud, pueden ser vistos como una generalización de los MLG, ya que solamente se asume una función de varianza conocida y una relación funcional entre dicha media que define el comportamiento de la varianza y el vector de parámetros β y no se requiere el conocimiento de la distribución de la variable respuesta. Bajo este escenario, se asume que

$$E(Y) = \mu(\beta) \quad \text{y} \quad \text{Var}(Y) = \sigma^2 V(\mu) \quad (2.15)$$

en donde σ^2 es un parámetro de dispersión [55]. Dada la estructura de la varianza, el modelo permite flexibilidad en el manejo de la misma, en donde su estructura no depende únicamente de la relación funcional $V(\mu)$ sino también del parámetro σ^2 . El logaritmo de la función de cuasi - verosimilitud es definido como

$$Q(\mu; y) = \frac{1}{\sigma^2} \int_y^\mu \frac{y - t}{V(t)} dt \quad (2.16)$$

en donde se puede verificar fácilmente que $Q(\mu; y)$ tiene propiedades semejantes a la función de verosimilitud usual (ver [55]). La distribución de la variable respuesta queda determinada por la especificación de $V(\mu)$ cuando esta coincide con una perteneciente a la familia exponencial. El artículo en cuestión muestra que la forma de estimación de los parámetros es hecha mediante la maximización de Q por medio del método iterativo de Fisher Scoring.

Como se mencionó anteriormente, estos métodos permiten una mayor flexibilidad en el manejo de la varianza. Por ejemplo cuando se especifica la función de varianza $V(\mu) = \mu$ o $V(\mu) = \mu(1 - \mu)$ en los respectivos casos Poisson o binomial, la varianza será $\text{Var}(Y) = \sigma^2 \mu$ y $\text{Var}(Y) = \sigma^2 \mu(1 - \mu)$. Se permite modelar sobredispersión si $\sigma^2 > 1$. Los resultados en las estimaciones de estos mismos, por su estructura y su construcción, son muy similares a los obtenidos por medio de los MLG clásicos; solamente difieren por supuesto en los valores de σ^2 . Este debe ser estimado separadamente y el método convencional para hacerlo es el de los momentos, que toma la forma general

$$\hat{\sigma}^2 = \frac{1}{n - p} \sum_{i=1}^n \frac{y_i - \hat{\mu}_i}{V(\hat{\mu}_i)}. \quad (2.17)$$

Si se supone que Y_1, \dots, Y_n son variables aleatorias independientes con logaritmo de función de cuasi - verosimilitud $Q(\mu_i; y_i)$, $i = 1, \dots, n$, se tiene que $Q(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^n Q(\mu_i; y_i)$ y además se supone que $g(\mu_i) = \eta_i$, se deduce que los MLG son un caso especial de los modelos de cuasi verosimilitud. Por lo tanto, algunas pero no todas, las funciones de cuasi verosimilitud corresponden a una verdadera función de verosimilitud para μ [55].

2.3.2. Familia doblemente exponencial

Este tipo de modelamiento puede ser usado para generalizar cualquier modelo de regresión, pero el principal interés del escrito de Efron en 1986 es en modelos binomial y Poisson [26]. Los modelos mencionados en este artículo, son basados en lo que denominan familia doblemente exponencial, lo que permite convenientemente modelar la media y la varianza mediante unas variables regresoras de manera simultánea e independiente.

Esta idea surge de revisar la restricción de que el parámetro de dispersión σ^2 es asumido constante en un modelo de cuasi verosimilitud. Consiste en generalizar un modelo de cuasi - verosimilitud de acuerdo a

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad \text{Var}(Y_i) = \sigma_i^2 V(\mu_i) \quad \text{y} \quad h(\sigma_i^2) = \mathbf{z}_i^T \boldsymbol{\gamma} \quad (2.18)$$

donde $h(\cdot)$ es otra función de enlace, positiva y doblemente diferenciable, $\boldsymbol{\gamma}$ es un vector de parámetros, \mathbf{z}_i es un vector de variables aleatorias, posiblemente diferentes a \mathbf{x}_i , y las funciones $g(\cdot)$, V , y los vectores \mathbf{x}_i y $\boldsymbol{\beta}$ como en la sección (2.1) [56]. En el caso específico de los MLG, si Y_1, Y_2, \dots, Y_n es una muestra aleatoria independiente con media μ_i podemos escribir esta generalización de la forma

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad \text{Var}(Y_i) = \phi_i V(\mu_i) \quad \text{y} \quad h(\phi_i) = \mathbf{z}_i^T \boldsymbol{\gamma} \quad (2.19)$$

Es claro que este enfoque es una generalización, ya que si $h(\phi_i) = \gamma$, es decir, si $\mathbf{Z} = [\mathbf{1}; \mathbf{0}]$ con $\mathbf{1}$ un vector columna con el valor 1 en todas sus entradas, cae en el caso de los modelos lineales generalizados (ver [31]).

2.4. Modelos de sobredispersión generalizados

Estos modelos que tratan la sobredispersión, son propuestos por Quintero, Cepeda y Nuñez en 2012 [57]. La propuesta básica consiste en asumir que el parámetro de sobredispersión varía como una función de un conjunto de valores tomados por algunas variables explicativas presentes en los datos, y así puede ser modelado como una función de esas variables. También, se consideran modelos con media aleatoria en los cuales la varianza del componente aleatorio del predictor lineal es modelado por una función de algunas variables explicativas [57].

2.4.1. Modelos asociados a la distribución Poisson

En la sección (2.2.1.1) se trataron los modelos usuales de sobredispersión para el caso de la respuesta Poisson. El primero considerado es el modelo binomial negativo. Se asume entonces que el parámetro de dispersión no es constante y que este varía de acuerdo a un conjunto de variables explicativas \mathbf{Z} [57]. Siguiendo la misma notación, la función de densidad de probabilidad resultante es

$$Pr(Y_i = y_i) = \binom{y_i + \phi_i - 1}{y_i} (1 - \pi_i)^{\phi_i} \pi_i^{y_i} \quad (2.20)$$

con $\pi_i = \mu_i/(\mu_i + \phi_i)$, de forma tal que un modelamiento conjunto de la media y la dispersión es considerado. Por ejemplo, se puede tener

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \quad \text{y} \quad \log(\phi_i) = \mathbf{z}_i^T \boldsymbol{\gamma} \quad (2.21)$$

siguiendo con la función de enlace canónica en el predictor lineal de la media.

De igual manera, el otro modelo considerado en esa sección es el Poisson normal. Se recuerda que se tiene la inclusión en el predictor lineal de un coeficiente aleatorio de forma

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i \quad (2.22)$$

en donde en este caso en específico, $\alpha_i \sim N(0, \varsigma_i)$, y ς_i varía como una función de algún conjunto de variables explicativas al igual que en el caso anterior. De manera similar al caso de la binomial negativa, se piensa en una función de enlace para la varianza como el logaritmo natural.

El uso de estos modelos debe ser considerado cuando hay signos de sobredispersión no constante en los datos. Para efectos de verificación si la sobredispersión varía como una función de un conjunto de variables explicativas, es suficiente chequear si los parámetros asociados a estas variables son estadísticamente significativos [57]. Los parámetros que definen estos modelos pueden ser ajustados desde el enfoque clásico o Bayesiano; en este caso el punto de partida es el último mencionado.

2.4.2. Modelos asociados a la distribución binomial

En la sección (2.2.1.2) se trataron los modelos usuales de sobredispersión para el caso de la respuesta binomial. El primer modelo en ser generalizado es el beta binomial. Allí se revisó que $Y_i | \pi_i \sim Bin(m_i, \pi_i)$ y que la distribución de Y_i incondicionada es beta binomial de forma que su varianza es de la forma (2.9). El objetivo es describir esta varianza en la forma extendida por $\text{Var}(Y_i) = m_i \pi_i (1 - \pi_i) (1 + (m_i - 1) \delta_i)$. En esta vía, se permite que el parámetro de dispersión δ_i varíe de acuerdo a un conjunto de variables explicativas. El objetivo subyacente de esta nueva propuesta es el modelamiento conjunto de la media y del parámetro de sobredispersión bajo el supuesto de que la variable respuesta sigue una distribución beta binomial, un supuesto que permite la posibilidad de tener una varianza más grande que la considerada por el MLG estándar o el modelo de sobredispersión beta binomial de la sección (2.2.1.2) [57].

El segundo modelo generalizado es el modelo binomial normal. En este caso, la variación extra es incorporada por medio de la inclusión del coeficiente aleatorio como en la expresión (2.13) con $\text{Var}(\alpha_i) = \sigma_i^2$. En tal caso, σ_i^2 puede ser modelado como una función de un conjunto de variables explicativas al igual que el caso de la Poisson normal (ver [57], [13], [11]).

2.5. Modelos de sobredispersión autocovariantes

Si se observa la estructura de todos los modelos que se han revisado anteriormente, capaces de resumir y controlar el problema de la sobredispersión, se debe tener en cuenta que el fenómeno en cuestión puede aparecer como consecuencia de la disposición geográfi-

ca de los datos, es decir, por la autocorrelación espacial (no independencia de la variable respuesta). Entre muchas otras causas, esta es la que mayor interés presenta en nuestro estudio, por los objetivos y características que las variables bajo estudio enmarcan a lo largo de la estructura espacial de la información. Sin embargo, pese a esto, cada uno de los modelos sobredispersos usuales que se revisaron con anterioridad, atienden de manera suficiente este problema, corrigiendo por sí mismos la presencia de la correlación entre las unidades espaciales. No obstante, cuando evidentemente hay un objetivos de cuantificar la presencia de la sobredispersión de acuerdo a un esquema de vecindad generado, es imprescindible usar un modelo capaz de capturar mediante un parámetro toda la información disponible que la estructura espacial definida pueda brindar. Es en este sentido en que estos modelos propuestos toman sentido. El principal interés es cuantificar la afectación espacial de una variable bajo estudio cuando esta presenta una respuesta Poisson o binomial, y está claramente acompañada de una estructura sobredispersa. En tal caso, el modelo soporta, recoge y cuantifica esta característica intrínseca que por sí sola la autocovariable definida no es capaz de capturar. La estimación de los parámetros de este modelo, puede ser hecha entre otros mediante el enfoque clásico vía máxima verosimilitud y el enfoque Bayesiano. El presente escrito se ocupa de la estimación Bayesiana.

2.5.1. Modelos asociados a la distribución Poisson

Sea S un sistema espacial con n unidades y sea $Y(\cdot)$ un proceso medido en cada una de las unidades espaciales, es decir se tiene $Y(R_i)$ en donde $\{R_1, R_2, \dots, R_n\}$ forma S . Por simplicidad, se hace $Y(R_i) = Y_i$. Dada una estructura de vecindad, y una matriz de pesos espaciales \mathbf{W} con las características mencionadas en el capítulo anterior, de manera general, se supone que

$$Pr(Y_i = y_i | Y_{(-i)}) = \frac{\exp(-\lambda_i) \lambda_i^{y_i}}{y_i!} \quad (2.23)$$

con $Y_{(-i)} = Y(R_{(-i)}) = \{Y(R_j) : j \neq i\}$. El modelo autocovariante Poisson es dado por la expresión

$$g(\lambda_i) = \mu_i + \rho \mathbf{W}_i \mathbf{Y} \quad (2.24)$$

con $E(Y_i | Y_{(-i)}) = \lambda_i$, en donde $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$, con \mathbf{X} una matriz diseño conocida, $\boldsymbol{\beta}$ un vector de parámetros a estimar, \mathbf{W} una matriz de pesos espaciales, \mathbf{Y} el vector del proceso Y y ρ un parámetro a ser estimado, denominado el parámetro de la autocovariable. Para el caso de la Poisson, usualmente $g(\cdot) = \log(\cdot)$. Por simplicidad también se hace $\mathbf{W}\mathbf{Y} = \mathbf{A}$. De esta manera el modelo autocovariante Poisson log lineal puede ser escrito como

$$\log(\lambda_i) = \mu_i + \rho A_i. \quad (2.25)$$

El siguiente modelo considerado es el modelo de sobredispersión binomial negativo. Este es el resultante de modelar en dos niveles, como se puede ver en la sección (2.2.1.1). En este caso se tiene que dados una estructura de vecindad y los valores del proceso de interés en las unidades espaciales, la distribución condicionada del proceso Y es binomial negativo.

Para esto se supone que $Y^* = Y | Y_{(-i)}$ y que $Y_i^* | Z = z \sim Pois(z)$, en donde $Z \sim G(\mu, \phi)$ en que ϕ no depende de μ . De esta manera, Y^* incondicionada tiene distribución binomial negativa como se puede apreciar en la sección (2.2.1.1). La función de densidad

de probabilidad resultante es

$$Pr(Y_i^* = y_i^*) = Pr(Y_i = y_i | Y_{(-i)}) = \binom{y_i + \phi - 1}{y_i} (1 - \pi_i)^\phi \pi_i^{y_i} \quad (2.26)$$

con $\pi_i = \mu_i / (\mu_i + \phi)$. Teniendo en cuenta que $E(Y_i | Y_{(-i)}) = \mu_i$, Se modela la parte sistemática

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho \mathbf{W}_i \mathbf{Y} \quad (2.27)$$

Usando la función de enlace logarítmica, se obtiene el modelo binomial negativo autocovariante con estructura

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho A_i \quad (2.28)$$

El siguiente modelo en presentar es el modelo Poisson normal autocovariante. Al igual que el caso inmediatamente anterior, suponemos que $Y^* = Y | Y_{(-i)}$ y que dado un vector de efectos aleatorios $\boldsymbol{\alpha}$, los conteos $Y_1^*, Y_2^*, \dots, Y_n^*$ son condicionalmente independientes tales que $Y_i^* | \boldsymbol{\alpha} \sim Pois(\mu_i)$ en donde

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho \mathbf{W}_i \mathbf{Y} + \mathbf{z}_i^T \boldsymbol{\alpha} \quad (2.29)$$

donde $\mu_i = E(Y_i | Y_{(-i)})$ con \mathbf{x}_i , \mathbf{W}_i y \mathbf{z}_i vectores fila de las matrices \mathbf{X} , \mathbf{W} y \mathbf{Z} respectivamente, las tres conocidas. Al igual que el caso simple relacionado en la sección (2.2.1.1), la función de enlace es $g(\mu) = \log(\mu)$ (función de enlace canónica), y para tratar sobredispersión se trabaja $\mathbf{Z} = \mathbf{I}_n$, con \mathbf{I}_n definida como antes. Para este caso específico

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho A_i + \alpha_i \quad (2.30)$$

con \mathbf{A}_i definida como arriba. Los siguientes modelos presentados son los modelos generalizados autocovariantes con respuesta Poisson. En la sección (2.4.1) se presentan los modelos propuestos por Quintero, Cepeda y Nuñez (véase [57]). De esta manera, se presenta el modelo binomial negativo generalizado autocovariante. Nuevamente se supone que $Y^* = Y | Y_{(-i)}$ y que $Y_i^* | z \sim Pois(z)$ con Z distribuido como antes. Se asume entonces que el parámetro de dispersión no es constante como en la sección (2.4.1) y que este varía de acuerdo a un conjunto de variables explicativas \mathbf{Z} . La función de densidad de probabilidad resultante es

$$Pr(Y_i^* = y_i^*) = Pr(Y_i = y_i | Y_{(-i)}) = \binom{y_i + \phi_i - 1}{y_i} (1 - \pi_i)^{\phi_i} \pi_i^{y_i} \quad (2.31)$$

con $\pi_i = \mu_i / (\mu_i + \phi_i)$ de forma tal que un modelamiento conjunto de la media y la varianza puede ser considerado. De este modo, se podría tener

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho \mathbf{W}_{1i} \mathbf{Y} \quad \text{y} \quad h(\phi_i) = \mathbf{z}_i^T \boldsymbol{\gamma} + \lambda \mathbf{W}_{2i} \mathbf{Y} \quad (2.32)$$

con $E(Y_i | Y_{(-i)}) = \mu_i$, $g(\cdot)$ y $h(\cdot)$ funciones de enlace, en donde \mathbf{W}_1 puede ser diferente de \mathbf{W}_2 , ambas matrices de pesos espaciales y λ un parámetro que define el comportamiento autocovariable o autorregresivo en la dispersión con base en la variable respuesta. Si se hace $A_{ji} = \mathbf{W}_{ji} \mathbf{Y}$ para $j = 1, 2$ y se usa la función de enlace logarítmica en ambos casos, se tiene

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho A_{1i} \quad \text{y} \quad \log(\phi_i) = \mathbf{z}_i^T \boldsymbol{\gamma} + \lambda A_{2i} \quad (2.33)$$

Nótese que el caso especial $\lambda \neq 0$ aparece cuando hay evidencias de que la dispersión de la variable respuesta depende en cierta medida de lo que ha pasado en los vecinos más cercanos con respecto a esa misma variable.

El siguiente modelo considerado es el modelo Poisson normal generalizado autocovariante. Se obtiene este modelo suponiendo que $Y^* = Y|Y_{(-i)}$ y que dado un vector de efectos aleatorios $\boldsymbol{\alpha}$, los conteos $Y_1^*, Y_2^*, \dots, Y_n^*$, son condicionalmente independientes tales que $Y_i^*|\boldsymbol{\alpha} \sim Pois(\mu_i)$ en donde el predictor lineal toma la forma (2.29). En tal caso, $\boldsymbol{\alpha} \sim N(0, \boldsymbol{\varsigma}_i)$ en donde $\boldsymbol{\varsigma}_i$ puede ser modelado por un conjunto de variables explicativas, incluso una nueva autocovariable, definida como en las estructuras del parámetro de dispersión en (2.33).

2.5.2. Modelos asociados a la distribución binomial

Siguiendo con la misma intención de la sección anterior, ahora se supone que el proceso Y dados los valores observados en la vecindad respectiva, sigue una distribución binomial. Esto es, $Y_i^* = Y_i|Y_{(-i)}$ tiene función de densidad de probabilidad

$$Pr(Y_i^* = y_i^*) = Pr(Y_i = y_i|Y_{(-i)}) = \binom{m_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m_i - y_i} \quad (2.34)$$

para $y_i = 0, 1, 2, \dots, m_i$, con $E(Y_i|Y_{(-i)}) = m_i \pi_i$. Un modelo logístico autocovariante, puede definirse desde la parte sistemática como sigue

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho \mathbf{W}_i \mathbf{Y} \quad (2.35)$$

con \mathbf{X} una matriz diseño conocida, $\boldsymbol{\beta}$ un vector de parámetros a estimar, \mathbf{W} una matriz de pesos espaciales, \mathbf{Y} el vector del proceso Y y ρ un parámetro a ser estimado, denominado el parámetro de la autocovariable, tal cual como el caso de la Poisson. La forma general de un modelo binomial autocovariante, es dado por la expresión (2.35) en donde $\log(\pi/(1-\pi))$ puede ser cambiado por una función $g(\cdot)$ que cumpla los requerimientos ya mencionados de una función de enlace. En cuanto a la matriz de pesos espaciales, en este escrito se presenta una forma de construcción basados en el conocimiento de la variable y_i^* y m_i para cada unidad espacial.

Para la construcción de un modelo autocovariante sobredisperso beta binomial, o modelo beta binomial autocovariante, se tiene lo siguiente. Suponemos nuevamente que los conteos $Y_i^* = Y_i|Y_{(-i)}$ siguen una distribución binomial con parámetros m_i y π_i . Como en la sección (2.2.1.2), se supone que $\Pi_i \sim Beta(a, b)$ con función de densidad de probabilidad dada por (2.10). De esta manera, la distribución de Y_i^* incondicionada tiene función de probabilidad

$$Pr(Y_i^* = y_i^*) = Pr(Y_i = y_i|Y_{(-i)}) = \frac{\binom{m_i}{y_i} B(a + y_i, m_i + b - y_i)}{B(a, b)} \quad (2.36)$$

para $y_i = 0, 1, 2, \dots, m_i$, con todo lo demás como en la sección (2.2.1.2) cuando se especificó que la distribución de Y_i^* es beta binomial. De esta manera se modela la media de Y_i^* por la parte sistemática

$$g(\pi_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho \mathbf{W}_i \mathbf{Y} \quad (2.37)$$

con $\pi_i = E(Y_i|Y_{(-i)})$. Usualmente en este caso $g(\pi_i) = \log(\pi_i/(1 - \pi_i))$ y por simplicidad de la notación, $A_i = \mathbf{W}_i \mathbf{Y}$ como antes. Luego la expresión (2.37) en específico para este caso, es

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho A_i \quad (2.38)$$

El siguiente modelo en ser presentado es el modelo binomial normal autocovariante. Se supone que dado un vector de efectos aleatorios $\boldsymbol{\alpha}$, las variables de conteo $Y_1^*, Y_2^*, \dots, Y_n^*$ son condicionalmente independientes tales que $Y_i|\boldsymbol{\alpha} \sim \text{Bn}(m_i, \pi_i)$ con Y_i^* como antes. De esta manera, la forma sistemática es dada por la ecuación

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho A_i + \mathbf{z}_i^T \boldsymbol{\alpha} \quad (2.39)$$

si la función de enlace escogida es la logística, con las especificaciones como antes. El modelo binomial normal es un caso especial de este que es un modelo binomial mixto. En tal caso, $\alpha_i \sim N(0, \sigma^2)$ y $\mathbf{Z} = \mathbf{I}_n$. El modelo queda entonces dado por la expresión

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho A_i + \alpha_i \quad (2.40)$$

Los siguientes modelos presentados son los modelos generalizados autocovariantes con respuesta binomial. En la sección (2.4.2) se muestra la propuesta hecha en el artículo de los autores Quintero, Cepeda y Nuñez (véase [57]). El primer modelo de sobredispersión con este tipo de respuesta en ser generalizado fue el beta binomial. Con base en la información de esta sección, suponga que se tienen los conteos Y_i^* como antes. De esta manera, $Y_i^*|\pi_i \sim \text{Bin}(m_i, \pi_i)$ y así, $\text{Var}(Y_i^*) = m_i \pi_i (1 - \pi_i) (1 + (m_i - 1) \delta_i)$. De esta forma se permite que δ_i varíe en función de algún conjunto de covariables. La propuesta en este término es un modelo con parte sistemática en media y varianza dada por

$$g(\pi_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho A_{1i} \quad \text{y} \quad h(\delta_i) = \mathbf{z}_i^T \boldsymbol{\gamma} + \lambda A_{2i} \quad (2.41)$$

respectivamente, con las especificaciones como en la sección anterior. En cuanto a la aplicación trabajada aquí por ejemplo, se tiene el modelo

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho A_{1i} \quad \text{y} \quad \log(\delta_i) = \mathbf{z}_i^T \boldsymbol{\gamma} + \lambda A_{2i} \quad (2.42)$$

Recuérdese que las matrices \mathbf{W}_j que definen las autocovariables A_j , $j = 1, 2$, pueden no ser iguales. El último modelo presentado en este escrito es el modelo binomial normal generalizado autocovariante. Este se obtiene suponiendo que las variables Y_i^* dado un vector de efectos aleatorios $\boldsymbol{\alpha}$ son distribuidas binomial. En el caso trabajado, se supone que $\alpha_i \sim N(0, \sigma_i^2)$ en donde es permitido modelar σ_i^2 por un conjunto de covariables. En tal caso

$$g(\pi_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \rho A_{1i} \quad \text{y} \quad h(\sigma_i^2) = \mathbf{z}_i^T \boldsymbol{\gamma} + \lambda A_{2i} \quad (2.43)$$

donde $m_i \pi_i = E(Y_i^*)$. En la aplicación, por ejemplo $g(\pi) = \log(\pi/(1 - \pi))$ y $h(\sigma^2) = \log(\sigma^2)$, con todo lo demás como el modelo anterior.

La siguiente sección muestra en breve los conceptos preliminares de la estadística Bayesiana y su uso en la estimación de los parámetros que definen la estructura de un modelo.

2.6. Estimación Bayesiana

Actualmente la inferencia Bayesiana surge como una propuesta alternativa a la inferencia clásica o frecuentista, atractiva por el desarrollo y la capacidad computacional adquirida en el campo científico durante las últimas décadas. Hacia los años 80, se creía que el enfoque bayesiano no tenía grandes ventajas prácticas en comparación a los demás, y se restringía su uso simplemente a las familias de distribuciones (modelos probabilísticos muy limitados) en donde su forma *a posteriori* resultara una distribución conocida por todos. Además de esto, se requería una especificación *a priori* del comportamiento probabilístico de los datos, lo que significaba esfuerzos por no cometer abusos en la limitada estructura computacional y en la información de carácter subjetivo por parte de los investigadores.

Sin embargo, la aparición de los métodos intensivos de simulación, como MCMC, y a su vez el desarrollo tecnológico en los dispositivos computacionales, ayudan a que en los últimos años, la estadística Bayesiana ocupe un lugar muy importante en el camino que disponen día a día los investigadores en el reto de encontrar herramientas que permitan un desarrollo probabilístico e inferencial de los fenómenos más sofisticadas y eficaces. Este capítulo reseña brevemente los principales puntos de la inferencia Bayesiana, y los métodos más usados de simulación estocástica, apoyándose en las referencias [52] y [28] entre algunos otros escritos. El software usado en general, para estos casos es WinBUGS y eventualmente, R. La primera sección se emplea para comentar sobre los métodos de simulación estocástica, pilar del uso contemporáneo de la metodología Bayesiana, y principales herramientas para el paso hacia la inferencia basada en la distribución de interés, la distribución *a posteriori*. En los siguientes se ocupa de la inferencia.

2.6.1. MCMC - Markov Chain Monte Carlo

Una cadena de Markov es un proceso estocástico de parámetro discreto con espacio de estados discreto, que satisface la propiedad de Markov, esto es, para cualquier entero $n \geq 0$, y para cualesquiera estados x_0, \dots, x_{n+1} , se cumple

$$p(x_{n+1}|x_0, \dots, x_{n+1}) = p(x_{n+1}|x_n) \quad (2.44)$$

Si el tiempo $n + 1$ se considera como un tiempo futuro, el tiempo n como el presente y los tiempos $0, 1, \dots, n - 1$ como el pasado, entonces la condición (2.44) establece que la distribución de probabilidad del estado del proceso al tiempo futuro $n + 1$ depende únicamente del estado del proceso al tiempo n , y no depende de los estados en los tiempos pasados [59]. Con base a esta definición, se supone que se está interesado en simular una distribución de probabilidad, denotada por $\pi(x)$. Aunque la función $\pi(\cdot)$ pueda ser escrita en forma cerrada, se supone que se está interesado en los momentos de π los cuales no pueden ser computados analíticamente. Además la simulación directa de π puede ser difícil o imposible, tal vez porque es de una dimensión muy alta, y así el uso del método simple de Monte Carlo para evaluar esos momentos, no es posible. MCMC ataca este problema, simulando de una cadena de Markov, cuya distribución invariante es π [21]. Como se verá, hay dos técnicas básicas usadas en MCMC, uno llamado el algoritmo *Metropolis Hasting* y el otro llamado el muestreador de *Gibbs*

2.6.1.1. Muestreador de Gibbs

El muestreador de Gibbs es uno de los algoritmos estocásticos más usados para simular datos de una distribución mediante cadenas de Markov. La idea de este algoritmo es básicamente descomponer un problema de simulación de dimensión alta a un problema más tratable; es tratar de reducirlo a una colección de simulaciones más sencillas de dimensiones más bajas [21]. El algoritmo es un esquema donde el kernel o probabilidad de transición es formado por todas las distribuciones condicionales completas [28]. El algoritmo procede como sigue:

Se asume que la estructura de probabilidad conjunta de una colección de variables aleatorias U_1, \dots, U_k , es tal que las formas de las densidades condicionales completas $[U_s|U_r, r \neq s]$, $s = 1, \dots, k$, son identificadas, al menos proporcionalmente. Dados valores de partida arbitrarios, $U^{(0)} = (U_1^{(0)}, \dots, U_k^{(0)})$ para las k variables aleatorias, se genera una variable aleatoria $U_1^{(1)}$ de $[U_1|U_2^{(0)}, \dots, U_k^{(0)}]$, seguido por $U_2^{(1)}$ de $[U_2|U_1^{(1)}, U_3^{(0)}, \dots, U_k^{(0)}]$ y así, hasta obtener $U_k^{(1)}$ de $[U_k|U_1^{(1)}, \dots, U_{k-1}^{(1)}]$. Esto completa una iteración del esquema, dando $U^{(1)} = (U_1^{(1)}, \dots, U_k^{(1)})$. Después de t iteraciones se tiene entonces, $U^{(t)} = (U_1^{(t)}, \dots, U_k^{(t)})$. La cadena de Markov así generada tiene distribución de equilibrio $[U_1, \dots, U_k]$, la distribución conjunta de U_1, \dots, U_k . Bajo débiles condiciones de regularidad, se tiene que cuando $t \rightarrow \infty$, $U^{(t)}$ tiende en distribución a la distribución de $[U_1, \dots, U_k]$, y el promedio ergódico $t^{-1} \sum_{j=1}^t g(U^{(j)})$ es un estimador consistente del valor esperado (con respecto a $[U_1, \dots, U_k]$) de cualquier función integrable $g(U)$ [67]. Para ver estas demostraciones, referirse por ejemplo, al apéndice del artículo [64]. Gamerman y Lopes [28], definen el algoritmo de la siguiente manera, un poco más precisa con respecto al enfoque de este escrito: Se tiene nuevamente el problema de simular la distribución de $\pi(\theta)$, esta vez, con $\theta = (\theta_1, \dots, \theta_d)$ un vector de parámetros. Se considera también al igual que como se mencionó en el caso anterior, que las distribuciones condicionales completas $\pi_i(\theta_i) = \pi(\theta_i|\theta_{-i})$, $i = 1, \dots, d$ están disponibles (esto significa que son completamente conocidas y que puede muestrearse de ellas). El problema a ser resuelto es muestrear de π cuando una generación directa es costosa complicada o simplemente cuando no es posible, pero cuando generaciones de π_i si es posible. El proceso puede ser descrito en los siguientes pasos:

Algoritmo 1 - El Muestreador de Gibbs

1. Inicie el contador de iteraciones de la cadena $j = 1$, y consiga un valor inicial $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$;
2. Obtenga un nuevo valor $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_d^{(j)})$ de $\theta^{(j-1)}$ a través de generaciones sucesivas de valores

$$\begin{aligned}
 \theta_1^{(j)} &\sim \pi(\theta_1|\theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}) \\
 \theta_2^{(j)} &\sim \pi(\theta_2|\theta_1^{(j-1)}, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}) \\
 &\vdots \\
 \theta_d^{(j)} &\sim \pi(\theta_d|\theta_1^{(j-1)}, \dots, \theta_{d-1}^{(j-1)})
 \end{aligned}$$

3. Cambie el contador de j a $j + 1$ y regrese al paso 2 hasta que la cadena converja.

Cuando la convergencia se tiene, el valor resultante $\theta^{(j)}$ es una muestra aleatoria de π . Como el número de iteraciones incrementa, la cadena se aproxima a su condición de equilibrio. La convergencia se mantiene aproximada [28]. El muestreador de Gibbs, como se había mencionado, ayuda en ocasiones cuando hay que simular datos de una función de alta complejidad, llevándolo a una dimensión más tratable. El muestreador de Gibbs, sin embargo, paga un precio por reducir dicha dimensión de las simulaciones que se realizan. Si los componentes, en este caso $\theta_1^{(j)}, \dots, \theta_d^{(j)}$ exhiben alta dependencia, el algoritmo convergerá lentamente y ocasionará resultados altamente correlacionados [21].

2.6.1.2. Algoritmo Metropolis - Hastings

El algoritmo básico de Metropolis - Hastings, ofrece una familia flexible de algoritmos MCMC, complementarios al muestreador de Gibbs. En esencia, el método usa un mecanismo de acepto/rechazo (accept / reject) para corregir una cadena de Markov arbitraria hasta alcanzar la distribución invariante de π [21]. Considérese de nuevo una distribución π (estacionaria) de la cual una muestra debe ser obtenida vía cadenas de Markov. Una vez más cabe mencionar que esta tarea solamente tomará sentido si la generación no iterativa de π es muy complicada o costosa [28]. Dada una arbitraria *densidad candidata generadora* $q(\theta, \phi)$ (Dellaportas, P. and Roberts, G.O. [21] la describen como probabilidad de transición arbitraria, es decir, la densidad de probabilidad de moverse de ϕ a θ) que sea fácil simular de ella, el algoritmo Metropolis - Hastings procede como sigue: Se comienza con un valor inicial arbitrario θ_0 y se simula ϕ de $q(\theta_0, \cdot)$ junto con una variable aleatoria uniforme u_0 sobre el intervalo $(0, 1)$. Si

$$u_0 < \alpha(\theta_0, \Phi) := \min \left[\frac{\pi(\phi)q(\theta_0, \phi)}{\pi(\theta_0)q(\phi, \theta_0)}, 1 \right] \quad (2.45)$$

es decir, con probabilidad $\alpha(\theta_0, \phi)$, se acepta ϕ como candidato y se hace $\theta_1 := \phi$. En otro caso, rechaza ϕ y se hace $\theta_1 = \theta_0$. Esto representa una iteración del método [63]. De manera más específica, Gamerman y Lopes, escriben el algoritmo de la siguiente manera:

Algoritmo 2 - El algoritmo Metropolis - Hastings

1. Inicie el contador de iteraciones de la cadena $j = 1$, y consiga un valor inicial arbitrario $\theta^{(0)}$.
2. Mueva la cadena a un nuevo valor ϕ generado de la densidad $q(\theta^{(j-1)}, \cdot)$.
3. Evalúe la probabilidad de aceptación del movimiento $\alpha(\theta^{(j-1)}, \phi)$. Si el movimiento es aceptado, $\theta^{(j)} = \phi$. Si no es aceptado $\theta^{(j)} = \theta^{(j-1)}$ y la cadena no se mueve.
4. Cambie el contador de j a $j + 1$ y retorne al paso 2 hasta obtener convergencia

El paso 3 es realizado después de la generación de una cantidad independiente u . Si $u \leq \alpha$, el movimiento es aceptado y si $u > \alpha$ el movimiento no es permitido [28]. Un

ejemplo de este algoritmo, es el muestreador independiente (*the independence sampler*). Consiste en la elección más simple posible de la distribución propuesta q , de forma que sea independiente de su primer argumento:

$$q(\boldsymbol{\theta}, \boldsymbol{\phi}) = q(\boldsymbol{\phi}) \quad (2.46)$$

donde el dominio de π está contenido en el de q . Aquí se puede escribir

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\phi}) = \min \left[\frac{\omega(\boldsymbol{\phi})}{\omega(\boldsymbol{\theta})}, 1 \right] \quad (2.47)$$

donde

$$\omega(\boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}. \quad (2.48)$$

Es claro que tomando q proporcional a π , ω es constante y el algoritmo se reduce, como se podría esperar, a una muestra aleatoria independiente e idénticamente distribuida de π [21].

Dejando de lado por ahora los métodos más comunes y usados de MCMC útiles para muestrear de funciones de probabilidad complejas, se reseñan algunas de las características más importantes de la inferencia Bayesiana en las siguientes secciones. Para un complemento detallado de algunos métodos MCMC implementados en R, y sus correspondientes resultados teóricos, se recomienda ver [61] y [28].

2.6.2. Inferencia Bayesiana

Como punto de partida del enfoque Bayesiano como herramienta inferencial estadística, está el teorema de Bayes. Antes de mencionarlo, se debe hacer mención a la información *a priori*, que se usará para desglosar el tema Bayesiano. Como se ha venido mencionando, información *a priori* debe estar disponible. Se supone que el modelo probabilístico que se desea establecer $p(\cdot)$, está determinado por un número, vector o una matriz θ de valor desconocido, que toma valores en un espacio definido Θ . Con base en este modelo $p(\cdot)$, lo que buscamos en realidad es tomar alguna decisión con respecto al fenómeno bajo estudio, que se ve afectado por θ . Dicha cantidad desconocida θ , la cual afecta el proceso de decisión, es comúnmente llamada el estado de la naturaleza [7].

Como el objetivo no es establecer una discusión sobre el impacto que puedan sufrir las inferencias, las estimaciones y las decisiones sobre lo subjetivo de la elección de la información, ya que posiblemente dicha fuente de información puede no ser experimental, se supone como punto de partida que se tiene información *a priori* del comportamiento de la cantidad θ (para una discusión ver por ejemplo [7] y [52]). Supóngase que H denota dicha información. Ahora se asume que esta información es expresada en términos probabilísticos. Puede ser resumido de forma que la información disponible es $p(\theta|H)$ y, si la información contenida en H es suficiente para los propósitos inferenciales, esto es todo lo que se necesita. En este caso, la descripción de la incertidumbre sobre θ es completa [52].

Se supone que un conjunto de observaciones \mathbf{X} relacionado con la cantidad desconocida, es observado. Se asume que la nueva información \mathbf{X} sigue un modelo probabilístico dado por $p(\mathbf{X}|\theta, H)$. Con base en esta información, se desea determinar el modelo probabilístico $p(\theta|\mathbf{X}, H)$ conocida como distribución *a posteriori* de la cantidad θ conocida como parámetro. Se nota así, que desde el enfoque Bayesiano, para hacer inferencia sobre

el parámetro, se le asume un modelo probabilístico y así, se considera una cantidad estocástica. Por comodidad, se omite H de la notación, ya que aparece en todos los términos, sin embargo, no debe ser olvidado [52].

2.6.2.1. Distribución a *posteriori* - Teorema de Bayes

El análisis Bayesiano es hecho combinando la distribución a *priori* y la información que brinda la muestra \mathbf{X} . La distribución a *posteriori* de θ dada la información de la muestra \mathbf{X} y dado el conocimiento a *priori* del parámetro, o simplemente la distribución a *posteriori*, es dada por el teorema de Bayes. Nótese que θ y \mathbf{X} tienen densidad conjunta dada por

$$p(\mathbf{X}, \theta) = p(\theta)p(\mathbf{X}|\theta) \quad (2.49)$$

y que \mathbf{X} tiene densidad de probabilidad marginal dada por

$$p(\mathbf{X}) = \int_{\Theta} p(\mathbf{X}, \theta) d\theta. \quad (2.50)$$

Teorema 1 (Teorema de Bayes). *Sean $p(\theta)$ y $p(\mathbf{X})$ las distribuciones de probabilidad a priori y marginal de \mathbf{X} respectivamente. Si $p(\theta, \mathbf{x}) = p(\mathbf{x}, \theta)$ es la distribución de probabilidad conjunta de las dos cantidades estocásticas, entonces la distribución de probabilidad de la expresión $\theta|\mathbf{X}$ es dada por*

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})} \quad (2.51)$$

Al término (2.51) se le conoce como *distribución a posteriori*.

La distribución a *posteriori* del parámetro, es la que brinda su información para realizar las inferencias necesarias del fenómeno bajo estudio. Para demostrar el teorema, basta con usar el hecho de que

$$p(\theta|\mathbf{x}) = \frac{p(\theta, \mathbf{x})}{p(\mathbf{x})} \quad (2.52)$$

y como

$$p(\mathbf{x}|\theta) = \frac{p(\theta, \mathbf{x})}{p(\theta)} \quad (2.53)$$

después de unos elementales movimientos algebraicos, se tiene lo que se quiere.

El resultado del teorema de Bayes, que permite encontrar la distribución a *posteriori* del parámetro, puede ser escrito de la siguiente manera:

$$p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta) \quad (2.54)$$

ya que $p(\mathbf{x})$ no depende de la cantidad θ y puede ser considerada como una constante. El teorema da una regla para actualizar probabilidades en función de θ , comenzando desde $p(\theta)$ y llegando a $p(\theta|\mathbf{x})$ después de observar una muestra \mathbf{x} que da información

sobre el comportamiento del parámetro. Esta es la razón por la cual las distribuciones son llamadas *distribución a priori* y *distribución a posteriori* respectivamente [52]. Sin embargo, el concepto de distribuciones *a priori* y *a posteriori* son siempre relativas a la observación considerada en el momento dado. Es posible que después de observar \mathbf{x} y obtener la densidad *a posteriori*, un nuevo conjunto de observaciones \mathbf{y} también relacionada a θ a través de una eventual función de verosimilitud distinta ($p(\mathbf{Y}|\theta)$ distinta a $p(\mathbf{X}|\theta)$) o en otros términos $p_Y(\mathbf{y}|\theta) \neq p_X(\mathbf{x}|\theta)$) esté disponible. En este caso, la distribución *a posteriori* (relativa a \mathbf{X}) es la distribución *a priori* (relativa a \mathbf{Y}) y una nueva información *a posteriori* puede ser obtenida por una nueva aplicación del teorema de Bayes [28]. Cabe mencionar que estos resultados se muestran para el caso univariado pero son extensivos a todas las dimensiones.

2.6.2.2. Distribución *a priori*

Para determinar la función de densidad o modelo probabilístico *a priori* de θ , se pueden tomar distintas vías. De manera muy general, cada una de ellas se puede clasificar o en técnicas informativas o en técnicas no informativas. Por lo general, como se había mencionado, muchas de las técnicas que ofrecen una distribución *a priori* informativa, resultan siendo subjetivas. Vale la pena recordar que [7] hace una discusión al respecto. Hay muchas maneras de determinar de manera subjetiva la distribución *a priori*. Si el espacio paramétrico Θ es discreto, el problema se reduce simplemente a determinar probabilidades a cada elemento de Θ . Si el espacio paramétrico es continuo, esta tarea se dificulta un poco [7]. Varias técnicas en este último caso, son mostradas en el texto de Berger, 1985, y aquí solamente se mencionan unas. Las más importantes son el enfoque del histograma. Se trata de dividir el espacio paramétrico Θ en intervalos, determinar la probabilidad subjetiva de cada uno de ellos, y luego dibujar un histograma de probabilidad (una densidad). En este caso, una densidad $p(\theta)$ puede ser obtenida. La segunda técnica dada por este autor, es la del enfoque relativo de la verosimilitud. Consiste simplemente en comparar la verosimilitud intuitiva de varios puntos de Θ y esbozar directamente una densidad *a priori* desde esas determinaciones. La tercera técnica es emparejando una forma funcional dada. El autor señala que simplemente se asume $p(\theta)$ es de una forma funcional predeterminada y que luego se asume un modelo probabilístico eligiendo la densidad que más forma similar tenga. Por último comenta un método que consiste en la determinación de la función de probabilidad acumulada o función de distribución *a priori*. Esto es hecho, determinando algunas probabilidades acumuladas α (percentiles o fractiles) y $z(\alpha)$, dibujando los puntos $(z(\alpha), \alpha)$ y bosquejando una curva suave uniéndolos. Antes de hablar de distribuciones *a priori* informativas y no informativas, es necesario recordar la noción de distribuciones conjugadas. La definición es tomada del texto de Migon y Gamerman.

Definición 1 (Distribuciones conjugadas). Sea $F = \{p(x|\theta), \theta \in \Theta\}$ una familia de distribuciones muestrales. Una clase de distribuciones P se dice que es una familia conjugada con respecto a F si para toda $p(\theta|x) \in F$ y $p(\theta) \in P$, se tiene que $p(\theta|x) \in P$.

En efecto, en otras palabras, esta situación se da cuando para un modelo observacional dado, las distribuciones *a priori* y *a posteriori* pertenecen a la misma clase de distribuciones. Esto brinda ventajas para el subsecuente proceso inferencial, debido a la simplificación del análisis que queda restringido a un conjunto pequeño de posibles distribuciones. Así, pasar de una distribución *a priori* a una *a posteriori* solo requiere un cambio en los hiperparámetros sin cálculo adicional [28]. La distribución de θ se puede actualizar si se

dispone de más información, y los cálculos no se complican. Las principales familias de distribuciones y sus conjugadas, son las siguientes:

- Para la distribución binomial, la familia de distribuciones de beta es su conjugada.
- La distribución normal es conjugada de sí misma, siempre y cuando la varianza sea conocida.
- La distribución Gamma es la conjugada de la distribución Poisson.
- La distribución Gamma es la conjugada de la distribución exponencial, un caso particular de sí misma.
- Para la distribución multinomial, la conjugada es la distribución de Dirichlet.
- La distribución Gamma es la conjugada de una distribución normal de varianza desconocida.
- La distribución normal - gamma o Normal - χ^2 es la conjugada de una normal con media y varianza desconocida.

Para ver mejor todos estos resultados, el lector debe referirse al texto de Migon y Gamerman.

Distribuciones a *priori* informativas

Cuando se hace referencia a distribuciones informativas, se establece que la información que recoge el modelo probabilístico a *priori*, influye de manera certera y precisa en la determinación del comportamiento a *posteriori* del parámetro. Cuando se establece una familia de distribuciones como información a *priori* del parámetro, de cualquier manera porta y brinda información sobre su comportamiento. Por eso, en este caso las distribuciones conjugadas pueden determinarse como informativas, ya que, como se mencionó anteriormente; están informando sobre el comportamiento probabilístico del fenómeno antes de que haya algún tipo de observación en el estudio.

Distribuciones a *priori* no informativas

Lo que se quiere establecer con estas distribuciones, es que el investigador pueda realizar sus inferencias sin la necesidad de trabajar con demasiada información subjetiva predispuesta en la distribución a *priori*. La primera idea que se trabajó fue la de proponer una distribución a *priori* uniforme, en donde, por la estructura de esta, se puede suponer que no hay información disponible o simplemente no se debe o quiere usar. En este caso, la función de probabilidad toma la forma $p(\theta) = k$ para θ variando en un conjunto determinado, lo que implica que no hay preferencia de ningún valor de θ . Las limitaciones con respecto a esta idea son realmente muchas. En particular, la especificación de los límites de la distribución, muestra que si hay algún tipo de información brindada. Una pequeña discusión con respecto a esta alternativa, se puede encontrar en el texto de Migon y Gamerman. Otra alternativa un poco más elaborada de trabajar con distribuciones a *priori* no informativas, es la de Jeffrey (1961). Lo que se busca con esta propuesta, es de dar la menor información a *priori* posible. La siguiente definición es la versión tomada del texto de Migon y Gamerman.

Definición 2 (La a priori No Informativa de Jeffrey). *Considere una observación x con función de densidad de probabilidad $p(x|\theta)$. La a priori No Informativa de Jeffrey tiene densidad dada por*

$$p(\theta) \propto [I(\theta)]^{1/2}, \quad \theta \in \Theta. \quad (2.55)$$

En el caso multivariado, la densidad es dada por

$$p(\boldsymbol{\theta}) \propto |\mathbf{I}(\boldsymbol{\theta})|^{1/2}, \quad \boldsymbol{\theta} \in \Theta. \quad (2.56)$$

Algunas buenas propiedades de esta propuesta son mostradas en [52], con algunos buenos ejemplos.

2.6.2.3. Modelos de regresión

Esta sección se ocupa de explicar de manera general el procedimiento de estimación de los parámetros en un modelo de regresión lineal. Sin embargo, en esencia es el mismo procedimiento para modelos más extensos, como el caso de los modelos lineales generalizados. Cepeda, E. y Gamerman, D. [12], proponen un algoritmo para estimar modelos con estructuras similares a los modelos lineales generalizados. Se propone como base de partida el modelo lineal general. Supóngase que se tiene un modelo de regresión, expresado de la siguiente manera:

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \quad (2.57)$$

en donde \mathbf{X} es la matriz diseño compuesta por n individuos (filas) y $p + 1$ columnas (p variables explicativas), $\boldsymbol{\beta}$ es un vector de $p + 1$ parámetros a ser estimados, \mathbf{I}_n es una matriz identidad de tamaño $n \times n$ y σ^2 el parámetro de dispersión a ser estimado. Desde el punto de vista Bayesiano, el modelo (4.14) se completa con una distribución a priori de los parámetros $(\boldsymbol{\beta}^T, \sigma^2)$ [28]. En este caso se pueden establecer las distribuciones

$$\boldsymbol{\beta}|\sigma^2 \sim N(b_0, \sigma^2 B_0) \quad \text{y} \quad \sigma^2 \sim IG\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \quad (2.58)$$

Bajo esta parametrización, se tiene que $n_0 S_0 / \sigma^2 \sim \chi_{n_0}^2$. Si se hace $\phi = \sigma^{-2}$. En términos de ϕ , la a priori es $NG(b_0, B_0, n_0, S_0)$ (normal - gamma) o equivalentemente

$$\boldsymbol{\beta}|\phi \sim N(b_0, \phi^{-1} B_0) \quad \text{y} \quad \phi \sim G\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \quad (2.59)$$

y así, la densidad del vector $(\boldsymbol{\beta}^T, \phi)$ es definida por $p(\boldsymbol{\beta}, \phi) \propto p(\boldsymbol{\beta}|\phi)p(\phi)$, y se tiene

$$p(\boldsymbol{\beta}, \phi) \propto \phi^{[(n_0+p+1)/2]-1} \exp\left\{-\frac{\phi}{2}[n_0 S_0 + (\boldsymbol{\beta} - b_0)^T B_0^{-1}(\boldsymbol{\beta} - b_0)]\right\} \quad (2.60)$$

El modelo (2.57) puede escribirse también como

$$\mathbf{y}|\boldsymbol{\beta}, \phi \sim N(\mathbf{X}\boldsymbol{\beta}, \phi^{-1} \mathbf{I}_n) \quad (2.61)$$

de donde se establece que $l(\boldsymbol{\beta}, \phi) = f(y|\boldsymbol{\beta}, \phi)$. De esta manera,

$$l(\boldsymbol{\beta}, \phi) = \phi^{-n/2} \exp \left\{ -\frac{\phi}{2} [Q(\boldsymbol{\beta}) + S_e] \right\} \quad (2.62)$$

donde $Q(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$, $\boldsymbol{\beta}$ es la solución habitual de mínimos cuadrados de las ecuaciones normales y $S_e = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$.

Ahora, para encontrar la distribución a *posteriori* de la conjunta de los parámetros de interés, se aplica la proporcionalidad $p(\boldsymbol{\beta}^T, \phi | \mathbf{y}, \mathbf{X}) \propto l(\boldsymbol{\beta}^T, \phi) p(\boldsymbol{\beta}^T, \phi)$, lo que resulta en una expresión como la que sigue:

$$\phi^{d/2} \exp \left\{ -\frac{\phi}{2} (\boldsymbol{\beta} - b_1)^T B_1^{-1} (\boldsymbol{\beta} - b_1) \right\} \phi^{\frac{n_1}{2}-1} \exp \left\{ -\frac{\phi}{2} n_1 S_1 \right\} \quad (2.63)$$

donde $n_1 = n_0 + n$, $n_1 S_1 = n_0 S_0 + (\mathbf{y} - \mathbf{X}b_1)^T \mathbf{y} + (b_0 - b_1)^T B_0^{-1} b_0$, $b_1 = B_1 (B_0^{-1} b_0 + \mathbf{X}^T \mathbf{Y})$ y $B_1^{-1} = B_0^{-1} + \mathbf{X}^T \mathbf{X}$. Si se compara (2.63) con (2.60), se puede asegurar que tienen la misma forma distribucional las densidades a *priori* y a *posteriori*. En tal caso, la distribución a *posteriori* de los parámetros de interés es $NG(b_1, B_1, n_1, S_1)$, es decir, las densidades a *posteriori* de cada uno de los parámetros separables son

$$\boldsymbol{\beta} | \phi \sim N(b_1, \phi^{-1} B_1) \quad \text{y} \quad \phi \sim G\left(\frac{n_1}{2}, \frac{n_1 S_1}{2}\right) \quad (2.64)$$

[28]. El objetivo final de esta sección, es observar que a partir de la distribución que se asume que tienen los datos y la distribución a *priori* especificada para los parámetros del modelo, se encuentra la distribución a *posteriori* que no siempre es tratable como en el caso simple de la regresión normal. Con base en ella, necesitamos hacer inferencia sobre los parámetros, usando entonces métodos intensivos de muestreo para generar muestras de esta distribución obtenida. Es exactamente el mismo caso para los modelos lineales generalizados, solo que algunas de las distribuciones pertenecientes a la familia exponencial no tienen una distribución conjugada, lo que al final resultará en distribuciones a *posteriori* que no pueden ser tratadas analíticamente, lo que dificulta la inferencia bayesiana exacta. Desde el enfoque clásico, en el caso de los modelos lineales generalizados, la distribución asintótica del vector de parámetros estimado por el método de máxima verosimilitud y maximizado por el método numérico de Fisher Scoring $\hat{\boldsymbol{\beta}}$ es normal de media el valor del vector paramétrico $\boldsymbol{\beta}$ y matriz de varianzas y covarianzas $\hat{\phi} \mathbf{I}^{-1}(\boldsymbol{\beta})$ con $\hat{\phi}$ un estimador consistente del parámetro de dispersión. Por ese hecho, aquí usamos la distribución a *priori* de los parámetros del modelo normales, con una varianza lo suficientemente grande y un valor esperado de cero. Esto también podemos verlo como una extensión, heredada desde los modelos lineales, como acabamos de ver anteriormente en esta sección.

2.6.2.4. Estimación en modelos de regresión - MLG

En este caso, la función de verosimilitud para $\boldsymbol{\beta}$ del modelo usual considerado (2.1) y (2.2) es,

$$L(\boldsymbol{\beta}) \propto \sum_{i=1}^n \phi \{y_i \theta_i - b(\theta_i)\} \quad (2.65)$$

en donde asumimos que el parámetro ϕ es fijo y conocido por facilidad de la deducción del procedimiento de estimación. Así, su vector de score es

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^n \phi \left\{ \sqrt{\frac{\omega_i}{V_i}} (y_i - \mu_i) x_{ij} \right\} \quad (2.66)$$

en donde $\omega_i = (d\mu_i/d\eta_i)^2/V_i$, con V_i definido como en la sección (2.1). Así, en forma matricial, la función de score puede escribirse como

$$\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \phi \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}) \quad (2.67)$$

con $\mathbf{W} = \text{diag}(\omega_1, \omega_2, \dots, \omega_n)$, $\mathbf{V} = \text{diag}(V_1, V_2, \dots, V_n)$ y $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$. Su respectiva matriz de información de Fisher es dada por la expresión

$$\mathbf{I}(\boldsymbol{\beta}) = \text{E} \left\{ -\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right\} = \phi \mathbf{X}^T \mathbf{W} \mathbf{X} \quad (2.68)$$

[55]. El estimador de $\boldsymbol{\beta}$ vía máxima verosimilitud, es obtenido por una solución iterativa de las ecuaciones $\mathbf{U}(\boldsymbol{\beta}) = 0$. Si la log verosimilitud es cóncava, entonces el estimador de máxima verosimilitud es única cuando existe al menos un $\hat{\boldsymbol{\beta}}$ en donde $\mathbf{U}(\boldsymbol{\beta})$ alcanza el máximo local o el máximo global. Bajo ciertas condiciones de regularidad, el estimador de máxima verosimilitud de $\boldsymbol{\beta}$ cumple con la siguiente propiedad distribucional

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow_d N(\mathbf{0}, \phi^{-1} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta})) \quad (2.69)$$

siempre que $n \rightarrow \infty$, en donde

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}) = \lim_{n \rightarrow \infty} \frac{\mathbf{I}(\boldsymbol{\beta})}{n}$$

[55]. Para un modelo Bayesiano asociado con la verosimilitud (2.65), necesitamos la especificación de una distribución a priori para $\boldsymbol{\beta}$. Una elección usada comúnmente es $N(\boldsymbol{\beta}_0, \mathbf{V}_0)$ donde $\boldsymbol{\beta}_0$ y \mathbf{V}_0 son conocidos. Así, después de observar $\mathbf{y} = (y_1, \dots, y_n)^T$, la distribución a posteriori de $\boldsymbol{\beta}$ es dada por

$$\pi(\boldsymbol{\beta} | \mathbf{y}) \propto \exp \left[\sum_{i=1}^n \phi \{ y_i s(\mathbf{x}_i^T \boldsymbol{\beta}) - b(s(\mathbf{x}_i^T \boldsymbol{\beta})) \} - \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{V}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right] \quad (2.70)$$

en donde la función que relaciona el parámetro canónico θ con la media μ (inversa de la función $b'(\cdot)$) es denotado por $\theta = q(\mu)$ y $\mu = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$, es decir, $s(\mathbf{x}_i^T \boldsymbol{\beta}) = q(g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}))$ (véase [16]). Como es claro, la distribución a posteriori de la ecuación (2.70) no es analíticamente tratable. De hecho, no existe una expresión cerrada para la constante normalizadora. Además, encontrar medidas a posteriori por integración numérica no es una tarea sencilla. El enfoque más conveniente es entonces la integración numérica MCMC, las cuales requieren, como se ha visto anteriormente, generación de muestras de la distribución a posteriori. Los algoritmos que pueden ser usados, son el Metropolis-Hastings, pero si la log verosimilitud a posteriori es cóncava, entonces se podría también usar el enfoque adaptive rejection sampling [16].

2.6.2.5. Estimación en modelos autocovariantes sobredispersos

En el presente capítulo en la sección (2.5) se hizo uso de la variable condicionada $Y_i^* = Y_i|Y_{(-i)}$ y se mencionó la estructura que toma en cada una de los casos considerados en los modelos autocovariantes sobredispersos. Se denota la función de probabilidad de esa variable condicionada de forma general por $f(y_i^*, \theta_i, \phi_i)$, con θ_i y ϕ_i parámetros asociados a la media y a la dispersión respectivamente. De manera general, se supone que la función de verosimilitud para los parámetros considerados en el modelo es

$$L^*(\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \lambda) = \prod_{i=1}^n f(y_i^*, \theta_i, \phi_i) = \prod_{i=1}^n f^+(y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \lambda) \quad (2.71)$$

cuando las observaciones y_i^* son independientes, para una función $f^+(\cdot)$ desconocida, con $\theta_i = g(\mu_i)$, $\mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \rho A_{1i})$, $\phi_i = h^{-1}(\mathbf{z}_i^T \boldsymbol{\gamma} + \lambda A_{2i})$, $E(Y_i^*) = \mu_i$, $\text{Var}(Y_i^*) = \phi_i V(\mu_i)$ para A_{ji} , $j = 1, 2$ y todo lo demás es como en la sección (2.5). Para establecer un modelo Bayesiano asociado a la verosimilitud (2.71), se necesita la especificación de una distribución a priori para $\boldsymbol{\beta}^* = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \lambda)$. Se supone que esa distribución es $\pi(\boldsymbol{\beta}^*)$ y entonces la distribución a posteriori es dada por

$$\pi(\boldsymbol{\beta}^*|\mathbf{y}) = \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \lambda|\mathbf{y}) \propto L^*(\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \lambda)\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \lambda). \quad (2.72)$$

Como la distribución (2.72) posiblemente no es tratable analíticamente, el enfoque más conveniente es la integración numérica MCMC, la cual requiere generación de muestras de la distribución a posteriori. Los algoritmos que pueden ser usados son el Metropolis-Hastings (Gibbs Sampler), pero si la log verosimilitud a posteriori es cóncava, entonces se podría también usar el enfoque adaptive rejection sampling [22].

2.7. Aplicación: Modelación espacial de la tasa de mortalidad infantil colombiana

Con base en la información explicada al inicio del capítulo que define cada una de las variables, es importante conocer el comportamiento espacial de la variable de interés que define la respuesta en el modelo considerado; el número de niños menores de cinco años que murieron durante el período de 2000 al 2005 (variable NF). El mapa de la variable NF es mostrado en la figura 2.1

Los departamentos con más índice de la variable NF son Chocó y Guajira, Amazonas, Vaupés, Guaviare, Guainía, Vichada, Casanare y Arauca, en general hacia los llanos, gran parte del departamento del Amazonas; cada uno con más de 5165 niños menores de cinco años fallecidos durante el período de 2000 al 2005. De otro lado, los departamentos con el menor número de niños menores de cinco años fallecidos entre el período 2000 a 2005, son Tolima, Quindío, Risaralda, Cundinamarca, Bogotá, Santander, Atlántico y Bolívar.

En un análisis preliminar, se tiene que el valor de la estadística I de moran es $I = 0.592$ y el p valor de la prueba es 7.797×10^{-11} en donde la hipótesis nula es que no hay correlación espacial por parte de esta variable (*rechazamos H_0 en favor de H_1*).

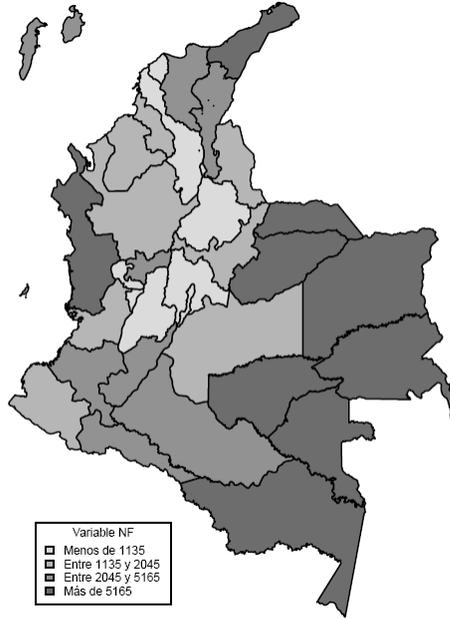


FIGURA 2.1. Mapa de Colombia por departamentos para la variable NF

2.7.1. Modelo Poisson

En esta sección se asume que la variable de interés condicionada a los valores de la variable en el sistema, el número de niños (*en cientos*) menores de 5 años que murieron en el período 2000 a 2005 en la i -ésima área (*cada departamento, además de cuatro ciudades*), sigue una distribución Poisson con parámetro λ_i , cuyo modelo de la media es dado por

$$\log(\lambda) = \beta_0 + \beta_1 \text{NMadres} + \beta_2 \text{Viol} + \beta_3 \text{NBI} + \beta_4 \text{Rec} + \beta_5 \text{ES} + \beta_6 \text{Vac} \quad (2.73)$$

Este es el ya conocido modelo Poisson, cuya función canónica de enlace es el logaritmo natural, que asume que la media y la varianza de la variable respuesta es la misma. En este caso, para la variable de interés NF se tiene que la media es 100.11 y la desviación estándar es 214.05 (*en cientos*), resultados que sugieren empíricamente sobredispersión en la información. En efecto, el supuesto que $E(Y) = \lambda$ y $\text{Var}(Y) = \lambda$, no se cumple, a pesar de que se asume que la variable sigue dicha distribución.

Para este modelo se tiene $-2\log L = 710.11$ y $BIC = 735.20$. En la tabla 2.1 se reportan las estimaciones de los parámetros y sus desviaciones estándar.

	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Est</i>	1.544	4.969×10^{-4}	0.068	0.006	-8.283×10^{-4}	-0.043	0.004
<i>d.e</i>	0.324	2.178×10^{-5}	0.005	0.002	7.883×10^{-5}	0.005	0.002

TABLA 2.1. Estimadores a posteriori, media y desviación estándar del modelo (2.73)

Las distribuciones a priori para cada uno de estos parámetros que modelan la media, es normal de media cero y varianza del orden 10^6 , es decir que para cada parámetro se

considera $\beta_i \sim N(0, 10^{-6})$ como distribución. Se tuvo en cuenta la actualización de la cadena en cada 20 pasos con el fin de eliminar la autocorrelación de la cadena.

2.7.2. Modelo autocovariante Poisson

Al tener en cuenta la estructura de correlación espacial capturada por una nueva variable \mathbf{WY} , construida con la matriz de contigüidad \mathbf{W} definida en el capítulo anterior, este nuevo modelo presenta los valores $-2 \log L = 468.30$ y $BIC = 496.97$. Haciendo uso del BIC se observa que el modelo que incluye la variable \mathbf{WY} tiene un mejor ajuste para los datos en comparación al simple modelo Poisson. El modelo se formula de la siguiente manera:

$$\log(\lambda) = \beta_0 + \beta_1 \text{NMadres} + \beta_2 \text{Viol} + \beta_3 \text{NBI} + \beta_4 \text{Rec} + \beta_5 \text{ES} + \beta_6 \text{Vac} + \rho A \quad (2.74)$$

Las estimaciones de los parámetros y sus respectivas desviaciones estándar, son reportadas en la tabla 2.2

	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Est</i>	-0.845	3.045×10^{-4}	0.082	0.015	-3.992×10^{-4}	-0.035	0.013
<i>d.e</i>	0.361	2.450×10^{-5}	5.078×10^{-3}	1.788×10^{-3}	8.082×10^{-5}	5.509×10^{-3}	0.002

TABLA 2.2. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.74)

En el ajuste del modelo, se encuentra que la estimación del parámetro de la autocorrelación es $\hat{\rho} = 3.574 \times 10^{-3}$ con desviación estándar 2.337×10^{-4} , lo que indica que el parámetro es significativamente distinto de cero y, que la información ponderada de los vecinos está afectando la respuesta en cada área. La distribución a priori para cada uno de los parámetros nuevamente es normal de media cero y varianza del orden de 10^6 , esto es $\beta_i \sim N(0, 10^{-6})$ para $i = 0, 1, \dots, 6$ y $\rho \sim N(0, 10^{-6})$. Con la información disponible en la tabla 2.2 y lo mencionado anteriormente, se concluye que todos los parámetros son significativamente distintos de cero. Así, se observa que el modelo mejora considerablemente cuando se le introduce este término que recoge la información ponderada de los vecinos.

2.7.3. Modelo de sobredispersión binomial negativo

En esta sección se asume para los datos un modelo de sobredispersión binomial negativo, es decir, se ajusta una distribución Poisson en donde su media es modelada como una gamma capturando mayor variabilidad en la respuesta. Desde este enfoque, se consigue que $NF \sim BN(\alpha_i, \tau)$, en donde $\alpha_i = \tau / (\tau + \lambda_i)$. El modelo para la media es expresado de la siguiente manera

$$\log(\lambda) = \beta_0 + \beta_1 \text{NMadres} + \beta_2 \text{Viol} + \beta_3 \text{NBI} + \beta_4 \text{Rec} + \beta_5 \text{ES} + \beta_6 \text{Vac} \quad (2.75)$$

Los valores estimados para cada uno de los parámetros son presentados en la tabla 2.3.

El BIC para este caso es 330.32. Las distribuciones a priori siguen siendo las mismas para los β 's, es decir, $N(0, 10^{-6})$. En cuanto al parámetro τ la distribución considerada es

	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Est</i>	0.897	6.910×10^{-4}	0.062	0.016	-3.329×10^{-4}	-0.033	0.003
<i>d.e.</i>	1.198	1.846×10^{-4}	0.020	0.009	1.649×10^{-4}	0.017	0.009

TABLA 2.3. Media y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.75)

una gamma de ambos parámetros 0.0001, es decir, $\tau \sim G(0.0001, 0.0001)$. La estimación es dada por $\hat{\tau} = 2.942$ y la desviación estándar es 0.835. Las pruebas de normalidad para los parámetros incluyentes, a excepción del parámetro de la varianza, son hechas mediante Kolmogorov - Smirnov y en todos los casos los p valores son menores que 0.05

2.7.4. Modelo binomial negativo autocovariante

En este caso la formulación del modelo binomial negativo autocovariante discutido en la sección (2.5.1) es:

$$\log(\lambda) = \beta_0 + \beta_1 \text{NMadres} + \beta_2 \text{Viol} + \beta_3 \text{NBI} + \beta_4 \text{Rec} + \beta_5 \text{ES} + \beta_6 \text{Vac} + \rho A \quad (2.76)$$

Para este modelo $-2 \log L$ es 294.813 y en consecuencia el BIC es 327.06. Nótese que el valor del BIC, baja un poco con respecto al modelo anterior. La estimación y la desviación estándar para la distribución a posteriori de cada parámetro son reportadas en la tabla 2.4.

	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Est</i>	0.891	2.935×10^{-04}	0.040	0.024	-3.412×10^{-04}	-0.014	3.393×10^{-03}
<i>d.e</i>	0.891	2.270×10^{-04}	0.019	9.550×10^{-03}	1.558×10^{-04}	0.018	0.007

TABLA 2.4. Media y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.76)

La estimación de los parámetros τ y ρ son 3.621 y 3.902×10^{-3} respectivamente, con desviaciones estándar 1.126 y 1.754×10^{-3} . El modelo obtenido luego de haber hecho un proceso de selección de variables con respecto a los resultados mostrados en la tabla 2.4, está dado por la siguiente ecuación

$$\log(\lambda_i) = \beta_0 + \beta_3 \text{NBI}_i + \beta_4 \text{Rec}_i + \rho A_i \quad (2.77)$$

En consecuencia, el nuevo valor de $-2 \log L$ es 300.064 y así, el BIC es 317.98. Como el valor del BIC disminuye un poco con respecto al anterior, se considera este modelo como el mejor, además de reducir el número de variables explicativas. Así, todos los parámetros son significativamente distintos de cero a un 95 %, en particular el ρ llamado el parámetro de la autocovariable **A**. La tabla 2.5 presenta estos resultados.

La prueba de normalidad para cada una de las distribuciones a posteriori de los parámetros, no rechaza únicamente para el parámetro β_0 .

Podemos ver que se ha encontrado un buen modelo, no solo por el valor del BIC que presenta, sino por su composición y su poca cantidad de parámetros (*parsimonia*).

	β_0	β_3	β_4	ρ	τ
<i>Est</i>	1.992	0.031	-4.150×10^{-04}	0.006	3.427
<i>d.e</i>	0.290	0.006	1.417×10^{-04}	8.573×10^{-04}	0.991

TABLA 2.5. Media y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.77)

2.7.5. Modelo Poisson Normal

La siguiente alternativa para trabajar la sobredispersión es el modelo Poisson - normal. La propuesta principal es adicionar en el predictor lineal un efecto aleatorio que sigue una distribución normal de media cero y varianza γ . De esta forma, la varianza explicada por Y que hace de la variable respuesta, se incrementa de acuerdo a la varianza del coeficiente aleatorio γ , además de darle peso a la covarianza entre dos observaciones. La fórmula del modelo es

$$\log(\lambda) = \beta_0 + \beta_1 \text{NMadres} + \beta_2 \text{Viol} + \beta_3 \text{NBI} + \beta_4 \text{Rec} + \beta_5 \text{ES} + \beta_6 \text{Vac} + \varsigma \quad (2.78)$$

en donde $\varsigma_i \sim N(0, \gamma)$. Para γ la distribución a priori es $\gamma \sim G(0.001, 0.001)$. Una vez determinadas las distribuciones a posteriori de cada uno de los parámetros, se encuentra que $-2 \log L = 189.04$ y así el BIC para este caso es 217.70. El valor estimado para γ es 2.44 (*la mediana de la distribución a posteriori*). La tabla 2.6 contiene las estimaciones y la desviación estándar para cada parámetro.

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	γ
<i>Est</i>	0.666	6.700×10^{-04}	0.055	0.019	-3.900×10^{-04}	-0.029	4.881×10^{-03}	2.440
<i>d.e</i>	1.778	-2.000×10^{-04}	0.029	0.012	1.910×10^{-04}	0.019	0.010	0.859

TABLA 2.6. Media y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.78)

Se procede a una depuración del modelo, es decir, a extraer variables explicativas que no influyan de forma certera en el ajuste. De este procedimiento se llega a la forma funcional

$$\log(\lambda) = \beta_0 + \beta_1 \text{NMadres} + \beta_2 \text{Viol} + \beta_4 \text{Rec} + \varsigma \quad (2.79)$$

del cual se encuentra que $-2 \log L = 187.12$ y así $BIC = 205.04$. La tabla 2.7 relaciona las estimaciones de cada parámetro y su desviación estándar.

	β_0	β_1	β_2	β_4	γ
<i>Est</i>	1.776	8.600×10^{-04}	0.042	-4.900×10^{-04}	1.689
<i>d.e</i>	0.490	1.000×10^{-04}	0.013	1.000×10^{-04}	0.492

TABLA 2.7. Media y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.79)

En la siguiente sección se considera el modelo autocovariante Poisson normal que incluye la autocovariable A_i .

2.7.6. Modelo autocovariante Poisson normal

El modelo autocovariante Poisson normal es el modelo de sobredispersión discutido en la sección (2.5.1). La forma funcional del modelo es dada por

$$\log(\lambda) = \beta_0 + \beta_1 \text{NMadres} + \beta_2 \text{Viol} + \beta_3 \text{NBI} + \beta_4 \text{Rec} + \beta_5 \text{ES} + \beta_6 \text{Vac} + \rho A + \varsigma \quad (2.80)$$

y el supuesto distribucional sobre ς_i se mantiene normal, de media cero y varianza γ . Para este, se tiene $-2 \log L = 189.65$, luego $\text{BIC} = 221.90$. La tabla 2.8 muestra las estimaciones y la desviación estándar para los parámetros involucrados en este modelo

	β_0	β_1	β_2	β_3	β_4	β_5	β_6
<i>Est</i>	0.853	2.176×10^{-04}	0.038	0.025	-3.726×10^{-04}	-0.013	2.033×10^{-03}
<i>d.e.</i>	1.032	2.230×10^{-04}	0.022	9.650×10^{-03}	1.620×10^{-04}	0.018	0.080

TABLA 2.8. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.80)

Las estimaciones de los parámetros ρ y γ son 0.004 y 3.192 respectivamente, con desviaciones estándar dadas por 1.818×10^{-03} y 1.076. El modelo más adecuado para este caso es dado por la expresión

$$\log(\lambda) = \beta_0 + \beta_3 \text{NBI} + \beta_4 \text{Rec} + \rho A + \varsigma \quad (2.81)$$

en donde la estimación de γ es 3. Para este ajuste, $-2 \log L = 188.98$ y así, se encuentra que $\text{BIC} = 206.90$.

Se calculan los p valores para la prueba de normalidad de las distribuciones a posteriori de cada uno de los parámetros, obteniendo, 0.00816, 2.353×10^{-11} , 0.4824 y 2.405×10^{-11} para β_0 , β_3 , β_4 y ρ respectivamente.

En seguida se tratan los modelos de sobredispersión generalizados. Únicamente se establece el modelo Poisson normal, ya que el binomial negativo generalizado no presenta un buen comportamiento de acuerdo al criterio de comparación escogido. Cualquiera de los modelos ajustados en este caso, varía de acuerdo al mencionado anteriormente, ya que el nuevo interés es observar el comportamiento de alguna de las variables en el caso del parámetro de dispersión.

2.7.7. Modelo Poisson normal generalizado

En esta sección se modela la varianza con la variable NBI . Se asume entonces que la media de la variable NF denotada por λ_i , sigue el modelo

$$\log(\lambda) = \beta_0 + \beta_1 \text{NMadres} + \beta_2 \text{Viol} + \beta_3 \text{NBI} + \beta_4 \text{Rec} + \beta_5 \text{ES} + \varsigma \quad (2.82)$$

donde $\varsigma_i \sim N(0, \gamma_i)$ y el modelo asociado a γ_i es dado por

$$\log(\gamma) = \theta_0 + \theta_1 \text{NBI} \quad (2.83)$$

La tabla 2.9 muestra tanto las estimaciones como la desviación estándar de cada uno de los parámetros que aparecen en el modelo.

	β_0	β_1	β_2	β_3	β_4	β_5
<i>Est</i>	1.192	6.300×10^{-04}	0.056	0.018	-3.900×10^{-04}	-3.218×10^{-02}
<i>d.e.</i>	0.582	1.000×10^{-04}	0.015	6.136×10^{-03}	1.594×10^{-04}	0.016

TABLA 2.9. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.82) y (2.83)

Las estimaciones de θ_0 y θ_1 son respectivamente dados por 0.750 y 7.462×10^{-03} , con desviaciones estándar 0.907 y 0.022. El valor para este caso de $-2 \log L$ es 189.07 y así, el BIC es 217.74. De la tabla 2.9 se observa que los parámetros que modelan la varianza no son significativos. Con respecto a esto, se observa que toda la relación o capacidad explicativa de la variable *NBI* está siendo capturada por la media o por la varianza. Se procede a la búsqueda de la mejor relación funcional y se llega al modelo

$$\log(\lambda) = \beta_0 + \beta_4 \text{Rec} + \beta_5 \text{ES} + \varsigma \tag{2.84}$$

manteniendo la estructura de la varianza. El valor de $-2 \log L$ es 185.98 y el *BIC* es 203.89. Con respecto al modelo (2.82), el BIC señala que hay un mejor ajuste del (2.84). Las estimaciones de este modelo están en la tabla 2.10

	β_0	β_4	β_5	θ_0	θ_1
<i>Est</i>	4.771	-5.821×10^{-04}	-7.283×10^{-02}	0.522	-0.0606
<i>d.e.</i>	0.510	2.539×10^{-04}	0.022	1.043	0.025

TABLA 2.10. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.84)

A pesar de los resultados obtenidos de los parámetros θ_0 y θ_1 mostrados por la tabla 2.10, y según el comportamiento del BIC, que en este caso baja considerablemente, se concluye que los parámetros no deben salir del modelo. Los *p* valores de las pruebas de normalidad para cada una de las distribuciones a posteriori de los parámetros son 0, 9.407×10^{-10} , 0.07012, 0.01323 y 0, para β_0 , β_4 , β_5 , θ_0 y θ_1 .

Hasta aquí se ha involucrado una variable que modela la varianza, siendo el caso de *NBI*, y cuya forma funcional es (2.83). Otra opción es pensar que el modelamiento de la varianza puede hacerse mediante la variable *ES*, es decir, ahora la varianza toma la forma funcional

$$\log(\gamma) = \theta_0 + \theta_1 \text{ES} \tag{2.85}$$

Ahora, con esta nueva estructura en la varianza, como primera medida se estudia el modelo donde el predictor lineal de la media tiene la siguiente forma

$$\log(\lambda) = \beta_0 + \beta_1 \text{NMadres} + \beta_2 \text{Viol} + \beta_3 \text{NBI} + \beta_4 \text{Rec} + \beta_5 \text{ES} + \beta_6 \text{Vac} + \varsigma \tag{2.86}$$

En cuanto a esta nueva propuesta, el valor de $-2 \log L$ es 188.32 y así, el BIC correspondiente es 220.57.

Al realizar el proceso de depuración de parámetros, la forma funcional de modelo más adecuado tiene la siguiente forma

$$\log(\lambda) = \beta_0 + \beta_4 \text{Rec} + \varsigma \quad (2.87)$$

con la misma estructura de dispersión para el coeficiente aleatorio, dada en la ecuación (2.85). Para este modelo, el valor de $-2 \log L$ es 188.26, y su respectivo *BIC* es de 202.32.

Hasta ahora se ha considerado un modelo que involucre en la varianza la variable *NBI* y la variable *ES*. Ahora se incluye en el modelo de la sobredispersión la covariable A_i al igual que en los modelos SAR y CAR para cuando se trabajan modelos lineales normales, de forma que este coeficiente queda expresado de forma funcional como

$$\log(\gamma) = \theta_0 + \theta_1 A \quad (2.88)$$

con $A_i = \mathbf{W}_i \mathbf{Y}$.

El siguiente predictor lineal es el del modelo que mejor se comporta, en cuanto al criterio BIC,

$$\log(\lambda) = \beta_0 + \beta_1 \text{Rec} + \varsigma \quad (2.89)$$

En este caso el *BIC* es de 200.78 y el valor de $-2 \log L$ es 186.45. Para cada uno de los parámetros, el *p* valor de la hipótesis de normalidad para sus distribuciones a posteriori es menor que 0.05 a pesar de que sus histogramas muestran mucha similitud con la normalidad. La información de cada uno de los parámetros, su estimación, su varianza, están relacionadas en la tabla 2.11.

	β_0	β_1	θ_0	θ_1
<i>Est</i>	3.283	-4.564×10^{-04}	0.522	-6.617×10^{-03}
<i>d.e.</i>	0.271	2.091×10^{-04}	0.352	2.069×10^{-03}

TABLA 2.11. Estimación y desviación estándar de la distribución a posteriori de los parámetros de los modelos (2.88) y (2.89)

Nuevamente, a pesar que θ_0 no es significativo, se prefiere este modelo por su buen comportamiento con respecto al BIC.

Cada uno de los modelos presentados en las secciones anteriores, muestra distintas características y presenta distintos componentes sistemáticos. El comportamiento de cada modelo y su calidad de ajuste, está representado y cuantificado por el valor del BIC. Se consideran varias propuestas mediante la suposición de que la variable respuesta es Poisson en la aplicación de la tasa de mortalidad en los niños menores de cinco años durante el período 2000 - 2005. El nombre de cada una de ellas y su respectivo BIC son mostrados en la tabla 2.12

Claramente, la inclusión de un término que recoge la información de la vecindad, baja considerablemente el BIC como se muestra anteriormente. Puede apreciarse en el comportamiento del criterio en los dos modelos más simples Poisson log - lineal y Poisson log - lineal autocovariante. Sin embargo, puede apreciarse que el modelo más adecuado escogido con el este criterio es el Poisson normal generalizado con estructura

$$\log(\lambda) = \beta_0 + \beta_1 \text{Rec} + \varsigma$$

<i>Modelo</i>	$-2 \log L$	k	BIC
Poisson Log - Lineal	710.118	7	735.20
Poisson Log - Lineal Autocovariante	468.302	8	496.97
Binomial Negativo	301.655	8	330.32
Binomial Negativo Autocovariante	300.064	5	317.98
Poisson Normal	187.126	5	205.04
Poisson Normal Autocovariante	188.987	5	206.90
Poisson Normal Generalizado	186.450	4	200.78

TABLA 2.12. BIC de los modelos con distribución asociada a Poisson

para la media, y

$$\log(\gamma) = \theta_0 + \theta_1 A$$

para la dispersión. De esta manera, el número de niños fallecidos por departamentos en Colombia, es una variable dependiente de los recursos asignados por el estado y su comportamiento variable también puede ser explicado por el comportamiento regional, en parte por la concentración de pobreza en algunos sectores colombianos en específico y por el conflicto armado que la genera, reunido de manera más perceptible en el sector rural, como se muestra en el mapa para la variable NF, figura 2.1. Esta propuesta es considerada para trabajar la sobredispersión y aquí se muestra la opción de no solamente modelar la media como función de un grupo de variables explicativas, sino también la varianza, haciendo uso de que la información disponible presenta cualidades de heterogeneidad espacial y que considerar una varianza igual para todas las unidades resulta en una opción no tan adecuada.

2.8. Aplicación: Modelación espacial del período postnatal de cribado en Colombia

Es importante conocer el comportamiento espacial de la variable de interés que define la respuesta en los próximos modelos considerados; el número de madres (*en miles*) en la i -ésima área que pasaron por un período postnatal de cribado (N_{scree}). El mapa de esta variable es mostrado en la figura 2.2

Los departamentos de la Guajira, Amazonas, Vaupés, Guaviare, Guainía, Vichada, Casanare, Putumayo y la región peninsular son las regiones o departamentos del país que presentan mayor índice de la variable N_{Scree} . Por su parte los departamentos de Cundinamarca, Bogotá, Santander, Norte de Santander, Córdoba, Bolívar, Magdalena y Atlántico, muestran menores índices, ya que el número de madres que pasaron por un período postnatal de cribado es inferior a 21190 para cada uno. En este caso la estadística I de Moran es $I = 0.540$ y el p valor de la prueba cuya hipótesis nula es que no hay correlación espacial por parte de esta variable es de 2.289×10^{-10} . Con esta información se procede a realizar el respectivo modelo para cada uno de los casos considerados.

El objetivo ahora es proponer un modelo que explique la proporción de madres que han dado a luz su último hijo entre 1999 y 2005, y que pasaron por período postnatal de cribado. Se supone entonces que el número de madres (en miles) en la i -ésima área que pasaron este cribado (es decir la variable N_{Scree}) dados los valores de la variable en el sistema considerado, sigue una distribución binomial con probabilidad de éxito dada por

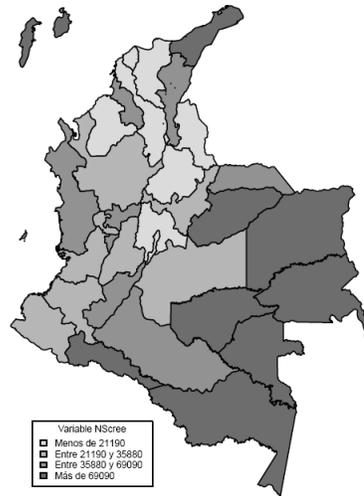


FIGURA 2.2. Mapa de Colombia por departamentos para la variable $NScree$

π_i , y su correspondiente número de ensayos dado por el número de madres (en miles) que han tenido su último hijo después de 1999.

2.8.1. Modelo logístico (Binomial)

En esta sección, se propone como punto de partida el modelo de regresión logístico dado por

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 \text{Viol} + \beta_2 \text{NBI} + \beta_3 \text{ES} + \beta_4 \text{Pay} \quad (2.90)$$

con π_i la probabilidad de éxito de la variable $NScree$ (la probabilidad que una madre en la i -ésima área pase por un período post natal de cribado). Los valores correspondientes a $-2 \log L$ y al BIC son 326.85 y 344.77 respectivamente. La información de las estimaciones de cada uno de los parámetros está relacionada en la tabla 2.13

	β_0	β_1	β_2	β_3	β_4
<i>Est</i>	1.628	-0.026	-4.166×10^{-03}	0.013	-0.045
<i>d.e.</i>	0.244	5.841×10^{-03}	2.524×10^{-03}	5.763×10^{-02}	8.014×10^{-03}

TABLA 2.13. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.90)

Los p valores para las pruebas de normalidad de Kolmogorov - Smirnov son 0.055, 0.043, 0.005, 0.278 y 0.246 para los parámetros β_0 , β_1 , β_2 , β_3 y β_4 respectivamente, a pesar de mostrar similitud con la distribución normal. Cabe recordar que se está frente a un caso de sobredispersión, y se debe seguir en la búsqueda del modelo más adecuado. Se trabaja ahora con el modelo autocovariante logístico definido en la sección (2.5.2). Para esto necesitamos un coeficiente que recoja la información de los vecinos de la i -ésima área y lo hacemos mediante la construcción una nueva variable denominada A_i , resultado de la expresión:

$$A_i = \frac{\hat{\pi}_i}{1 - \hat{\pi}_i} \tag{2.91}$$

en donde $\hat{\pi}_i$ tiene la forma del estimador de máxima verosimilitud del parámetro π_i . En general, se define el estimador $\hat{\pi}_i = \frac{Y_i^*}{n_i^*}$ con Y_i^* definido por $\mathbf{W}_i Y = Y_i^*$ y n_i^* definido por $\mathbf{W}_i n = n_i^*$ en donde \mathbf{W}_i es la ya mencionada fila i - ésima de la matriz de pesos espaciales \mathbf{W} de contigüidad 1, $Y_i = N\text{Scree}_i$ y $n_i = N\text{Madres99}_i$ (n es el vector de la variable $N\text{Madres99}$), características que definen la distribución binomial de la variable respuesta.

2.8.2. Modelo autocovariante logístico (Binomial)

El propósito al igual que en el caso de la Poisson, es incluir en el predictor lineal del modelo, un coeficiente que resuma y aporte información relacionada con los vecinos más cercanos, en cuanto a información sobre la variable $N\text{Scree}$. Se trata de observar el comportamiento del modelo cuando se involucra una variable que ayude a controlar la variabilidad extra causada por los datos. El modelo propuesto en la sección (2.5.2) toma la forma de la siguiente ecuación

$$\log\left(\frac{\pi}{1 - \pi}\right) = \beta_0 + \beta_1 \text{Viol} + \beta_2 \text{NBI} + \beta_3 \text{ES} + \beta_4 \text{Pay} + \rho A \tag{2.92}$$

en donde la variable A_i es definida como en (2.91). En este nuevo caso, se tiene $-2 \log L = 313.5$ y además $BIC = 335.06$, de donde se observa que hay un mejor ajuste en este modelo que en caso del modelo logístico simple, pues, hay una reducción importante en el valor del BIC. Se muestra la tabla 2.14 que relaciona los valores de las estimaciones de los parámetros y sus respectivas desviaciones estándar.

	β_0	β_1	β_2	β_3	β_4	ρ
<i>Est</i>	1.978	-1.661×10^{-02}	-4.070×10^{-03}	1.088×10^{-02}	-0.045	-1.197
<i>d.e.</i>	0.251	5.900×10^{-03}	2.300×10^{-03}	5.500×10^{-03}	7.900×10^{-03}	0.493

TABLA 2.14. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.92)

Se realizan las pruebas de normalidad de todos los parámetros. Para los parámetros en su orden, $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ y ρ , los p valores son 0.00004, 0.016, 0.8339, 0.8992, 0.4224 y 0.2991.

2.8.3. Modelo Beta - Binomial

La información respectiva a la construcción de este modelo, puede ser encontrada en la sección (2.2.1.2). El objetivo de modelar los datos con esta propuesta, es atacar la sobredispersión presente. La idea principal, es fijar un modelo en donde la variable respuesta tenga distribución beta binomial. Esto es producto de asumir que la variable respuesta tiene distribución binomial de parámetros π_i^* y m_i , y que los π_i^* siguen una distribución beta con primer momento π_i y segundo momento $\tau \pi_i(1 - \pi_i) + \pi_i^2$, de modo que la varianza está dada por $\tau \pi_i(1 - \pi_i)$. De esta manera, la variable respuesta Y_i sigue una distribución beta binomial con esperanza y varianza dadas por $m_i \pi_i$ y $m_i \pi_i(1 - \pi_i)[1 + \tau(m_i - 1)]$. Para este caso el modelo propuesto es dado por la siguiente expresión

$$\log\left(\frac{\lambda}{1-\lambda}\right) = \beta_0 + \beta_1 \text{Viol} + \beta_2 \text{NBI} + \beta_3 \text{ES} + \beta_4 \text{Pay} \quad (2.93)$$

en donde $N\text{Scree} \sim \text{Bin}(N\text{Madres99}_i, \pi_i)$, $\pi_i \sim \text{Beta}(a_i, b_i)$ y λ_i es la media de la distribución beta, es decir, $\lambda_i = \frac{a_i}{a_i+b_i}$ y $\tau = a_i + b_i$. En principio, $\log(\tau_i) = \theta$, constante para todas las observaciones. La tabla 2.15 muestra las estimaciones y las desviaciones estándar de las estimaciones del modelo.

	β_0	β_1	β_2	β_3	β_4	θ
<i>Est</i>	2.468	-0.041	-8.962×10^{-03}	4.015×10^{-03}	-0.051	3.781
<i>d.e.</i>	0.485	0.014	0.007	0.012	0.017	0.425

TABLA 2.15. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.93)

En este caso, y a menos que se especifique lo contrario, cabe recordar que las distribuciones a priori son dadas por normales independientes de la forma

$$\beta_j \sim N(0, 1.0 \times 10^{-6})$$

dando como resultado un fuerte proceso de convergencia, por el tamaño que se le da a la varianza. Aquí por la forma de construcción del modelo, la distribución a priori para el parámetro θ es $N(0, 1.0 \times 10^{-6})$, igual que para un coeficiente del predictor lineal. La estimación de la expresión $\log \tau_i$ es 3.78 con una desviación estándar de 0.4254, lo que sugiere que el parámetro es significativamente distinto de cero. Los p valores de las pruebas de normalidad para cada parámetro, son todos dados por valores menores que 9.911×10^{-5} , a excepción del de β_4 , cuyo valor p es de 0.2410. Para este modelo, se tiene que $-2 \log L = 182.71$ y $\text{BIC} = 204.21$, mostrando una reducción importante en el valor del BIC, lo que de inmediato sugiere una preferencia sobre este.

Después de realizar el procedimiento de examinar las variables, se llega a un modelo con forma funcional

$$\log\left(\frac{\lambda}{1-\lambda}\right) = \beta_0 + \beta_2 \text{NBI} + \beta_4 \text{Pay} \quad (2.94)$$

cuyo valor de $-2 \log L$ es 183.13 y el BIC es 197.46, mostrando así nuevamente, una mejoría con respecto al anterior. La tabla 2.16 muestra las estimaciones y las desviaciones estándar de las estimaciones cada uno de los parámetros.

	β_0	β_2	β_4	θ
<i>Est</i>	1.089	-0.014	-0.022	3.602
<i>d.e.</i>	0.186	4.278×10^{-03}	0.011	0.404

TABLA 2.16. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.94)

Los p valores de la prueba de normalidad para cada uno de los parámetros, obteniendo 0.2122, 8.778×10^{-11} , 6.387×10^{-6} y 0.0001625 para β_0 , β_2 , β_4 y θ respectivamente.

2.8.4. Modelo autocovariante Beta - Binomial

Este se aborda de manera que la ecuación resultante que formula el mejor modelo que incluye la variable rezagada, está dada por la expresión:

$$\log\left(\frac{\lambda}{1-\lambda}\right) = \beta_0 + \beta_2\text{NBI} + \rho A \tag{2.95}$$

cuyas estimaciones de los parámetros y de las varianzas de las estimaciones son dadas en la tabla 2.17.

	β_0	β_2	ρ	θ
<i>Est</i>	1.629	-0.016	-1.689	3.457
<i>d.e.</i>	0.364	4.363×10^{-03}	0.911	0.358

TABLA 2.17. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.95)

Cabe mencionar que el valor de $-2\log L$ para este nuevo modelo es de 180.19 y asimismo el valor del BIC baja considerablemente a 194.52, en comparación con el modelo descrito en la sección anterior. Los p valores de la prueba de normalidad en este caso son menores o iguales a 2.373×10^{-9} para β_0 , β_2 y ρ , y de 0.02468 para el parámetro θ que define al τ que aumenta la varianza recogida por el modelo. Hasta el momento se han considerado los modelos logístico y beta binomial, cada uno involucrando una variable de rezago que permita recoger información de los vecinos más cercanos. Ahora, es el turno de los modelos con efectos aleatorios.

2.8.5. Modelo Binomial Normal

La característica principal del modelo binomial normal, está en adicionar al predictor lineal un efecto aleatorio, que permita aumentar la variabilidad que explica el modelo. Notado por ς_i al coeficiente aleatorio para la i -ésima unidad espacial, se asume que es distribuido normalmente con media cero y varianza γ y que son independientes con ς_j , para todo $i \neq j$. De esta manera, la forma funcional del modelo queda dada por la siguiente expresión:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1\text{Viol} + \beta_2\text{NBI} + \beta_3\text{ES} + \beta_4\text{Pay} + \varsigma \tag{2.96}$$

con $\varsigma_i \sim N(0, \gamma)$. Aquí se asume una distribución a priori para el γ como $G(0.0001, 0.0001)$, la cual solo admite valores positivos y además presenta una varianza suficientemente grande. Los coeficientes del modelo se muestran en la tabla 2.18

	β_0	β_1	β_2	β_3	β_4	γ
<i>Est</i>	2.546	-0.044	-7.907×10^{-03}	5.508×10^{-03}	-0.055	9.340
<i>d.e.</i>	0.751	0.020	8.242×10^{-03}	0.013	0.019	4.952

TABLA 2.18. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.96)

En este caso, $-2\log L = 180.802$ y $\text{BIC} = 202.30$. Sin embargo se realiza el procedimiento de búsqueda del mejor ajuste, y se llega al modelo que tiene forma funcional dada por

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_4\text{Pay} + \varsigma \tag{2.97}$$

Este modelo es el que mejor se comporta de todos los evaluados, con respecto a la categoría en estudio Binomial Normal. Su valor de $-2 \log L$ es 181.545 y Su BIC es de 192.9. Se muestra en la tabla 2.19 los valores de los parámetros estimados y sus respectivas desviaciones estándar.

	β_0	β_4	γ
<i>Est</i>	0.600	-0.030	5.760
<i>d.e.</i>	0.119	0.013	2.464

TABLA 2.19. Estimación y desviación estándar de la distribución a posteriori de los parámetros (2.97)

Los p valores de las pruebas de normalidad para cada parámetro son 0.1709 y 0.5362.

Como la propuesta es trabajar con modelos autocovariantes, se presenta el resultado de modelar con la variable rezagada A_i .

2.8.6. Modelo autocovariante Binomial Normal

Las especificaciones del modelo se encuentran en la sección (2.5.2). Al realizar el procedimiento detenidamente de selección de variables, se encuentra un modelo que tiene por particularidad contar solo en el predictor lineal con A_i . De esta manera, el modelo resultante está dado por la siguiente expresión:

$$\log \left(\frac{\pi}{1 - \pi} \right) = \beta_0 + \rho A + \varsigma \tag{2.98}$$

en donde $\varsigma_i \sim iidN(0, \gamma)$, al igual que el caso anterior. Las distribuciones a priori son las mismas con las que se trabajó el modelo binomial normal. De esta forma, se muestra la tabla 2.20 que relaciona el estimador de cada parámetro y su correspondiente desviación estándar.

	β_0	ρ	γ
<i>Est</i>	1.042	-1.868	4.816
<i>d.e.</i>	0.381	1.066	1.847

TABLA 2.20. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.98)

Puede notarse de la tabla 2.20 que el parámetro ρ resulta siendo no significativo. Sin embargo, al extraer la variable rezagada A_i del predictor lineal, el BIC aumenta. En este caso, el valor de $-2 \log L$ es 179.14 y el valor del BIC es 189.89. Los p valores de las pruebas de normalidad para cada distribución a posteriori, son más pequeños que 1.086×10^{-9} .

2.8.7. Modelo beta binomial generalizado

La forma funcional para la media del modelo está dada por la siguiente expresión:

$$\log \left(\frac{\lambda}{1 - \lambda} \right) = \beta_0 + \beta_1 \text{Viol} + \beta_2 \text{NBI} + \beta_3 \text{ES} + \beta_4 \text{Pay} \tag{2.99}$$

y para la varianza se tiene

$$\log(\gamma) = \theta_0 + \theta_1 \text{NBI} \quad (2.100)$$

en donde se asume que la variabilidad no constante de la variable respuesta a través de los departamentos puede ser explicada por el comportamiento de la variable NBI . El valor de $-2 \log L$ en este caso es de 184.593 y el BIC es 209.67. Haciendo el proceso de depuración en las variables, el modelo queda reducido en la media por

$$\log\left(\frac{\lambda}{1-\lambda}\right) = \beta_0 + \beta_4 \text{Pay} \quad (2.101)$$

manteniéndose la estructura funcional del parámetro de la dispersión γ_i dada en (2.100). Para este caso, $-2 \log L = 180.844$ y $\text{BIC} = 194.4$, lo que muestra una mejoría con respecto al modelo inmediatamente anterior. Las estimaciones de los parámetros y sus respectivas desviaciones estándar son mostradas en la tabla 2.21

	β_0	β_4	θ_0	θ_1
<i>Est</i>	0.526	-0.032	2.017	0.032
<i>d.e.</i>	0.114	0.011	1.240	0.031

TABLA 2.21. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.100) y (2.101)

Aunque la tabla 2.21 muestra que los parámetros que modelan la dispersión no son significativos, se aprecia que ayudan a que el modelo presente algo de mejoría con respecto al mostrado anteriormente. Sin embargo, al observar esta característica, se piensa en incluir en el modelamiento de la varianza otra variable explicativa, es decir, ahora se asume que la variabilidad no constante de la variable respuesta a través de los departamentos puede ser explicada por el comportamiento de la variable ES . De esta forma, el modelo más adecuado que se encontró en el caso de tener una forma funcional para la varianza de la siguiente manera

$$\log(\gamma) = \theta_0 + \theta_1 \text{ES} \quad (2.102)$$

es el modelo mostrado en (2.101), pero no presenta grandes cambios con respecto al BIC , que sigue siendo de 194.4.

Se presenta la tabla 2.22 que resume el nombre de cada modelo para el caso de la aplicación de la proporción de madres que pasaron por un período postnatal de cribado y su respectivo valor del BIC (*modelos asociados a la distribución binomial*).

<i>Modelo</i>	$-2 \log L$	k	BIC
Logístico Binomial	326.857	5	344.77
Logístico Binomial Autocovariante	313.561	6	335.06
Beta Binomial	183.130	4	197.46
Beta Binomial Autocovariante	180.190	4	194.52
Binomial Normal	181.545	3	192.9
Binomial Normal Autocovariante	179.145	3	189.89
Beta Binomial Generalizado	180.265	4	194.4
Binomial Normal Generalizado	179.338	4	193.32

TABLA 2.22. BIC modelos distribución Binomial

El primer modelo considerado, el logístico, muestra un BIC de 344.77. Es importante señalar, como en este caso y en los demás, la baja en este valor en cuanto se agrega al pre-

dictor lineal la autocovariable A_i definida anteriormente. Cabe resaltar el comportamiento del modelo Beta Binomial generalizado en comparación con el usual considerado desde el enfoque de la sobredispersión Beta Binomial; muestra un BIC un poco menos alto. El modelo más plausible, con mejor BIC , es el Binomial Normal Autocovariante, definido por la expresión

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \rho A + \varsigma$$

en donde $\varsigma_i \sim N(0, \gamma)$. En este caso, la información que define la probabilidad de encontrar una mujer que ha dado a luz su último hijo entre 1999 y 2005, y que pasó por período postnatal de cribado, depende de los valores que tome la región. El comportamiento de esta variable puede definirse por las condiciones de los vecinos circundantes o más cercanos. Mediante las técnicas econométricas usadas, puede encontrarse un modelo parsimonioso, capaz de explicar el comportamiento de esta variable, únicamente teniendo en cuenta los valores de la misma en los lugares más cercanos, haciendo uso de uno de los modelos más usados para abordar la sobredispersión, en este caso provocada por la correlación y por la heterogeneidad espacial.

2.9. Aplicación: Otras matrices de pesos espaciales

Esta sección se ocupa de observar el comportamiento de algunos de los modelos más relevantes considerados, con la diferencia de que la matriz de contigüidad se construye de otra manera a la de primer orden. El procedimiento es igual que en la primera aplicación, se usa entonces la matriz de contigüidad física de segundo orden y la construcción de vecindad mediante las regiones naturales.

2.9.1. Matriz de contigüidad física de segundo orden

El primer modelo probabilístico considerado es el Poisson.

2.9.1.1. Modelos con respuesta Poisson

Con esta nueva estructura espacial, el I de Moran toma un valor de 0.2717 y haciendo uso de las pruebas de aleatorización se encuentra un p valor de 4×10^{-05} que rechaza la hipótesis nula de que la autocorrelación espacial es cero. Bajo esta nueva estructura, la correlación espacial es mucho más baja que si se considera la matriz de contigüidad física de primer orden. Se muestra el diagrama de Moran univariado, en donde se dibuja la variable de interés y la misma variable rezagada en simultáneo. la figura 2.3 muestra que existe una relación no lineal entre la variable rezagada y la variable misma. Esto sugiere que a pesar de que I es pequeña, se puede decir que la correlación es significativa, además de observar el p valor de la prueba, teniendo en cuenta que la función de enlace es la logarítmica. La tabla que resume el BIC de todos los modelos que consideramos se muestra en el cuadro 2.23

El cuadro 2.23 contiene los BIC para los modelos autocovariantes, con la matriz de pesos espaciales definida de segundo orden. Se aprecia claramente que en esta aplicación, esta nueva covariable definida, afecta de manera negativa el modelo en comparación a la definición de contigüidad básica binaria. En efecto, cuando el modelo bajo consideración es

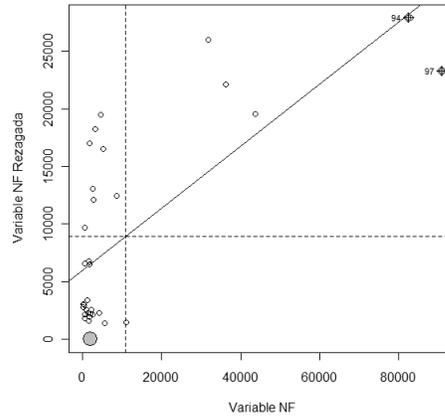


FIGURA 2.3. Diagrama de Moran Univariado para la variable NF

<i>Modelo</i>	$-2 \log L$	k	BIC
Poisson Log - Lineal Autocovariante	680.288	8	708.95
Binomial Negativo Autocovariante	315.246	5	333.16
Poisson Normal Autocovariante	186.26	5	207.76
Poisson Normal Generalizado	185.83	4	200.16

TABLA 2.23. BIC modelos distribución Poisson - Matriz de segundo orden

sobredisperso, la diferencia de BIC puede resultar despreciable, asimismo cuando el modelo considerado es el Poisson log - lineal, ya que el cambio que sufre el BIC es de 735.20 a 708.95. Esto sugiere que la matriz de contigüidad física de primer orden es mucho más adecuada en este caso que la matriz de contigüidad física de segundo orden. La siguiente tabla relaciona estas conclusiones para el caso de la aplicación binomial.

2.9.1.2. Modelos con respuesta binomial

En esta sección, al igual que la inmediatamente anterior, se muestran los resultados cuando se considera una matriz de pesos espaciales de segundo orden. En tal caso el I de moran para la variable

$$OR = \frac{\hat{\pi}}{1 - \hat{\pi}}$$

(con $\hat{\pi}$ el estimador de máxima verosimilitud $NScree/NMadres99$) toma un valor de -0.0023 y el p valor de la prueba de aleatorización es 0.2589 , lo que indica que estadísticamente, ese coeficiente es significativamente igual a cero. También se hace la prueba de aleatorización para la variable $NScree$, que muestra un I de moran de 0.2654 y un p valor de la prueba de aleatorización de 0.00004 , indicando así que se rechaza la hipótesis nula de que la autocorrelación espacial es igual que cero. La figura 2.4 muestra los diagramas de moran para ambos casos.

La figura anterior sugiere de manera similar que el caso anterior, que la relación entre la variable de interés y su comportamiento con respecto a la covariable construida con la matriz de contigüidad de segundo orden, es no lineal. Los respectivos BIC para cada modelo en consideración son mostrados en la tabla 2.24

El cuadro 2.24 muestra que cuando se trata de un modelo con sobredispersión, es indistinta la definición de matriz de contigüidad. En tal caso, según el BIC, no hay diferencias

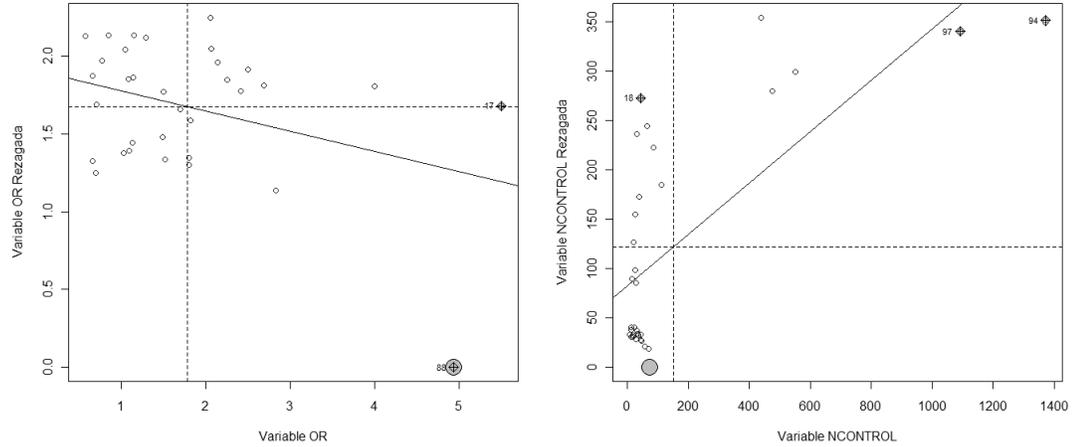


FIGURA 2.4. Diagrama de Moran Univariado para la variable OR y la variable NScree

<i>Modelo</i>	$-2 \log L$	k	BIC
Logístico Binomial Autocovariante	281.612	6	303.11
Beta Binomial Autocovariante	180.336	4	194.67
Binomial Normal Autocovariante	179.226	3	189.97
Beta Binomial Generalizado	183.396	4	197.73

TABLA 2.24. BIC modelos distribución Binomial - Matriz de segundo orden

importantes entre un modelo sobredisperso autocovariante y su matriz que define la vecindad. Sin embargo, es importante señalar que si el modelo considerado no es sobredisperso, el BIC baja considerablemente, mostrando así que hay un mejor ajuste cuando se trabaja con una vecindad de segundo orden que de primer. El BIC baja de 344.77 a 303.11, lo que valida nuestra conclusión.

2.9.2. Matriz de contigüidad física por regiones naturales

Como antes, el primer modelo probabilístico considerado es el Poisson.

2.9.2.1. Modelos con respuesta Poisson

Considerando la estructura espacial de regiones naturales, el I de Moran toma un valor de 0.3943 y haciendo uso de las pruebas de aleatorización se presenta un p valor de 0.00245, rechazando la hipótesis nula de que la autocorrelación espacial es cero en favor de que es positiva. Se muestra el diagrama de Moran univariado, en donde se dibuja la variable de interés y la misma variable rezagada en simultáneo, de acuerdo a esta definición de vecindad. la figura 2.5 muestra que no hay una relación lineal o al menos tan clara entre la variable rezagada y la variable misma. Si se tiene en cuenta que la función de enlace es la logarítmica, se podría llegar a pensar que esta relación podría mejorar considerablemente. La tabla que resume el BIC de todos los modelos que consideramos se muestra en el cuadro 2.25

El cuadro 2.25 contiene los BIC para los modelos autocovariantes, con la matriz de pesos espaciales definida por regiones naturales. Cuando el modelo bajo consideración es sobredisperso, la diferencia de BIC puede resultar despreciable como ocurrió en los ca-

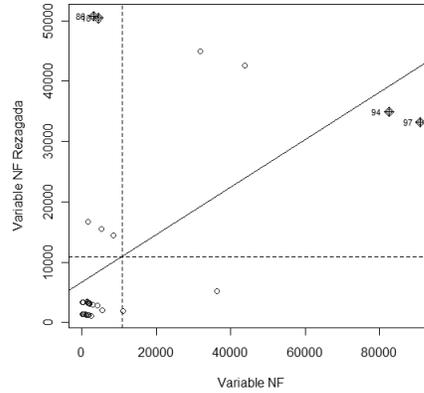


FIGURA 2.5. Diagrama de Moran Univariado para la variable NF - Matriz de contigüidad por regiones naturales

<i>Modelo</i>	$-2 \log L$	k	BIC
Poisson Log - Lineal Autocovariante	572.386	8	601.05
Binomial Negativo Autocovariante	322.870	5	340.78
Poisson Normal Autocovariante	185.55	5	203.46
Poisson Normal Generalizado	185.28	4	199.62

TABLA 2.25. BIC modelos distribución Poisson - Matriz de contigüidad por regiones naturales

Los modelos anteriores y cuando el modelo considerado es el Poisson log - lineal, aunque mejora con respecto a la vecindad de segundo orden, no es más adecuado que el modelo considerando la matriz de contigüidad física de primer orden. Esto sugiere que la matriz de contigüidad física de primer orden sigue siendo más adecuada. La siguiente tabla relaciona estas conclusiones para el caso de la binomial.

2.9.2.2. Modelos con respuesta binomial

Aquí se muestran los resultados cuando se considera una matriz de pesos espaciales mediante la definición de vecindad por regiones naturales. En tal caso el I de Moran para la variable OR es de -0.0105 y el p valor de la prueba de aleatorización es 0.3431 , lo que indica que estadísticamente, ese coeficiente es significativamente igual a cero. De forma similar que cuando lo hicimos para la matriz de contigüidad física de segundo orden, también hacemos la prueba de aleatorización para la variable N_{Scree} , que nos muestra un I de Moran de 0.3591 y un p valor de 0.00434 , indicando así que se rechaza la hipótesis nula de que la autocorrelación espacial es igual a cero. La figura 2.6 muestra los diagramas de Moran para ambos casos, igual que el caso anterior.

La figura 2.6 sugiere de manera similar que el caso inmediatamente anterior, que la relación entre la variable de interés y su comportamiento con respecto a la covariable construida con la matriz de contigüidad por regiones naturales, no es tan lineal o no es tan clara. Los respectivos BIC para cada modelo en consideración son mostrados en la tabla 2.26

Si el modelo considerado no es sobredisperso, el BIC muestra que la definición de vecindad más adecuada es la de segundo orden. Si el modelo es considerado sobredisperso,

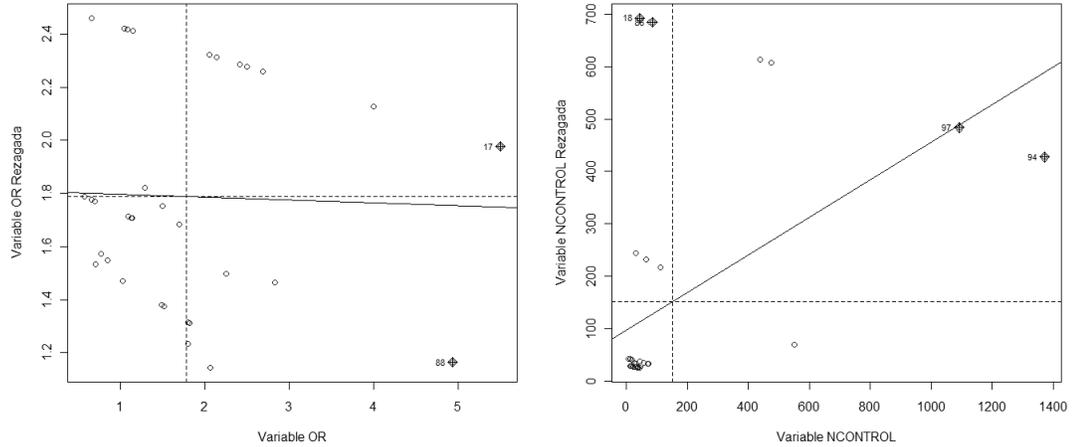


FIGURA 2.6. Diagrama de Moran Univariado para la variable OR y la variable NScree - Matriz de contigüidad por regiones naturales

<i>Modelo</i>	$-2 \log L$	k	<i>BIC</i>
Logístico Binomial Autocovariante	320.83	6	342.33
Beta Binomial Autocovariante	179.487	4	193.82
Binomial Normal Autocovariante	178.484	3	189.23
Beta Binomial Generalizado	184.577	4	198.91

TABLA 2.26. BIC modelos distribución Binomial - Matriz de contigüidad por regiones naturales

entonces, las conclusiones y los resultados no varían mucho, indistintamente de la matriz de contigüidad y la vecindad definida.

2.10. Aplicación: Volumen de atención hospitalaria en Colombia

En el capítulo 1 se usaron las técnicas econométricas espaciales para modelar el volumen de atención hospitalaria en Colombia. Se hizo uso de la variable que mide la tasa de eventos de hospitalización, asistencia por urgencias y asistencia por consulta externa. Es de tener en cuenta que estos datos están relacionados con todos los tipos de entidades (IPS); privada, mixta o pública. Aquí se usan los modelos propuestos y se modela la variable VOLAT definida previamente en el capítulo 1.

Se supone que la variable VOLAT en la i -ésima región condicionada a los demás valores en el sistema espacial, sigue una distribución Poisson de media λ_i . El predictor lineal del modelo es dado por la expresión

$$\log(\lambda) = \beta_0 + \beta_1 \text{Displc} + \beta_2 \text{IndIPS} + \beta_3 \text{Vac} + \beta_4 \text{NBI} + \beta_5 \text{Rec} + \beta_6 \text{RNtr} \quad (2.103)$$

En esta sección solamente se considera la matriz de contigüidad espacial de orden uno. Los valores del BIC son mostrados en la tabla 2.27

La tabla muestra que el modelo que mejor se comporta de acuerdo a esta vecindad establecida es el modelo Poisson normal. Sin embargo se destaca que la diferencia entre

<i>Modelo</i>	$-2 \log L$	k	BIC
Poisson Log - Lineal	384.437	7	408.912
Poisson Log - Lineal Autocovariante	342.791	8	370.763
Binomial Negativo	240.066	4	254.052
Binomial Negativo Autocovariante	231.537	5	249.019
Poisson Normal	157.706	4	171.692
Poisson Normal Autocovariante	157.214	5	174.696
Poisson Normal Generalizado	157.824	5	175.306

TABLA 2.27. BIC de los modelos para la variable VOLAT

el modelo Poisson y Poisson autocovariante es importante, señalando que la inclusión en el modelo de la covariable que recoge la información de los vecinos más cercanos mejora el ajuste. Lo mismo ocurre con el modelo binomial negativo. La diferencia del BIC en el caso de los modelos Poisson normal y Poisson normal autocovariante es de casi tres puntos, una diferencia posiblemente despreciable; sin embargo, se destaca el hecho de que el coeficiente que define la correlación espacial es significativamente distinto de cero. Por esta razón decimos que el modelo tiene la ventaja de capturar la correlación espacial en términos de un factor fijo, y corregir por medio del coeficiente aleatorio la sobredispersión que ρ no es capaz de capturar.

Por esta razón se presenta el modelo Poisson normal autocovariante como el modelo de mejores características. Cabe destacar que en todos los modelos sobredispersos, los parámetros β_1 , β_2 , β_3 y β_6 no son significativos. El modelo Poisson normal tiene entonces la siguiente estructura

$$\log(\lambda) = \beta_0 + \beta_4 \text{NBI} + \beta_5 \text{Rec} + \rho A + \varsigma \quad (2.104)$$

con $\varsigma \sim N(0, \gamma)$. Las estimaciones y desviaciones estándar de las distribuciones a posteriori de los parámetros son reportadas en la tabla 2.28

	β_0	β_4	β_5	ρ	γ
<i>Est</i>	2.791	-0.042	0.131	0.017	4.379
<i>d.e.</i>	0.335	0.006	0.019	0.006	1.710

TABLA 2.28. Estimación y desviación estándar de la distribución a posteriori de los parámetros del modelo (2.104)

De acuerdo con esto, en los departamentos en donde la variable NBI es más alta, se tiene una variable VOLAT más baja y en los departamentos en donde la variable Rec es más alta, la variable VOLAT es más alta. Esto es, en los departamentos en donde el porcentaje de la población con necesidades básicas insatisfechas es más alto, se espera un número de eventos relacionados con el volumen de atención hospitalaria bajo y asimismo, en los departamentos en donde los recursos otorgados para la educación y atención integral para niños pequeños son más altos, se espera un número de eventos relacionados con el volumen de atención hospitalaria más alto.

Se observa que los parámetros son significativos a un 99 %, en particular el parámetro ρ . Según el valor estimado, se espera que entre más alto sea el valor de la variable de los departamentos vecinos, más alto es el valor de la variable respuesta, es decir, entre más eventos relacionados con el volumen de atención hospitalaria se presenten en los vecinos más próximos, más eventos se presentan en la unidad espacial de interés.

Los modelos generalizados no mostraron un ajuste adecuado, razón por la cual no se mencionaron; sin embargo, al igual que en las aplicaciones anteriores, podemos concluir que el criterio de vecindad usado puede cambiar el comportamiento del modelo; nosotros para esta aplicación, solamente consideramos la matriz de contigüidad física de primer orden.

2.11. Conclusiones

Como se ha comentado a través de las aplicaciones, esta es una clase de modelos que permite capturar la correlación espacial de acuerdo con una estructura de vecindad establecida. En muchas formas, la presencia de esta estructura en los datos genera sobre-dispersión. Aunque los modelos de sobredispersión capturan y controlan este fenómeno, incluyendo la autocorrelación entre las unidades, el interés va más allá de controlarla; es en un sentido, cuantificarla y darle característica de factor fijo. En los casos mostrados en el presente capítulo, la inclusión al modelo de una variable que recoja información espacial cercana, permite mejoras en el comportamiento de cada modelo. Sin embargo, debe tenerse en cuenta que por el carácter de la información, la estructura de vecindad definida en la ecuación (1.1) no puede ser conocida por la naturaleza de los datos, sino debe ser demarcada por los objetivos establecidos por el investigador. De acuerdo con esto, el efecto producido por la autocorrelación espacial puede ser totalmente controlado por la autocovariable, siempre y cuando la estructura (1.1) quede totalmente determinada por la realización estructural y probabilística del sistema S ; algo muy complejo de determinar en la práctica.

De otra mano, la variable Y usada en la definición de los modelos sobredispersos en todos los sentidos, puede ser considerada condicionada en dos vías. La primera, antes de establecer el modelo, considerando solamente la estructura espacial S y el proceso $Y(\cdot)$ medido en cada unidad espacial que compone S . En este aspecto, claramente juega un papel muy destacado la definición de vecindad (1.1), obteniendo por notación $Y_i^* = Y_i | Y_{(-i)}$. Después, cuando se considera la construcción de un modelo, el proceso Y^* puede ser condicionado por diferentes vías, como es el caso de los modelos sobredispersos usuales cuando se condiciona por el valor de un parámetro o por el conjunto de efectos aleatorios, como el caso más usual.

APÉNDICE A

Algunos resultados en inferencia para las estadísticas que definen correlación espacial

Cliff y Ord [14] encuentran los siguientes momentos para las dos estadísticas (1.8) y (1.9), como apoyo para la inferencia sobre ρ .

$$E(I) = -1/(n-1) \quad (\text{A.1})$$

$$E(I^2) = (n^2 S_1 - n S_2 + 3\Omega^2)/\{(n^2 - 1)\Omega^2\} \quad (\text{A.2})$$

$$E(C) = 1 \quad (\text{A.3})$$

$$\text{Var}(C) = [(2S_1 + S_2)(n-1) - 4\Omega^2]/\{2(n+1)\Omega^2\} \quad (\text{A.4})$$

donde $\Omega = \sum w_{ij}$

$$S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2 \quad \text{para } i \neq j \quad (\text{A.5})$$

y

$$S_2 = \frac{1}{2} \sum_{i=1}^n (w_{i.} + w_{.i})^2 \quad (\text{A.6})$$

Getis y Ord [30] proponen una importante familia de coeficientes G capaz de medir la asociación espacial usando estadísticas de distancia. En ese escrito se puede encontrar una interesante comparación de la estadística I de Moran y la propuesta G . Explícitamente, la estadística es dada por la expresión

$$G_i(d) = \frac{\sum_{j=1}^n w_{ij}(d)Y(A_j)}{\sum_{j=1}^n Y(A_j)} \quad \text{para } i \neq j \tag{A.7}$$

con $w_{ij}(d)$ es una matriz de pesos espaciales de unos y ceros, con unos para todas las entradas que pertenecen a un conjunto que señala que la ubicación de cada unidad espacial está a menos de una distancia d del punto i . Todas las otras entradas son cero, incluyendo la entrada del punto mismo, es decir, la diagonal de la matriz [30]. Es decir que esta estadística mide el grado de asociación que resulta de la concentración de puntos ponderados (o área representada por un punto muestral) y todos los otros puntos ponderados incluido entre un radio de distancia d del punto original ponderado. Esto quiere decir, que esta estadística tal cual como está descrita, mide la correlación espacial, pero a nivel local, existiendo el valor de cada uno de los puntos de importancia. Así, una generalización, una estadística útil para encontrar correlación global para todo el conjunto de datos, es dada por

$$G(d) = \frac{\sum_{j=1}^n \sum_{i=1}^n w_{ij}(d)Y(A_i)Y(A_j)}{\sum_{j=1}^n \sum_{i=1}^n Y(A_i)Y(A_j)} \quad \text{para } i \neq j \tag{A.8}$$

En el escrito se muestran los momentos de la estadística, como punto de partida para la inferencia de los datos. En efecto, para la estructura general de la estadística, que es la de interés, se tiene

$$E[G(d)] = W/[n(n - 1)] \tag{A.9}$$

en donde

$$W = \sum_{i=1}^n \sum_{j=1}^n w_{ij}(d) \quad \text{para } j \neq i \tag{A.10}$$

La varianza para esta estadística, es dada por

$$\text{Var}(G) = E(G^2) - \{W/[n(n - 1)]\}^2 \tag{A.11}$$

en donde ayudándose de la notación dada en Cliff and Ord [14], se tiene que

$$E(G^2) = \frac{1}{(m_1^2 - m_2)^2 n^{(4)}} [B_0 m_2^2 + B_1 m_4 + B_2 m_1^2 m_2 + B_3 m_1 m_3 + B_4 m_1^4] \tag{A.12}$$

donde $m_j = \sum_{i=1}^n Y(A_i)^j$, $j = 1, 2, 3, 4$, y $n^{(r)} = n(n - 1)(n - 2) \dots (n - r + 1)$. Los coeficientes son

$$\begin{aligned}
 B_0 &= (n^2 - 3n + 3)S_1 - nS_2 + 3W^2 \\
 B_1 &= -[(N^2 - n)S_1 - 2nS_2 + 3W^2] \\
 B_2 &= -[2nS_1 - (n + 3)S_2 + 6W^2] \\
 B_3 &= 4(n - 1)S_1 - 2(n + 1)S_2 + 8W^2 \\
 B_4 &= S_1 - S_2 + W^2
 \end{aligned}$$

Para la estadística $G_i(d)$ que observa correlación local, se tienen los siguientes resultados como apoyo de la inferencia:

$$E(G_i) = W_i/(n - 1) \quad (\text{A.13})$$

con $W_i = \sum_{j=1}^n w_{ij}(d)$ y además

$$\text{Var}(G_i) = \frac{W_i(n - 1 - W_i)Y_{i2}}{(n - 1)^2(n - 2)Y_{i1}^2} \quad (\text{A.14})$$

en donde se tienen las relaciones $Y_{i1} = \frac{\sum_{j=1}^n Y(A_j)}{(n-1)}$ y $T_{i2} = \frac{\sum_{j=1}^n Y(A_j)^2}{(n-1)} - Y_{i1}^2$. De esta manera, asumiendo que $G_i(d)$ es aproximadamente distribuido normal estándar, cuando

$$Z_i = \{G_i(d) - E[G_i(d)]\}/\sqrt{\text{Var}(G_i(d))} \quad (\text{A.15})$$

es positivamente o negativamente más grande que algún nivel de significancia específico, decimos que obtenemos asociación espacial positiva o negativa [30]. La estadística $G(d)$ debe ser usada en variables positivas con origen natural. Por eso no debe ser usada para probar la hipótesis de no correlaciones en los residuales de un eventual modelo lineal. Sin embargo, cabe mencionar que la estadística I tiene un problema con respecto a la G , y es que esta presenta debilidad en cuanto a discriminar patrones que tienen grandes valores dominantes o bajos valores dominantes dentro de la distancia d . Ambas estadísticas tienen dificultad con respecto a patrones aleatorios en los cuales hay una pequeña desviación con respecto de la media [30].

Otra propuesta para inferir sobre el comportamiento del parámetro que define la autocovariable, es que partiendo de lo anterior, en previo análisis al involucrar una autocovariable en cada uno de los modelos, puede probarse simplemente si hay autocorrelación espacial, haciendo uso de las pruebas de aleatorización. La base lógica para este tipo de pruebas es que se considera la hipótesis $\{\rho = 0\}$, es decir, la autocorrelación espacial es cero. Se supone que el plan (*orden en que se encuentran posicionadas cada una de las unidades espaciales*) que se ha usado, es escogido aleatoriamente de los

$$s = n! \quad (\text{A.16})$$

posibles planes (*posibles ubicaciones de las unidades espaciales*). Se enumeran estos planes por Π_γ ($\gamma = 1, 2, \dots, s$). Así, si la correlación espacial fuera cero, se debería tener exactamente el mismo resultado, es decir, el mismo valor de I o de C para cada Π_γ . En otras palabras, se encuentran los valores $\{G_\gamma, \gamma = 1, 2, \dots, s\}$ con G denominado un criterio,

que es el valor de uno de los estadísticos I o C . Sea G_0 el valor del criterio en el plan denominado actual (*en otras palabras* $G_0 = \rho$), que es el plan en donde las unidades espaciales conservan su estructura original. Así, el nivel de significancia de la prueba de que la autocorrelación espacial es cero contra la alternativa que es diferente, está dado por

$$NS = \frac{1}{s}[\text{Número de } G_\gamma(\gamma = 1, 2, \dots, s) \geq G_0] \quad (\text{A.17})$$

o de otra forma

$$NS = \frac{1}{s}[\text{Número de } I_\gamma(\gamma = 1, 2, \dots, s) \geq I_0] \quad (\text{A.18})$$

en donde I_γ es el coeficiente de Moran calculado con la estructura espacial definida por el plan γ e I_0 es el coeficiente de Moran calculado en el denominado plan original, que es la estructura espacial original. Para ver mas detalles, véase el capítulo 5 de [40].

Estimación por máxima verosimilitud e inferencia para el modelo SAR

Estimación por máxima verosimilitud

Anselin en algunas de sus publicaciones muestra las limitaciones de usar mínimos cuadrados ordinarios, en modelos que contengan algún tipo de procesos espaciales (En modelos CAR es usual y recurrente usar mínimos cuadrados).

En primer lugar, se hace $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$ y $\mathbf{B} = \mathbf{I} - \lambda\mathbf{W}_2$. Como $E[\boldsymbol{\xi}\boldsymbol{\xi}^T] = \boldsymbol{\Omega}$, existe un vector de errores aleatorios homocedasticos \boldsymbol{v} tales que $\boldsymbol{v} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\xi}$ y los errores definidos en (1.28) quedan

$$\boldsymbol{\epsilon} = \mathbf{B}^{-1}\boldsymbol{\Omega}^{1/2}\boldsymbol{v} \tag{B.1}$$

También se tiene con estas nuevas definiciones, que la ecuación (1.27) queda convertida en

$$\mathbf{A}\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{B.2}$$

luego, esto es

$$\mathbf{A}\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}^{-1}\boldsymbol{\Omega}^{1/2}\boldsymbol{v} \tag{B.3}$$

o alternativamente,

$$\boldsymbol{\Omega}^{1/2}\mathbf{B}(\mathbf{A}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \boldsymbol{v} \tag{B.4}$$

donde \boldsymbol{v} es normalmente distribuido con matriz de covarianzas \mathbf{I} . La expresión (B.4) es una función de los parámetros del modelo con estructuras no lineales. Aunque el término \boldsymbol{v} tiene una distribución, no puede ser observado, y la función de distribución tiene que ser basada en la de \mathbf{Y} . Así, es necesario hacer uso del concepto de jacobiano, el cual permite que la distribución para \mathbf{Y} sea derivada desde el vector \boldsymbol{v} , a través de la misma relación funcional (B.4) [4]. El jacobiano para esta transformación es:

$$J = \det(\partial\boldsymbol{v}/\partial\mathbf{Y})$$

lo que da de dicha expresión, como

$$|\boldsymbol{\Omega}^{1/2}\mathbf{B}\mathbf{A}| = |\boldsymbol{\Omega}^{1/2}||\mathbf{B}||\mathbf{A}|$$

y la función de log verosimilitud asociada por

$$L = -(N/2) \log(\pi) - (1/2) \log |\boldsymbol{\Omega}| + \log |\mathbf{B}| + \log |\mathbf{A}| - (1/2)\mathbf{v}^T\mathbf{v} \quad (\text{B.5})$$

con

$$\mathbf{v}^T\mathbf{v} = (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{B}^T\boldsymbol{\Omega}^{-1}\mathbf{B}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (\text{B.6})$$

como una suma de cuadrados de aproximados errores transformados [3]. Las expresiones resultantes para el vector de score igualado a cero son $(\mathbf{d} = (\partial L/\partial \boldsymbol{\theta}) = \mathbf{0}$, con $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta}^T, \lambda, \sigma^2, \boldsymbol{\alpha}^T)$ el vector de todos los $p + k + 3$ posibles parámetros del modelo)

$$\partial L/\partial \boldsymbol{\beta} = \mathbf{v}^T(\boldsymbol{\Omega}^{-1/2}\mathbf{B}\mathbf{X}) \quad (\text{B.7})$$

$$\partial L/\partial \rho = -\text{tr}\mathbf{A}^{-1}\mathbf{W}_1 + \mathbf{v}\boldsymbol{\Omega}^{-1/2}\mathbf{B}\mathbf{W}_1\mathbf{Y} \quad (\text{B.8})$$

$$\partial L/\partial \lambda = -\text{tr}\mathbf{B}^{-1}\mathbf{W}_2 + \mathbf{v}\boldsymbol{\Omega}^{-1/2}\mathbf{W}_2(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (\text{B.9})$$

$$\partial L/\partial \alpha_a = -(1/2)\text{tr}\boldsymbol{\Omega}^{-1}\mathbf{H}_a + (1/2)\mathbf{v}^T\boldsymbol{\Omega}^{-3/2}\mathbf{H}_a\mathbf{B}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (\text{B.10})$$

para $a = 1, 2, \dots, p$. Las expresiones para las entradas de la matriz de varianzas asintótica (matriz de información esperada), se pueden encontrar en [3] y [4]. Como se observa, ese sistema de ecuaciones es no lineal y no tiene una solución analítica, pero puede ser resuelta por métodos numéricos.

Pruebas de hipótesis usando ML

Un resultado muy conocido de los estimadores vía máxima verosimilitud es la normalidad asintótica. Anselin [4] lo escribe como

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow n \left(\mathbf{0}, \lim_{n \rightarrow \infty} (I(\boldsymbol{\theta})/n)^{-1} \right) \quad (\text{B.11})$$

resultado indicando que el estimador de máxima verosimilitud, converge en probabilidad al parámetro, con esperanza el parámetro y con varianza asintótica la inversa de la matriz de información esperada. Las hipótesis de interés en el caso lineal pueden ser formuladas de acuerdo a la expresión

$$H_0 : \mathbf{C}^T\boldsymbol{\theta} = \boldsymbol{\delta} \quad \text{versus} \quad H_1 : \mathbf{C}^T\boldsymbol{\theta} \neq \boldsymbol{\delta} \quad (\text{B.12})$$

para una matriz \mathbf{C} de tamaño $p + k + 3 \times g$ de rango g con coeficientes conocidos y $\boldsymbol{\delta}$ un vector con entradas $\boldsymbol{\delta} = (\delta_1, \dots, \delta_g)$. Se asume que $\text{ran}(\mathbf{C}) = g$, porque de otra manera algunas de las relaciones de H_0 serían redundantes y podrían ser obtenidas de otras. En la practica, se espera que dichas eventuales hipótesis redundantes sobre los parámetros, hayan sido eliminadas [58]. Anselin propone una forma más general para las hipótesis, asumiendo que pueden haber combinaciones no lineales de los parámetros, y asumiendo que dichas pueden ser estimables. Sin embargo, el enfoque de este escrito puede tener solamente tres puntos de referencia de manera particular, ellos son expresados por las siguientes hipótesis:

$$H_0 : \rho = 0 \quad \text{versus} \quad H_1 : \rho \neq 0 \quad (\text{B.13})$$

y

$$H_0 : \lambda = 0 \quad \text{versus} \quad H_1 : \lambda \neq 0 \quad (\text{B.14})$$

juegos que prueban la presencia o no de la correlación en la variable respuesta o en los errores, respectivamente. Otra hipótesis de interés es

$$H_0 : \boldsymbol{\alpha} = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\alpha} \neq \mathbf{0} \quad (\text{B.15})$$

probando así la presencia de heterocedasticidad. Para cada una de estas hipótesis, el mecanismo de prueba es el de siempre; hay que establecer una distancia que permita identificar que tan diferentes son los parámetros estimados bajo la hipótesis nula y los parámetros estimados sin ninguna restricción. Anselin, de manera muy general y sencilla, propone un juego de hipótesis de forma que el vector paramétrico pueda escribirse como una partición, es decir, $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]$, y que el eventual juego de hipótesis puede escribirse de manera general como

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0 \quad \text{versus} \quad H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^0 \quad (\text{B.16})$$

aunque en nuestro caso particular, se debería escribir la hipótesis nula como

$$H_0 : \boldsymbol{\theta}_1 = \mathbf{0}$$

Así, se deben estimar dos parámetros; el estimador de máxima verosimilitud sin ninguna restricción $\hat{\boldsymbol{\theta}}$ y el estimador de $\boldsymbol{\theta}$ restringido bajo la hipótesis nula, es decir, asumiendo que la partición $\boldsymbol{\theta}_1$ es cero. Este se denota como $\hat{\boldsymbol{\theta}}_R$. Cada una de las pruebas, deben basarse entonces en identificar que tan lejos o cerca (medida de distancia, usualmente usando chi cuadrado) están estos parámetros, y si es viable entonces aceptar que $\boldsymbol{\theta}_1$ sea cero, o si definitivamente rechazar la hipótesis nula. En efecto, se dan a conocer las estructuras de los estadísticos para las respectivas pruebas, a saber la de Wald, de Score y la de la Razón de Verosimilitud.

- **Test de Wald:** La estadística de Wald, para probar el juego de hipótesis (B.16), es expresado por la cantidad

$$\xi_W = [\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^0]^T \hat{\text{Var}}^{-1}(\hat{\boldsymbol{\theta}}_1) [\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^0] \quad (\text{B.17})$$

en que $\hat{\boldsymbol{\theta}}_1$ sale del vector $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ (Paula, G., 2010). La varianza aquí mostrada es proporcional a la inversa de la matriz de información de Fisher esperada, y basta con tomar la partición de interés. La estadística de Wald es asintóticamente distribuida chi cuadrado χ^2 con q grados de libertad, donde q corresponde al número de restricciones. Como la raíz de la estadística corresponde a una variable normal estándar, es equivalente asintóticamente a una variable distribuida t . En el contexto de probar la significancia de los parámetros del modelo, el test de Wald es comúnmente señalado como una t asintótica [4]. Para el juego de hipótesis (B.13), la estadística queda

convertida en

$$\xi_W = \hat{\rho}^2/v_{11} \quad (\text{B.18})$$

cuya distribución es chi cuadrado con un grado de libertad, en donde $\hat{\rho}$ es el estimador de máxima verosimilitud del parámetro ρ y v_{11} es el elemento de la diagonal correspondiente al elemento ρ en la matriz de varianzas y covarianzas asintótica. La deducción de las otras estadísticas de prueba de interés, son, para el juego de hipótesis (B.14)

$$\xi_W = \hat{\lambda}^2/v_{p+3,p+3} \quad (\text{B.19})$$

donde $\hat{\lambda}$ es el estimador de máxima verosimilitud de λ y $v_{p+3,p+3}$ es el elemento de la diagonal correspondiente al parámetro λ en la matriz de varianzas y covarianzas asintótica, al igual que el caso anterior, en este caso la distribución es chi cuadrado con un grado de libertad; y para el juego de hipótesis (B.15) la estadística es

$$\xi_W = \hat{\boldsymbol{\alpha}}^T \hat{\text{V}}\text{ar}^{-1}(\hat{\boldsymbol{\alpha}}) \hat{\boldsymbol{\alpha}} \quad (\text{B.20})$$

en donde $\hat{\boldsymbol{\alpha}}$ es el estimador de máxima verosimilitud de $\boldsymbol{\alpha}$ y $\hat{\text{V}}\text{ar}^{-1}(\hat{\boldsymbol{\alpha}})$ es la inversa de la matriz de varianzas y covarianzas asintótica particionada en los g parámetros de interés que definen la heterocedasticidad del modelo $\alpha_1, \dots, \alpha_g$, cuya distribución es chi cuadrado con g grados de libertad.

- **Test de Score:** Esta prueba se basa en la estadística de score. La forma general queda determinada por

$$\xi_S = \mathbf{U}_1(\hat{\boldsymbol{\theta}}^0)^T \hat{\text{V}}\text{ar}_0(\hat{\boldsymbol{\theta}}_1) \mathbf{U}_1(\hat{\boldsymbol{\theta}}^0) \quad (\text{B.21})$$

en que $\mathbf{U}_1(\boldsymbol{\theta}) = \partial L / \partial \boldsymbol{\theta}_1$, $\hat{\boldsymbol{\theta}}^0 = (\hat{\boldsymbol{\theta}}_1^0, \hat{\boldsymbol{\theta}}_2^0)$ en que $\hat{\boldsymbol{\theta}}_2^0$ es la estimativa de máxima verosimilitud de $\boldsymbol{\theta}_2$ bajo la hipótesis nula (recuerde que estamos trabajando bajo el juego de hipótesis (B.16)) [55]. La varianza mencionada en esa estadística, es la misma que el caso anterior, solo que es evaluada en la estimación de máxima verosimilitud para $\boldsymbol{\theta}_1$ restringida bajo la hipótesis nula.

- **Test de razón de verosimilitudes:** La función de verosimilitud del vector paramétrico $\boldsymbol{\theta}$ es

$$L(\boldsymbol{\theta}) = -(N/2) \log(\pi) + \log |\boldsymbol{\Omega}^{-1/2} \mathbf{B} \mathbf{A}| - (1/2) \mathbf{v}^T \mathbf{v} \quad (\text{B.22})$$

la verosimilitud restringida bajo H_0 (para el juego de hipótesis (B.13)), es

$$L(\boldsymbol{\theta}_R) = -(N/2) \log(\pi) + \log |\boldsymbol{\Omega}_R^{-1/2} \mathbf{B}_R| - (1/2) \mathbf{u}^T \mathbf{u} \quad (\text{B.23})$$

con

$$\mathbf{u}^T \mathbf{u} = (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}_R)^T \mathbf{B}_R^T \boldsymbol{\Omega}_R^{-1} \mathbf{B}_R (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}_R) \quad (\text{B.24})$$

La estadística de razón de verosimilitudes es:

$$\xi_{RV} = (\mathbf{u}^T \mathbf{u} - \mathbf{v}^T \mathbf{v}) + 2\{\log |\boldsymbol{\Omega}^{-1/2} \mathbf{B} \mathbf{A}| - \log |\boldsymbol{\Omega}_R^{-1/2} \mathbf{B}_R|\} \quad (\text{B.25})$$

con los coeficientes evaluados respectivamente en sus ecuaciones restringidas y no restringidas. La estadística, es entonces asintóticamente distribuida como χ^2 con un grado de libertad [4]. En este caso se usa el hecho que bajo H_0 , $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1 = \mathbf{I}$. Bajo el mismo procedimiento, la estadística para probar el juego de hipótesis (B.14) es

$$\xi_{RV} = (\mathbf{u}^T \mathbf{u} - \mathbf{v}^T \mathbf{v}) + 2\{\log |\boldsymbol{\Omega}^{-1/2} \mathbf{B} \mathbf{A}| - \log |\boldsymbol{\Omega}_R^{-1/2} \mathbf{A}_R|\} \quad (\text{B.26})$$

con

$$\mathbf{u}^T \mathbf{u} = (\mathbf{A}_{RY} - \mathbf{X}\boldsymbol{\beta}_R)^T \boldsymbol{\Omega}_R^{-1} (\mathbf{A}_{RY} - \mathbf{X}\boldsymbol{\beta}_R) \quad (\text{B.27})$$

y como el caso anterior, la estructura de $L(\boldsymbol{\theta}_R)$ queda como

$$L(\boldsymbol{\theta}_R) = -(N/2) \log(\pi) + \log |\boldsymbol{\Omega}_R^{-1/2} \mathbf{A}_R| - (1/2) \mathbf{u}^T \mathbf{u} \quad (\text{B.28})$$

Así, la distribución de la estadística ξ_{RV} es χ^2 con un grado de libertad. Aquí usamos el hecho que bajo H_0 , se tiene $\mathbf{B} = \mathbf{I} - \lambda\mathbf{W}_2 = \mathbf{I}$.

Anselin en [4] y [3] muestra una alternativa muy interesante para pruebas de hipótesis de este tipo bajo técnicas de estimación restringida por el método de multiplicadores de lagrange.

Residuales Modelo CAR Residual

La primera parte del apéndice, muestra los residuales del modelo lineal normal, que tiene la estructura (1.32). En principio, al realizar un simple análisis de residuales, encontramos que hay un ajuste adecuado, un R^2 de 0.5346 y un AIC de 338.81. De acuerdo al análisis residual, se usa la prueba de Shapiro Wilk para chequear normalidad, encontrando un p valor de 0.7263, comprobando normalidad. Esto se puede apreciar de acuerdo a la figura C.1. La figura C.1 muestra un análisis gráfico residual básico del modelo (1.32).

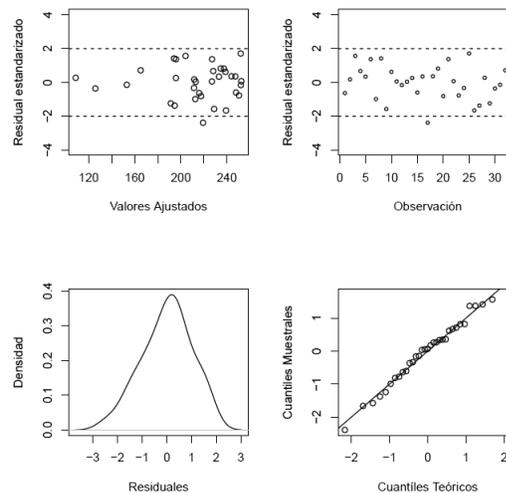


FIGURA C.1. Análisis gráfico residual del modelo (1.32)

De ahí, se observa una clara no correlación entre los valores estimados y los residuales, una normalidad aceptable, y una homocedasticidad no tan clara. De hecho, para valores ajustados mayores, mayor variabilidad en los residuales y viceversa, lo que hace pensar que no hay homocedasticidad en el modelo. Una característica no observable a simple vista, es que si se agrupan estos residuales por regiones naturales, se obtiene la estructura mostrada en la figura C.2. En tal caso, la región natural 1 es Amazonas, la 2 es Andina, la 3 es Caribe incluyendo Insular, la 4 es Orinoquía y la 5 es Pacífica: En cuanto a la figura C.2, se observa que los residuales se comportan de manera heterogénea de acuerdo a la región natural. Esto sugiere desde el principio que definitivamente falta un factor por incluir en el modelo, de manera que se cumplan todos los supuestos. Para este efecto, se

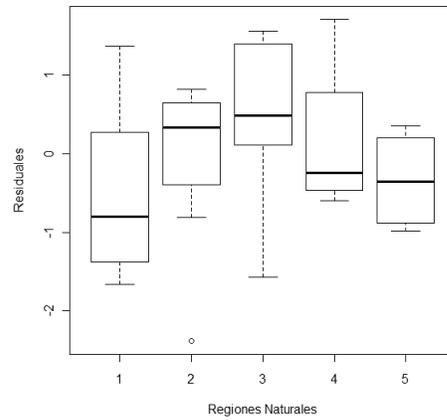


FIGURA C.2. Residuales del modelo (1.32) por regiones naturales

propone un modelo lineal mixto de la siguiente manera:

$$\mu = \beta_0 + \beta_1 \text{Displc} + \beta_2 \text{IndIPS} + \beta_3 \text{Vac} + \beta_4 \text{NBI} + \beta_5 1/(\text{Rec}) + \beta_6 \text{RNtr} + \alpha \quad (\text{C.1})$$

con α un efecto aleatorio para el factor Región Natural. En este caso, el valor de $\hat{\sigma}_\alpha^2 = 213.57$ y el valor de σ^2 cambia a 1201.40. El AIC es de 322.6, menor que el modelo lineal normal. Nótese que el factor aleatorio hace recoger un poco más de información al modelo que el anterior. Sin embargo, el modelo sigue pareciendo no tan adecuado en relación a los residuales. La figura C.3 muestra de nuevo el comportamiento de estos agrupados por regiones naturales. Es claro entonces, que no es conveniente modelar las regiones naturales

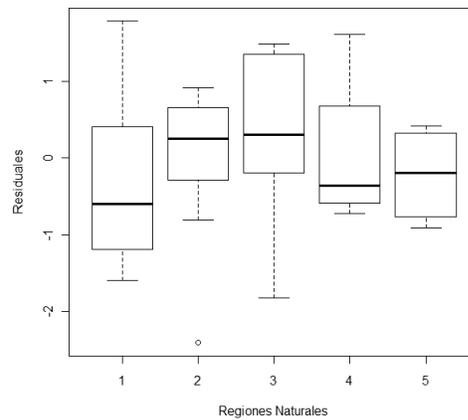


FIGURA C.3. Residuales del modelo (C.1) por regiones naturales

como un factor en la media. En tal caso, al parecer es conveniente pensar en un modelo de la forma (1.32) para la media, y para la varianza un factor que recoja y modele el comportamiento diferente de las regiones. El modelo CAR, que es el más adecuado, hace esto. La figura C.4 compara el comportamiento de los residuales del modelo lineal normal y lo que modela el CAR residual, usando la definición de vecindad por regiones naturales. Nótese que son muy similares en comportamiento.

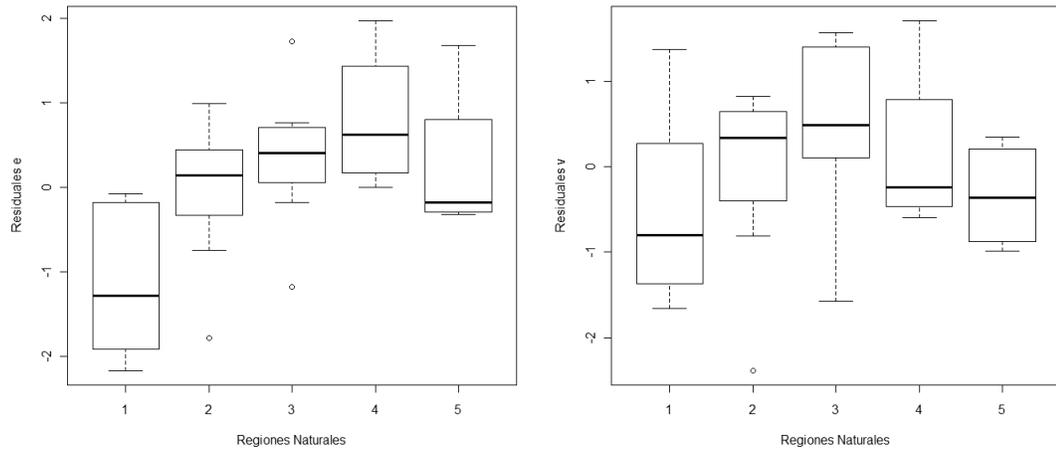


FIGURA C.4. Box Plot para los residuales del modelo lineal normal y el modelo error CAR de acuerdo a las regiones naturales

APÉNDICE D

Códigos R y Winbugs

Las aplicaciones del primer capítulo se realizaron en el paquete R y las del segundo en WinBugs. A continuación se muestran los códigos en R de los modelos del capítulo 1:

```
library(spdep)

### Modelo SAR Espacial con estructura espacial en el término de errores.
model.fit1 <- errorsarlm( TASA ~ DISPLC + INDIPS100 + VACUNA + NBI +
RECINV + RIESGO, data = Colombia, listw = Col, zero.policy = TRUE)

### Modelo SAR Espacial con estructura espacial en el proceso Y
model.fit2 <- lagsarlm(TASA ~ DISPLC + INDIPS100 + VACUNA + NBI +
RECINV + RIESGO ,data = Colombia, listw = Col, zero.policy=TRUE)

### Modelo SAR Espacial mixto
model.fit3 <- lagsarlm(TASA ~ DISPLC + INDIPS100 + VACUNA + NBI +
RECINV + RIESGO ,data = Colombia, listw = Col, zero.policy=TRUE,
type = "mixed")

### Modelo CAR Residual
model.fit4 <- spautolm(TASA ~ DISPLC+ INDIPS100 + VACUNA + NBI +
RECINV + RIESGO, data = Colombia, listw = ColC, zero.policy=TRUE,
family = "CAR")
```

A continuación se muestran algunos de los códigos empleados en WinBugs en el capítulo 2.

```
### Modelo Poisson

model
{
for( i in 1 : N ) {
VA[i] ~ dpois(lambda[i])
log(lambda[i]) <- beta[1]+ beta[2]*Desplazamiento[i] + beta[3]*INDIPS[i] +
```

```

beta[4]*Vac[i] + beta[5]*NBI[i] + beta[6]*Rec[i] + beta[7]*RiesgosN[i]
sresid[i] <- (VA[i] - lambda[i])/(sqrt(lambda[i]))
}
for(j in 1:7){
beta[j] ~ dnorm(0.0,1.0E-6)
}
}

### Modelo Poisson autocovariante

model
{
for( i in 1 : N ) {
VA[i] ~ dpois(lambda[i])
log(lambda[i]) <-beta[1]+ beta[2]*Desplazamiento[i] + beta[3]*INDIPS[i] +
beta[4]*Vac[i] + beta[5]*NBI[i] + beta[6]*Rec[i] + beta[7]*RiesgosN[i] +
beta[8]*Wy[i]
sresid[i] <- (VA[i] - lambda[i])/(sqrt(lambda[i]))
}
for(j in 1:8){
beta[j] ~ dnorm(0.0,1.0E-6)
}
}

### Modelo binomial

model
{
for( i in 1 : N ) {
Ncontrol[i] ~ dbin(p[i],NMamas99[i])
logit(p[i]) <- beta[1] + beta[2]*Viol[i] + beta[3]*NBI[i] + beta[4]*ES[i]
+ beta[5]*pago[i]
sresid[i]<-(Ncontrol[i]-NMamas99[i]*p[i])/sqrt(NMamas99[i]*p[i]*(1-p[i]))
}
for (j in 1:5){
beta[j] ~ dnorm(0.0,1.0E-6)
}
}

### Modelo binomial autocovariante

Model
{
for( i in 1 : N ) {
Ncontrol[i] ~ dbin(p[i],NMamas99[i])
logit(p[i]) <- beta[1] + beta[2]*Viol[i] + beta[3]*NBI[i] + beta[4]*ES[i]
+ beta[5]*pago[i] + beta[6]*Wy[i]
sresid[i]<-(Ncontrol[i]-NMamas99[i]*p[i])/sqrt(NMamas99[i]*p[i]*(1-p[i]))
}
}

```

```
for (j in 1:6){  
beta[j] ~ dnorm(0.0,1.0E-6)  
}  
}
```

Bibliografía

- [1] H. Ahn and J. J. Chen, *Generation of Over-Dispersed and Under-Dispersed Binomial Variates*, Journal of Computational and Graphical Statistics **4** (1995), no. 1, 55 – 64.
- [2] L. Anselin, *Estimation Methods for for Spatial Autoregressive Structures* , Regional Science Dissertation and Monograph Series **8** (1980).
- [3] ———, *Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity*, Geographical Analysis **24** (1988), no. 3, 189 – 206.
- [4] ———, *Spatial econometrics: Methods and models*, second ed., Kluwer Academic Publishers, University of California, Santa Barbara, 1988.
- [5] ———, *The future of spatial analysis in the social sciences*, Geographic Information Sciences **5** (1999), no. 2, 67–76.
- [6] N. H. Augustin, M.A. Mugglestone, and S.T. Buckland, *An autologistic model for the spatial distribution of wildlife*, J. Appl. Ecol. **33** (1996), 339 – 347.
- [7] J.O. Berger, *Statistical decision theory and bayesian analysis*, Springer, 1985.
- [8] J. Besag, *Spatial Interaction and the Statistical Analysis of Lattice Systems*, Journal of the Royal Statistical Society **36** (1974), no. 2, 192 – 236.
- [9] P. Bodson and Peeters. D., *Estimation of the Coefficients of a Linear Regression in the Presence of Spatial Autocorrelation: An Aplication to a Belgian Labour-Demand Function*, Environment and Planning **7** (1975), 445–472.
- [10] N. Breslow, *Extra Poisson Variation in log linear models*, Applied Statistics **33** (1984), no. 1, 38 – 44.
- [11] E. Cepeda and J. Achcar, *Heterocedastic nonlinear regression models*, Communications in Statistics - Simulation and Computation **39** (2010), 405–419.
- [12] E. Cepeda and D. Gamerman, *Bayesian Modelling Of Variance Heterogeneity In Normal Regression Models*, Brazilian Journal of Probability and Statistics (2001), no. 14, 207–221.
- [13] ———, *Bayesian methodology for modeling parameters in the two parameter exponential family*, Estadística **57** (2005), no. 168-169, 93–105.
- [14] A.D. Cliff and J.K. Ord, *Spatial Autocorrelation*, London:Pion (1973).

-
- [15] A.D. Cliff and J.K. Ord, *Spatial Processes, Models and Applications*, London:Pion (1981).
- [16] G.M. Cordeiro and C.G.B. Demétrio, *Modelos lineares generalizados*, 2007.
- [17] D.R. Cox and E.J. Snell, *Analysis of binary data*, Chapman and Hall, London, 1989.
- [18] M. J. Crowder, *Beta- Binomial anova for propoortiones*, Applied Statistics **27** (1978), no. 4, 34 – 37.
- [19] M. Dacey, *A Review of Measures of Contiguity for Two and K - Color Maps* , In Spatial Analysis: A Reader in Statistical Geography (1968), 479–495.
- [20] C. B. Dean, *Testing for overdispersion in Poisson and binomial regression models*, Journal of the American Statistical Association **87** (1992), no. 418, 451 – 457.
- [21] P. Dellaportas and G.O. Roberts, *An introduction to mcmc*, 2002.
- [22] D. Dey, S.K. Ghosh, and B.K. Mallick, *Generalized linear models: A bayesian perspective*, Marcel Dekker, 2000.
- [23] P.J. Diggle and P.J. Ribeiro, *Model-based geostatistics*, Springer, New York, 2007.
- [24] C.F. Dormann, J.M. McPherson, M.B. Araújo, R. Bivand, J. Bolliger, G. Carl, R.G. Davies, A. Hirzel, W. Jetz, W.D. Kissling, Kühn I., R. Ohlemüller, Peres-Neto P.R., B. Reineking, B. Schröder, F.M. Schurr, and R. Wilson, *Methods to account for spatial autocorrelation in the analysis of species distributional data*, Ecography **30** (2007), 609–628.
- [25] A. Dorothy and A. Murray, *Variance Component Models with Binary Response: Interviewer Variability*, Journal of the Royal Statistical Society **47** (1985), no. 2, 203 – 210.
- [26] B. Efron, *Double Exponential Families and Their Use in Generalized Linear Regression*, Journal of the American Statistical Association **81** (1986), no. 395, 709–721.
- [27] D. Gamerman, *Markov chains monte carlo: Stochastic simulation for bayesian inference.*, Chapman and Hall, London, 1997.
- [28] D. Gamerman and H.F. Lopes, *Markov chain monte carlo: Stochastic simulation for bayesian inference*, Chapman and Hall/CRC Texts in Statistical Science, 2006.
- [29] P. Geary, *The Contiguity Ratio and Statistical Mapping*, The Incorporated Statisticians **5** (1954), 115 – 145.
- [30] A. Getis and J.K. Ord, *The analysis of spatial association by use of distance statistics*, Geographical Analysis **24** (1992), no. 3, 189 – 206.
- [31] K.S. Gordon, *Generalized linear models with varying dispersion*, Journal of the Royal Statistical Society - Series B **51** (1989), no. 1, 47–60.
- [32] D. A. Griffiths, *Maximum Likelihood Estimation for the Beta-Binomial Distribution and an Application to the Household Distribution of the Total Number of Cases of a Disease*, Biometrics **29** (1973), no. 4, 637 – 648.

-
- [33] R. Haining, *Model Specification In Stationary Random Fields*, Geographical Analysis **9** (1977), no.-, 107 – 129.
- [34] ———, *Estimating Spatial Interaction Models*, Environment and Planning **10** (1978), no.-, 305 – 320.
- [35] ———, *Interaction Modelling of Central Place Lattices*, Journal of Regional Science **18** (1978), no.-, 217 – 228.
- [36] ———, *Statistical Test and Process Generators for Random Field Models*, Geographical Analysis **11** (1979), no.-, 45 – 64.
- [37] ———, *Testing a Spatial Interacting Markets Hypotesis*, The Review of Economics and Statistics **66** (1984), no.-, 576 – 583.
- [38] ———, *Spatial data analysis - theory and practice*, Cambridge University Press, 2003.
- [39] J. Hinde and C.G.B Demétrio, *Overdispersion: Models and Estimation*, Computational Statistics and data Analysis **27** (1998), 151 – 170.
- [40] K. Hinkelmann and O. Kempthorne, *Design and analysis of experiments*, Wiley, 1994.
- [41] J. Jiang, *Linear and generalized linear mixed models and their applications*, Springer, University of California, California, 2007.
- [42] M. Kaboli, A. Guillaumet, and R. Prodon, *Avifaunal gradients in two arid zones of central Iran in relation of vegetation, climate and topography*, J. Biogeogr **33** (2006), 133–144.
- [43] T.H. Keitt, O.N. Bjornstad, P.M. Dixon, and S. Citron-Pousty, *Accounting for spatial pattern when modeling organism - environment interactions*, Ecography **25** (2002), 616 – 625.
- [44] D. S. Klute and et al , *Autologistic regression modeling of American woodcock habitat use with spatially dependent data*, Island Press (2002), 335 – 343.
- [45] R. A. Knapp, K.R. Matthews, H.K. Preisler, and R. Jellison, *Developing probabilistic models to predict amphibian site occupancy in a patchy landscape*, Ecol. Appl **13** (2003), 1069 – 1082.
- [46] I. Kühn, *Incorporating spatial autocorrelation may invert observed patterns*, Biodiversity letter **13** (2007), no.-, 66 – 69.
- [47] J.W. Lichstein, T.R. Simons, and K.E. Franzreb, *Spatial Autocorrelation and Autoregressive Models in Ecology*, Ecological Monographs **72** (2002), no. 3, 445–463.
- [48] M. Luoto, M. Kuussaari, H. Rita, J. Salminen, and T. von Bonsdorff, *Determinants of distribution and abundance in the clouded apollo butterfly: a landscape ecological approach*, Ecography **24** (2001), 601–617.
- [49] L. Magee, *R2 measures based on Wald and likelihood ratio joint significance tests*, Am. Statistician **44** (1990), 250 – 253.
- [50] B.H. Margolin, N. Kaplan, and E. Zeiger, *Statistical Analysis of the Ames Salmonella/Mocrosome test*, Proc. Nat. Acad. Sci **76** (1981), 3779 – 3783.

-
- [51] P. McCullagh and J. A. Nelder, *Generalized linear models*, Chapman and Hall, London, 1989.
- [52] H. Migon and D. Gamerman, *Statistical inference: An integrated approach*, Arnold, New York, 1999.
- [53] P. Moran, *The Interpretation of Statistical Maps*, Journal of the Royal Statistical Society **10** (1948), 243 – 251.
- [54] N. J. D. Nagelkerke, *A Note on a General Definition of the Coefficient of Determination*, Biometrika **78** (1991), no. 3, 691–692.
- [55] G.A. Paula, *Modelos de regressão com apoio computacional*, Universidade de São Paulo, São Paulo, 2010.
- [56] D. Pregibon, *Review of generalized linear models, by p. mccullagh and j.nelder*, The Annals of Statistics **12** (1984), 1589–1596.
- [57] L.A. Quintero, E. Cepeda, and V. Nuñez, *Estimating infant mortality in colombia: some overdispersion modelling approaches*, Journal of Applied Statistics **39** (2012), no. 5, 1011–1036.
- [58] N. Ravishanker and D.K. Dey, *A first course in linear model theory*, Chapman and Hall / CRC, 2002.
- [59] L. Rincón, *Introducción a los procesos estocásticos*, Facultad de Ciencias UNAM, 2011.
- [60] B. D. Ripley, *Spatial statistics*, John Wiley And Sons, University of London, London, 1981.
- [61] C.P. Robert and G. Casella, *Introducing monte carlo methods with r*, Springer, 2010.
- [62] A. Sen and S. Soot, *Rank Test for Spatial Correlation*, Environment and Planning **9** (1977), 897 – 903.
- [63] M. Sköld and G.O. Roberts, *Density estimation for the metropolis-hastings algorithm*, Scandinavian Journal of Statistics **30** (2003), no. 4, 699–718.
- [64] A.F.M. Smith and G. O. Roberts, *Bayesian computation via the gibbs sampler and related markov chain monte carlo methods (with discussion)*, Journal of the Royal Statistical Society, Series B **55** (1993), no. 1, 3–23.
- [65] P. A. Smith, *Autocorrelation in logistic regression modelling of species’ distributions*, Global. Ecol. Biogeogr **4** (1994), 47 – 61.
- [66] J.W. Tukey, *Data analysis and regression : a second course in statistics*, Addison-Wesley, -, 1977.
- [67] J.C. Wakefield, A.F.M. Smith, A. Racine-Poon, and A.E. Gelfand, *Bayesian analysis of linear and non-linear population models by using the gibbs sampler*, Journal of the Royal Statistical Society, Series C **43** (1994), no. 1, 201–221.
- [68] P. Whittle, *On Stationary Processes in the Plane*, Biometrika **41** (1954), no. 3/4, 434 – 449.

-
- [69] D. A. Williams, *The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity*, *Biometrics* **31** (1975), no. 4, 949 – 952.
- [70] ———, *Extra - binomial variation in logistic linear models*, *Applied Statistics* **31** (1982), no. 2, 144 – 188.
- [71] B.A. Wintle and D.C. Bardos, *Modeling Species-Habitat Relationships with Spatially Autocorrelated Observation Data*, *Ecological Applications* **16** (2006), no. 5, 1945–1958.