UNIVERSIDAD DEL VALLE

TRABAJO DE GRADO

Modelación de la Razón Mortalidad Materna en Colombia: Un enfoque Multinivel

Autores:
José Alejandro ECHEVERRY
Sebastian Howard PACHECO

Director:
Ph.D Luz Adriana PEREIRA
Co Director:
Ph.D Cristian Eduardo GARCÍA

Trabajo de grado presentado como requisito parcial para optar al título de Estadístico en

Faculta de Ingenierías
Escuela de Estadística

17 de febrero de 2026

Declaración

Nos permitimos afirmar que hemos realizado el presente Trabajo de Grado de manera autónoma y con la única ayuda de los medios permitidos y no diferentes a los mencionados en el propio trabajo. Todos los pasajes que se han tomado de manera textual o figurativa de textos publicados y no publicados, los hemos reconocido. Ninguna parte del presente trabajo se ha empleado en ningún otro tipo de Tesis o Trabajo de Grado.

Igualmente declaramos que los datos utilizados en este trabajo están protegidos por las correspondientes cláusulas de confidencialidad.

Santiago de Cali, 14/12/2025

Jose Alejandro Echeverry Medina

Firmado digitalmente por Jose Alejandro Echeverry Medina
Fecha: 2026.02.17 12:20:46 -05'00'

Jose Alejandro Echeverry Medina

Sebastian Howard Pacheco Caicedo

Firmado digitalmente por Sebastian Howard Pacheco Caicedo
Fecha: 2026.02.17 12:18:35 -05'00'

Sebastian Howard Pacheco Caicedo

Resumen

La mortalidad materna constituye un indicador clave del estado de salud pública y del nivel de desarrollo social, al reflejar desigualdades en las condiciones socioeconómicas, el acceso a servicios y las características demográficas de los territorios. En Colombia, a pesar de los avances observados en la reducción de la Razón de Mortalidad Materna (RMM) a nivel nacional, persisten marcadas diferencias entre departamentos y municipios. En este contexto, el presente trabajo tiene como objetivo analizar el comportamiento de la RMM en Colombia durante el año 2023 en relación con factores sociodemográficos, excluyendo variables clínicas o físicas de salud, con el fin de identificar patrones territoriales y comprender cómo las condiciones estructurales se asocian con la ocurrencia y magnitud de la mortalidad materna.

Para abordar este objetivo, se adopta un enfoque bayesiano multinivel que permite modelar explícitamente la estructura jerárquica de los datos, con municipios anidados en departamentos. Dada la presencia de un elevado número de ceros en la variable de respuesta, se propone un modelo *hurdle* con inflación de ceros. La estimación del modelo se realiza exclusivamente mediante métodos bayesianos implementados en el lenguaje probabilístico *Stan*, utilizando simulación Monte Carlo por cadenas de Markov, lo que permite obtener inferencias completas sobre los efectos fijos y aleatorios, así como una adecuada cuantificación de la incertidumbre asociada a las estimaciones.

Los resultados evidencian una marcada heterogeneidad territorial en la RMM, con variabilidad significativa tanto a nivel departamental como municipal. En la componente binaria del modelo, las covariables asociadas a condiciones socioeconómicas desfavorables, como mayores niveles de pobreza, menor cobertura educativa y limitaciones en el acceso a servicios básicos, se relacionan con una mayor probabilidad de observar mortalidad materna. En la componente continua, una vez ocurren los eventos, los territorios con peores indicadores sociodemográficos presentan niveles esperados de RMM más elevados.

En conjunto, los hallazgos resaltan la importancia de considerar exclusivamente los determinantes sociodemográficos y la estructura jerárquica del territorio en el análisis de la mortalidad materna, aportando evidencia estadística relevante para la formulación de políticas públicas focalizadas en los contextos de mayor vulnerabilidad.

Palabras clave: Mortalidad Materna; Razón de Mortalidad Materna; Modelos Multinivel; Datos Multinivel; Modelo Hurdle; Análisis Bayesiano; Factores Sociodemográficos; Salud Pública; Colombia.

abstract

Maternal mortality constitutes a key indicator of public health status and social development, as it reflects inequalities in socioeconomic conditions, access to services, and demographic characteristics of territories. In Colombia, despite the observed progress in reducing the Maternal Mortality Ratio (MMR) at the national level, marked differences persist between departments and municipalities. In this context, the present study aims to analyze the behavior of the MMR in Colombia during 2023 in relation to sociodemographic factors, excluding clinical or physical health variables, in order to identify territorial patterns and understand how structural conditions are associated with the occurrence and magnitude of maternal mortality.

To address this objective, a Bayesian multilevel approach is adopted, which explicitly models the hierarchical structure of the data, with municipalities nested within departments. Given the presence of a high number of zeros in the response variable, a hurdle model with zero inflation is proposed. The model estimation is carried out exclusively using Bayesian methods implemented in the probabilistic language Stan, employing Markov Chain Monte Carlo simulation, which allows for complete inferences on both fixed and random effects, as well as an adequate quantification of the uncertainty associated with the estimates.

The results show marked territorial heterogeneity in the MMR, with significant variability at both the departmental and municipal levels. In the binary component of the model, covariates associated with unfavorable socioeconomic conditions, such as higher levels of poverty, lower educational coverage, and limitations in access to basic services, are related to a higher probability of observing maternal mortality. In the continuous component, once events occur, territories with poorer sociodemographic indicators present higher expected levels of MMR.

Overall, the findings highlight the importance of considering exclusively sociodemographic determinants and the hierarchical structure of the territory in the analysis of maternal mortality, providing relevant statistical evidence for the formulation of public policies targeted at the contexts of greatest vulnerability.

Keywords: Maternal Mortality; Maternal Mortality Ratio; Multilevel Models; Multilevel Data; Hurdle Model; Bayesian Analysis; Sociodemographic Factors; Public Health; Colombia.

Agradecimientos

Agradecemos de manera especial a nuestros directores de tesis, la **Doctora Luz Adriana Pereira** y el **Doctor Cristian Eduardo García**, por su acompañamiento constante, orientación académica y rigurosidad metodológica a lo largo del desarrollo de este trabajo. Sus aportes fueron fundamentales para la correcta formulación del modelo estadístico y la interpretación de los resultados.

Asimismo, agradecemos a la Doctora Erika Johanna Cantor y al Doctor Jaime Mosquera Restrepo por el apoyo brindado, las observaciones realizadas y la disposición para orientar aspectos específicos de esta investigación, contribuyendo al fortalecimiento del enfoque metodológico adoptado.

Finalmente, expresamos nuestro agradecimiento a los docentes de la Escuela de Estadística, quienes, a través de su formación y acompañamiento académico, aportaron de manera directa e indirecta a nuestra formación profesional y al desarrollo de las competencias necesarias para la realización de este trabajo de grado.

Índice general

Declaración II

Resumen III

abstract IV

Agradecimientos V

1. Introducción **1**
1.1. Planteamiento del Problema 1
1.2. Justificación 4
1.3. Pregunta de investigación 6
1.4. Objetivo General 6
1.4.1. Objetivos Específicos 6

2. Marco Teórico **7**
2.1. Marco conceptual 7
2.1.1. Mortalidad materna y Razón de Mortalidad Materna 7
2.1.2. Factores sociodemográficos 8
2.2. Marco estadístico 8
2.2.1. Metodología Delphi 8
2.2.2. Datos Multinivel 8
2.2.3. Análisis y modelos multinivel 9
2.2.4. Modelo Hurdle Multinivel 9
Componente binaria 9
Componente continua 10
Priors jerárquicas 10
Verosimilitud general 10
Distribución a posteriori 10
2.2.5. Estimación Bayesiana 11
2.2.6. Análisis bayesiano objetivo 12
Parametrización no centrada en modelos jerárquicos bayesianos 12
Fundamentos de los métodos MCMC 13
Métodos avanzados: Hamiltonian Monte Carlo y NUTS 13
2.2.7. Modelos Gaussianos Latentes (MGLs) 15
2.2.8. Campos aleatorios de Markov gaussianos (GMRF) 15
2.2.9. Evaluación del modelo y selección del modelo 15
Criterios basados en la información (DIC y WAIC) 15

Transformación integral predictiva (PIT) . . . 16
Factor de reducción de escala potencial ($\hat{R}$) . . . 16
Probability of Direction (pd) . . . 17
Intervalo creible . . . 17
Tamaño efectivo de muestra en simulaciones MCMC . . . 18
2.2.10. Autocorrelación espacial . . . 19
Índice de Moran e Índice General de Getis & Ordon . . . 20
2.2.11. Inflación de ceros . . . 21
2.2.12. Análisis de Componentes Principales (ACP) . . . 21
Análisis de la nube de individuos ($\mathbb{R}^p$) . . . 22
Análisis de la nube de variables ($\mathbb{R}^n$) . . . 23
Representación Simultánea . . . 23
2.3. Antecedentes . . . 23

3. Metodología **26**
3.1. Descripción de los datos . . . 26
3.1.1. Análisis estadístico . . . 27
Análisis exploratorio de datos . . . 27
Selección de las variables predictoras . . . 28
Análisis via ACP . . . 28
3.2. Modelo . . . 28
3.2.1. Definición del modelo estadístico para la RMM: un enfoque bayesiano . . . 28
3.2.2. Uso del análisis bayesiano objetivo en la especificación del modelo . . . 29
3.2.3. Modelo hurdle para RMM con inflación de ceros . . . 29
3.2.4. Efectos aleatorios con dependencia espacial . . . 31
3.3. Modelo completo . . . 33
3.3.1. Verosimilitud del modelo Hurdle . . . 34
3.3.2. Distribuciones a priori . . . 34
3.3.3. Distribución a posteriori . . . 37
3.3.4. Diagnóstico del modelo . . . 38
3.3.5. Estimación e interpretación de los coeficientes del modelo . . . 39

4. Resultados y análisis **40**
4.1. Resultado 1: Análisis descriptivo, espacial y de agrupamiento territorial de la Razón de Mortalidad Materna en Colombia . . . 40
4.1.1. Introducción . . . 40
4.1.2. Análisis descriptivo a nivel departamental . . . 41
Análisis espacial exploratorio . . . 43
Análisis univariado . . . 44
Análisis de Componentes Principales (ACP) . . . 46
4.1.3. Análisis descriptivo a nivel departamental . . . 50
4.1.4. Evaluación de agrupamientos territoriales y análisis comparativo . . . 53
Correlación espacial . . . 53
Caracterización de los grupos de departamentos . . . 57
4.1.5. Síntesis del Resultado Uno . . . 58
4.2. Resultado 2: Modelado bayesiano de la Razón de Mortalidad Materna 60
4.2.1. Diagnóstico de supuestos: Indenpendencia condicional y linealidad . . . 61

Linealidad en el logit (parte Bernoulli) 61
Independencia condicional (parte Bernoulli) 62
Linealidad en el logit (parte Lognormal) 62
Independencia condicional (parte Lognormal) 63
4.2.2. Diagnóstico y convergencia del muestreo MCMC 64
Inspección visual del muestreo 65
Parámetros de referencia del modelo hurdle 67
4.2.3. Efectos aleatorios jerárquicos 68
4.2.4. Predicciones posteriores y evaluación del ajuste 72
4.3. Resultado 3. Factores sociodemográficos asociados a la Razón de Mortalidad Materna en Colombia, 2023 80
4.3.1. Resultados del componente Bernoulli: ocurrencia de mortalidad materna 81
4.3.2. Resultados del componente lognormal: magnitud de la Razón de Mortalidad Materna condicionada a su ocurrencia 82
4.3.3. Síntesis del Resultado tres 84

5. Conclusiones **85**

6. Discusión **88**

7. Anexos **90**

Índice de figuras

2.1. Estructuras de datos multinivel. 9

4.1. Razón de Mortalidad Materna(RMM) por Departamento 42
4.2. Mapa de calor (heatmap) de la RMM registrada por Departamento . . 43
4.3. Boxplots de las variables sociodemográficas asociadas a la RMM 44
4.4. Correlograma de variables sociodemográficas y RMM 45
4.5. ACP: valores propios de cada componente (izquierda), varianza acumulada (%) explicada por los componentes (derecha) 46
4.6. Circulo de correlación 47
4.7. número óptimo de clusters mediante WSS (Izquierda) y Gap Statistic (Derecha) 48
4.8. Mapa factorial de los departamentos según las variables sociodemográficas 49
4.9. Biplot simultáneo de individuos (departamentos) y variables sociodemográficas 49
4.10. Porcentaje de datos faltantes a nivel departamental 51
4.11. RMM categorizado por rangos 51
4.12. Mapa de calor (heatmap) de la RMM registrada por municipio 52
4.13. Mapa coroplético (choropleth map) - Departamentos 55
4.14. Mapa coroplético (choropleth map) - Municipios 55
4.15. RMM por Departamento con Identificación de Clústeres 57
4.16. Linealidad en el logit: Residuos agrupados (Binned Residuals) 61
4.17. Independencia condicional por departamento 62
4.18. Linealidad del logit: QQ-plot 63
4.19. Independencia condicional por departamento 63
4.20. Traceplots de los parámetros principales del modelo hurdle multinivel. 66
4.21. Densidades posteriores de los parámetros principales del modelo hurdle multinivel 66
4.22. Áreas de credibilidad al 90 % de los parámetros principales del modelo hurdle multinivel 67
4.23. Predicción posterior media de la RMM frente a los valores observados por municipio. 73
4.24. Distribución de los residuos posteriores (RMM observada menos predicción posterior media). 73
4.25. Predicción posterior media de la RMM a nivel municipal. Modelo hurdle multinivel bayesiano (Stan). 75

4.26. Incertidumbre posterior (amplitud del intervalo creíble del 90 %) de la RMM predicha a nivel municipal. 76
4.27. Clasificación municipal en quintiles según la media posterior de RMM. 77
4.28. Predicción posterior de la RMM a nivel departamental (promedio de la predicción municipal). 78

7.1. Densidades posteriores de los coeficientes de regresión correspondientes al componente Bernoulli del modelo hurdle multinivel. 92
7.2. Densidades posteriores de los coeficientes de regresión correspondientes al componente lognormal del modelo hurdle multinivel. 93

Índice de cuadros

3.1. Diccionario de datos. 27

4.1. Estadísticas descriptivas utilizadas 41
4.2. Estadísticas descriptivas por variable - Departamentos 41
4.4. Departamentos identificados como valores atípicos por variable 45
4.5. Pruebas globales de autocorrelación espacial para Municipios y Departamentos 54
4.6. Clasificación de municipios según el estadístico Local Moran (LISA) . 54
4.7. Estadísticas descriptivas por cluster (Clusters 1, 2 y 3) 57
4.8. Indicadores de información para los modelos ajustados (Parte continua) 60
4.9. Diagnóstico de convergencia para los parámetros principales del modelo hurdle multinivel. 65
4.10. Estimaciones posteriores de la variabilidad jerárquica y residual del modelo hurdle. 68
4.11. Media posterior para la probabilidad de presentar un mayor riesgo de mortalidad materna en los departamentos (efectos aleatorios, componente Bernoulli) 69
4.12. Media posterior de los efectos aleatorios departamentales en la parte continua (lognormal) del modelo hurdle, para los departamentos con mayor probabilidad de ocurrencia de mortalidad materna. 70
4.13. Media posterior para la probabilidad de presentar un mayor riesgo de mortalidad materna en los municipios (efectos aleatorios, componente Bernoulli). 71
4.14. Media posterior de los efectos aleatorios municipales en la parte continua (lognormal) del modelo hurdle, para los municipios con mayor probabilidad de ocurrencia de mortalidad materna. 71
4.15. Resumen de la distribución posterior de la RMM en municipios seleccionados. 72
4.16. Estimaciones posteriores de los efectos fijos de las covariables sociodemográficas sobre la probabilidad de ocurrencia de mortalidad materna (componente Bernoulli del modelo hurdle) 82
4.17. Estimaciones posteriores de los efectos fijos de las covariables sociodemográficas en la magnitud de la RMM (componente lognormal del modelo hurdle) 83

7.1. Tabla de Estadísticas por Departamento n, Media, Desviación estándar y Coeficiente de variación 91

7.2. Tabla de Estadísticas por Departamento n, Media, Desviación estándar y Coeficiente de variación 92

List of Abbreviations

MM	**M**ortalidad **M**aterna
RMM	**R**azón de **M**ortalidad **M**aterna
TMM	**T**asa de **M**ortalidad **M**aterna
PM	**P**roporción de **M**uertes maternas
LTR	**L**ifetime **R**isk (Riesgo de muerte materna a lo largo de la vida)
SD	**D**esviación **E**stándar (Standard Deviation)
CV	**C**oeficiente de **V**ariación
NBI	**N**ecesidades **B**ásicas **I**nsatisfechas
IPM	**Í**ndice de **P**obreza **M**ultidimensional
IICA	**Í**ndice de **I**nfraestructura y **C**onectividad **A**grícola
IRCA	**Í**ndice de **R**iesgo de la **C**alidad del **A**gua
PIB	**P**roducto **I**nterno **B**ruto
INS	**I**nstituto **N**acional de **S**alud
OMS	**O**rganización **M**undial de la **S**alud
ZIP	**Z**ero-**I**nflated **P**oisson
ZINB	**Z**ero-**I**nflated **N**egative **B**inomial
MCMC	**M**arkov **C**hain **M**onte **C**arlo
HMC	**H**amiltonian **M**onte **C**arlo
NUTS	**N**o-**U**-**T**urn **S**ampler
WAIC	**W**idely **A**pplicable **I**nformation **C**riterion
DIC	**D**eviance **I**nformation **C**riterion
R	Lenguaje y entorno estadístico **R**
Stan	Lenguaje para modelos **S**tatistical **An**alysis

Dedicamos este trabajo, de manera especial, a nuestras familias, quienes brindaron el mismo apoyo, acompañamiento y comprensión a lo largo de todo este proceso académico. Su respaldo constante, paciencia y confianza fueron fundamentales para la culminación de esta etapa de formación.

A ***Alba Gladys Medina Jiménez*** *y* ***Anyi Vanessa Monard Medina****, así como a* ***Amparo Caicedo Ibarra****,* ***Howard Pacheco Chicaiza*** *y* ***Lizeth Pacheco Caicedo****, por su tiempo, apoyo incondicional y motivación permanente, que hicieron posible el desarrollo y finalización de este trabajo de grado.*

De manera complementaria, agradecemos a nuestros compañeros y docentes con quienes compartimos aprendizajes y experiencias a lo largo de la carrera, y cuyo acompañamiento contribuyó de forma significativa a este proceso académico.

Capítulo 1

Introducción

1.1. Planteamiento del Problema

La mortalidad materna (MM) constituye una de las causas principales del fallecimiento entre las mujeres durante la etapa de gestación y posparto, ya sea a nivel nacional o internacional, siendo considerada una problemática de salud pública y derechos humanos, puesto que altos índices de mortalidad materna conllevan a la trasgresión de los derechos de la mujer y de las niñas a la vida, la salud, la igualdad, la no discriminación y el acceso a los avances del conocimiento científico y del mas alto estándar de salud alcanzable (Instituto Nacional de Salud (INS), 2022a)

Esta es definida por la Organización Mundial de la Salud - OMS - como "la muerte de una mujer durante su embarazo, parto o dentro de los 42 días después de su terminación por cualquier causa relacionada o agravada por el embarazo, parto o puerperio o su atención, pero no por causas accidentales o incidentales" (Departamento Administrativo Nacional de Estadística (DANE), 2021)

De Acuerdo a estudios previos en la república de Colombia, tales como: "Informe de Mortalidad Materna, Colombia, 2022" , se identifican características sociodemográficas y de acceso a servicios de salud asociadas a esta problemática, tales como el estrato socioeconómico, la edad, el nivel educativo, la pertenencia étnica, área de residencia (urbano o rural), no estar afiliado al Sistema General de Seguridad Social en Salud y no tener acceso a controles prenatales. (Instituto Nacional de Salud (INS), 2022b)

Asimismo, dicho informe evidenció una fuerte correlación entre la razón de mortalidad materna (RMM) y el índice de pobreza monetaria $(r = 0.72)$, destacando departamentos como La Guajira, Nariño y Chocó, donde coinciden altos niveles de pobreza y mortalidad materna. En general, los factores sociales y económicos vinculados a la desigualdad representan una de las principales causas de este problema en el país. Un hallazgo relevante es el efecto de la pandemia, pues se estimó que la RMM aumentó en 15.2 puntos en 2020, principalmente por la reducción en la asistencia a controles prenatales (Departamento Administrativo Nacional de Estadística (DANE), 2021). Dado que los departamentos menos favorecidos en términos socioeconómicos, tienden a poseer una razón de mortalidad más alta. Debido a esto, se busca caracterizar los departamentos en base al comportamiento de la razón de mortalidad materna, con el propósito de modelar este indicador, basado en los factores de mayor riesgo disponibles en Terridata y DANE.

Junto con los desafíos epidemiológicos y sociales, el estudio de la Razón de Mortalidad Materna (RMM) presenta dificultades estadísticas particulares que deben ser consideradas para lograr una comprensión adecuada del fenómeno. La información disponible a nivel municipal suele contener una alta proporción de valores iguales a cero, tanto por la baja frecuencia del evento como por posibles limitaciones en los sistemas de registro. Cuando no se registran muertes maternas en un municipio durante un período determinado, el conteo del evento es igual a cero y, en consecuencia, la Razón de Mortalidad Materna toma el valor de cero, dado que el número de muertes maternas constituye el numerador del indicador, tal como se define en la ecuación 2.1. Esta situación da lugar a una acumulación de valores nulos en el indicador de RMM, la cual puede deberse tanto a la ausencia de muertes maternas registradas como a limitaciones en los sistemas de información, procesos administrativos u otros factores que afecten el registro adecuado de los eventos.

De acuerdo con la definición presentada en la ecuación 2.1, la Razón de Mortalidad Materna se construye a partir del conteo de muertes maternas, considerando el número de nacidos vivos como variable de exposición. En la práctica, la modelación de este indicador suele abordarse mediante modelos de regresión para datos de conteo, en los cuales el número de muertes maternas constituye la variable respuesta y los nacidos vivos se incorporan como un término de ajuste que permite estimar el riesgo relativo asociado a distintos factores sociodemográficos y de acceso a servicios de salud.

Cuando estas aproximaciones se aplican a escalas territoriales desagregadas, como municipios o departamentos, la presencia de una proporción elevada de valores nulos en el conteo de muertes maternas plantea limitaciones para los enfoques convencionales de modelación. En particular, estas metodologías no permiten representar de manera diferenciada los procesos asociados a la ocurrencia del evento y a la magnitud del indicador en los casos en que se observan valores positivos, lo que motiva la consideración de estrategias de modelación alternativas.

Diversos autores han mostrado que este tipo de estructuras estructuras generan un exceso de ceros que no puede ser explicado por modelos clásicos de regresión para conteos, los cuales tienden a subestimar la probabilidad real de observar valores nulos y a producir estimaciones sesgadas en la parte positiva de la distribución (Min, Y. and Agresti, A., 2005). Este comportamiento ha sido identificado también en estudios de mortalidad materna en contextos comparables, como en Etiopía, donde la mayoría de distritos presentaban cero o muy pocas muertes anuales, generando dispersión elevada e inestables estimaciones cuando se aplicaron métodos tradicionales (Jabessa y Jabessa, 2021).

Adicionalmente, la estructura territorial agrega un desafío adicional. En escenarios donde los municipios se encuentran anidados en departamentos, y donde los territorios vecinos comparten condiciones sociales, económicas o de acceso a servicios, es habitual que los riesgos de mortalidad no sean independientes entre sí. En estudios realizados en Etiopía se observó que más del 50 % de la variabilidad total de la mortalidad materna se explica por diferencias entre regiones, lo que evidencia la existencia de heterogeneidad no observada y la necesidad de incorporar efectos aleatorios jerárquicos (Jabessa y Jabessa, 2021). De forma similar, en el Nordeste de Brasil se identificaron conglomerados territoriales persistentes mediante estadística espacial, mostrando que los municipios con mayor mortalidad tienden a agruparse geográficamente (Oliveira, I. V. G, et all, 2023).

La inestabilidad del indicador en municipios de baja población constituye otra

dificultad relevante. En Brasil se documentó que la RMM presenta fluctuaciones abruptas en áreas con pocos nacidos vivos, lo que llevó a aplicar métodos de suavizamiento basados en enfoques bayesianos para disminuir la variabilidad aleatoria y estabilizar las tasas a nivel municipal (Oliveira, I. V. G, et all, 2023). Este comportamiento es característico de los indicadores construidos en pequeñas áreas, donde pequeñas variaciones en los eventos pueden producir cambios desproporcionados en las tasas.

A lo anterior se suma que los valores positivos de la RMM suelen presentar marcada asimetría y dispersión, típica de razones epidemiológicas asociadas a eventos poco frecuentes. Este tipo de comportamiento ha sido destacado en modelos de conteo con truncamiento, donde la parte positiva del indicador requiere distribuciones adecuadas para capturar la variabilidad real observada (Min, Y. and Agresti, A., 2005).

Estas características configuran un problema estadístico complejo, en el que coexisten exceso de ceros, sobre-dispersión, dependencia territorial y variabilidad estructural asociada a niveles administrativos. Frente a este escenario, estudios recientes han mostrado que los modelos jerárquicos bajo un enfoque bayesiano permiten integrar de manera simultánea estas fuentes de variabilidad. En el caso etíope, por ejemplo, los modelos multinivel bayesianos permitieron capturar adecuadamente la variación entre regiones, producir estimaciones más estables y seleccionar modelos más apropiados mediante criterios como el DIC (Jabessa y Jabessa, 2021). De manera complementaria, los análisis realizados en Brasil destacaron que los métodos bayesianos empíricos contribuyen a estabilizar tasas en municipios pequeños y a reducir el efecto de fluctuaciones aleatorias (Oliveira, I. V. G, et all, 2023).

Así, el análisis de la RMM en Colombia plantea la necesidad de un enfoque metodológico que responda a las particularidades de los datos: un conjunto de observaciones con predominio de ceros, heterogeneidad territorial y estructuras jerárquicas. Un modelo de tipo hurdle con componentes multinivel, estimado dentro de un marco bayesiano, permite abordar estas características de manera integrada y coherente, y ofrece una representación más adecuada del riesgo materno y de su relación con los factores sociodemográficos disponibles en fuentes oficiales como Terridata y el DANE.

Finalmente, el presente documento se estructura en siete capítulos. El Capítulo 1 introduce el contexto epidemiológico y estadístico del problema, así como su pertinencia en el marco de la salud pública en Colombia. El Capítulo 2 desarrolla el marco teórico, en el que se revisan los conceptos centrales sobre mortalidad materna, modelos jerárquicos y métodos bayesianos utilizados en la literatura especializada. En el Capítulo 3 se describe la metodología empleada, incluyendo las fuentes de información, los procedimientos de selección de variables y la formulación del modelo hurdle multinivel. El Capítulo 4 presenta los resultados obtenidos a partir de la estimación del modelo y su interpretación. El Capítulo 5 expone las conclusiones principales derivadas del análisis, mientras que el Capítulo 6 discute los hallazgos en relación con estudios previos, las limitaciones del enfoque adoptado y posibles líneas de investigación futura. Finalmente, el Capítulo 7 reúne los anexos técnicos y complementarios que respaldan el desarrollo metodológico y empírico del estudio.

1.2. Justificación

De acuerdo a datos oficiales proporcionados por el Departamento Administrativo Nacional de Estadística (DANE) , la razón de mortalidad materna en Colombia ha presentado una tendencia decreciente con el paso de los años. En 2007 se registraron 73.3 casos por 100.000 nacidos vivos, mientras que en 2017 esta cifra disminuyó a 51 casos. En los años posteriores, como 2018 y 2019 la razón de mortalidad materna se mantuvo inferior a los 50 casos por 100.000 nacidos vivos. Por lo contrario, los años 2020 y 2021 significaron un aumento considerable en esta razón, tanto a nivel nacional como internacional, debido a la crisis sanitaria sin precedentes que desestabilizó la economía mundial y cambió drásticamente la vida de miles de millones de personas (Departamento Administrativo Nacional de Estadística (DANE), 2021), (Instituto Nacional de Salud (INS), 2022a)

El aumento, descenso o cualquier comportamiento observable en la RMM, tiende a estar relacionado con los factores sociales, económicos y demográficos, por consiguiente, se considera importante identificar las causas sociodemográficas asociadas a la mortalidad materna, ya que entender estos factores permite enfrentar el problema de manera integral.

Basándose en los Objetivos de Desarrollo Sostenible (ODS) (reducir para el 2030 la tasa mundial de mortalidad materna a menos de 70 por cada 100.000 nacidos vivos) desarrollados por la Organización Mundial de la Salud (OMS), planteados desde el año 2015, Colombia se fijó la meta de reducir la mortalidad materna de manera progresiva hasta llegar a 32 casos por cada 100.000 nacidos vivos mediante distintos planes sociales y otras alternativas (Instituto Nacional de Salud (INS), 2020)

Si bien se han registrado reducciones en la RMM, estas no se presentan de manera homogénea entre las entidades territoriales, como lo reflejan los valores reportados en departamentos como Vichada, Chocó y La Guajira (Instituto Nacional de Salud (INS), 2020)

Por lo anterior, comprender el comportamiento de la RMM y reconocer la complejidad de los factores que influyen en la misma, considerando especialmente la pluriculturalidad del país y las diversas realidades sociales, económicas y geográficas presentes en Colombia, que inciden directamente en esta problemática. En este sentido, además del análisis epidemiológico, resulta necesario incorporar una justificación de orden estadístico-metodológico que permita abordar adecuadamente las características propias de los datos disponibles y las exigencias analíticas del fenómeno.

La estimación de la RMM a nivel territorial enfrenta retos particulares derivados de la estructura de los datos. En numerosos municipios colombianos se registran valores iguales a cero en un año determinado, lo cual es esperable dada la baja ocurrencia del evento. Este patrón no es exclusivo del contexto colombiano; Min, Y. and Agresti, A., 2005 muestran que los conteos con una proporción elevada de ceros generan sobre-dispersión y dificultan el uso de modelos clásicos, recomendando estrategias específicas como los modelos de tipo hurdle para separar la ocurrencia del evento de la magnitud cuando este se presenta. A lo anterior se suma la heterogeneidad propia de las realidades territoriales, donde factores culturales, geográficos y de acceso a servicios de salud pueden modificar la probabilidad de registrar un caso materno y, por tanto, incidir en la distribución estadística del indicador.

Asimismo, en municipios con pocos nacidos vivos anuales, las tasas de mortalidad materna pueden fluctuar de manera considerable debido a la sensibilidad extrema del indicador a pequeños cambios en el número de eventos. En el estudio realizado en el Nordeste de Brasil, Oliveira et al. (2023) describen cómo estas variaciones generan tasas poco estables y dificultan la interpretación, motivo por el cual aplicaron procedimientos de suavizamiento basados en métodos bayesianos para mejorar la consistencia de las estimaciones territoriales (Oliveira, I. V. G, et all, 2023). Este tipo de estrategia es particularmente pertinente para Colombia, donde una proporción significativa de municipios presenta poblaciones pequeñas y patrones irregulares de registro.

En cuanto a la estructura territorial, si bien el planteamiento del problema identifica la importancia de las diferencias entre municipios y departamentos, desde la perspectiva metodológica es relevante destacar que las aproximaciones clásicas no siempre permiten capturar adecuadamente estas variaciones. Estudios como el desarrollado en Etiopía por Jabessa y Jabessa (2021) muestran que la incorporación explícita de estructuras jerárquicas mediante modelos multinivel mejora la capacidad para representar la variabilidad entre unidades administrativas, facilita la interpretación de los determinantes a diferentes niveles y ofrece estimaciones más coherentes con la organización del territorio (Jabessa y Jabessa, 2021).

La elección de un enfoque bayesiano responde también a consideraciones estadísticas fundamentales asociadas a la incertidumbre en los datos y a la complejidad de los modelos requeridos. El uso de métodos bayesianos, como lo exponen Jabessa y Jabessa (2021) en su análisis de mortalidad materna, permite integrar fuentes de variación entre niveles, tratar de manera natural parámetros de efectos aleatorios y evaluar el ajuste del modelo mediante criterios como el DIC (Jabessa y Jabessa, 2021). En el caso del análisis brasileño, Oliveira et al. (2023) ilustran que los métodos bayesianos empíricos son útiles para estabilizar tasas en áreas pequeñas, reducir fluctuaciones aleatorias y mejorar la calidad de los indicadores utilizados para la vigilancia epidemiológica (Oliveira, I. V. G, et all, 2023).

Considerando estos antecedentes, el uso de un modelo hurdle multinivel bajo un enfoque bayesiano constituye una opción metodológica adecuada para caracterizar la Razón de Mortalidad Materna en Colombia. Esta aproximación permite abordar de manera conjunta el exceso de ceros, la inestabilidad en municipios pequeños y la existencia de estructuras territoriales jerárquicas, proporcionando estimaciones más sólidas y coherentes con la complejidad del fenómeno y contribuyendo a generar evidencia útil para orientar políticas públicas en salud materna.

1.3. Pregunta de investigación

¿Cómo se comporta la razón de mortalidad materna (RMM) en los diferentes departamentos y municipios respecto a los factores sociodemográficos en Colombia durante el año 2023?

1.4. Objetivo General

Identificar y caracterizar la relación entre los factores sociodemográficos y la Razón de Mortalidad Materna (RMM) en Colombia durante 2023, considerando las variaciones entre departamentos y municipios a través de un modelo jerárquico multinivel.

1.4.1. Objetivos Específicos

1. Analizar el comportamiento geográfico general de la mortalidad materna en Colombia, con el fin de identificar posibles agrupaciones o patrones territoriales que sugieran áreas de interés para una intervención prioritaria.

2. Modelar la razón de mortalidad materna (RMM) utilizando métodos bayesianos, considerando los distintos niveles presentes en la estructura de los datos e identificando patrones y diferencias significativas tanto dentro como entre los niveles.

3. Identificar factores de riesgo sociodemográficos asociados a la razón de mortalidad materna en el año 2023.

Capítulo 2

Marco Teórico

El marco teórico que se desarrolla a continuación presenta los conceptos necesarios para el entendimiento del desarrollo de este trabajo de grado. Inicialmente, se plantean las definiciones de índice de razón de mortalidad materna (RMM) y los factores sociodemográficos relacionados. Posteriormente, se presentan técnicas de análisis bayesiano de modelos multinivel.

2.1. Marco conceptual

Colombia, oficialmente República de Colombia, es un país soberano situado en la región noroccidental de América del Sur, limitando al norte, con el océano Atlántico; al este, con Venezuela y Brasil; al sur con Perú y Ecuador; y al oeste, con el océano Pacífico y Panamá. Está organizada políticamente en treinta y dos departamentos descentralizados y el Distrito Capital de Bogotá, sede del Gobierno nacional.

Es uno de los países más poblados de América latina y el mundo, poseyendo una población de 52 millones de habitantes hispanohablantes. También se caracteriza por ser un país multicultural y pluriétnico, donde conviven múltiples cosmovisiones de las culturas indígenas asentadas en el país a la llegada de los españoles, la cultura europea (España y de otros lugares de Europa), y las culturas africanas importadas durante la Colonia son la base de la cultura colombiana (Marca País, 2025)

2.1.1. Mortalidad materna y Razón de Mortalidad Materna

La Mortalidad Materna (MM) es definida por la Organización Mundial de la Salud (OMS) como "la muerte de una mujer durante su embarazo, parto o dentro de los 42 días después de su terminación, por cualquier causa relacionada o agravada por el embarazo, parto o puerperio o su atención, pero no por causas accidentales o incidentales" (Instituto Nacional de Salud (INS), 2022b)

Asi mismo la OMS define la razon de mortalidad materna (RMM) como el "número de defunciones maternas ocurridas durante un periodo de tiempo dado por cada cien mil nacidos vivos en el mismo periodo". Este indicador es reconocido como un medida de desarrollo, ya que refleja las diferencias en ingresos, el acceso a los servicios de salud y la desigualdad entre los países y regiones. Así mismo, permite estimar el comportamiento de la mortalidad materna (MM), basado en ciertas características de la madre y su entorno (Instituto Nacional de Salud (INS),

2022b)(Sandoval-Vargas, Y. G. and Eslava-Schmalbach, J. H., 2013)(U.S. Census Bureau, 2022)

$$\text{Se define como: } RMM = \left(\frac{\text{Número de muertes maternas (año)}}{\text{Número de nacidos vivos (año)}} \right) \times 10\,000 \quad (2.1)$$

Otros indicadores para medir la Mortalidad Materna (MM) son la Tasa de Mortalidad Materna (TMM), la Proporción de muertes maternas (PM) y el Riesgo de muerte materna a lo largo de la vida (LTR). No se deben confundir la TMM y la RMM, ya que la RMM mide las muertes maternas por cada 100.000 nacidos vivos, mientras que la TMM se refiere a las muertes maternas por cada 100.000 mujeres en edad reproductiva, mostrando el riesgo general de morir por causas maternas en este grupo (U.S. Census Bureau, 2022)

2.1.2. Factores sociodemográficos

Según la OMS, los factores sociodemográficos son las características individuales y colectivas que influyen en la salud y el bienestar de las personas. Entre estos factores se incluyen la edad, género, raza, nivel de educación, ocupación, ingresos, estructura familiar y otros aspectos que influyen en la salud y el bienestar de las personas

2.2. Marco estadístico

2.2.1. Metodología Delphi

La metodología Delphi es una técnica estructurada de consenso empleada para recopilar, organizar y depurar el juicio de expertos mediante rondas sucesivas de consulta anónima. Según Varela-Ruiz et al., 2012, esta técnica se orienta a abordar problemáticas complejas a través de un proceso iterativo que integra preguntas guiadas, retroalimentación controlada y análisis progresivo de las respuestas con el fin de alcanzar un nivel de acuerdo entre los participantes. Una característica central del método es el anonimato de las respuestas, lo que reduce la influencia de jerarquías o presiones sociales dentro del grupo experto. Su uso se ha extendido en las ciencias de la salud debido a su capacidad para integrar conocimiento especializado sin requerir la presencia física de los participantes y a su utilidad para la identificación y priorización de variables, criterios o escenarios en contextos con información limitada o alta incertidumbre.

2.2.2. Datos Multinivel

El término multinivel hace referencia a una estructura jerárquica o anidada de los datos, en la cual las unidades de análisis se organizan en diferentes niveles. Esta estructura implica que las observaciones dentro de un mismo grupo tienden a ser más similares entre sí que las de grupos diferentes, lo cual viola el supuesto de independencia de los modelos estadísticos tradicionales (Hritcu, R. O. S., 2015)

El enfoque multinivel permite modelar simultáneamente la variabilidad entre y dentro de los niveles de análisis, teniendo en cuenta la estructura jerárquica de los datos. Estas estructuras pueden ser jerárquicas, cuando cada unidad inferior pertenece a una sola unidad superior; cruzadas, cuando se relaciona con más de un nivel

superior; o de membresía múltiple, cuando los individuos pertenecen a varios grupos de forma simultánea (Hritcu, R. O. S., 2015)

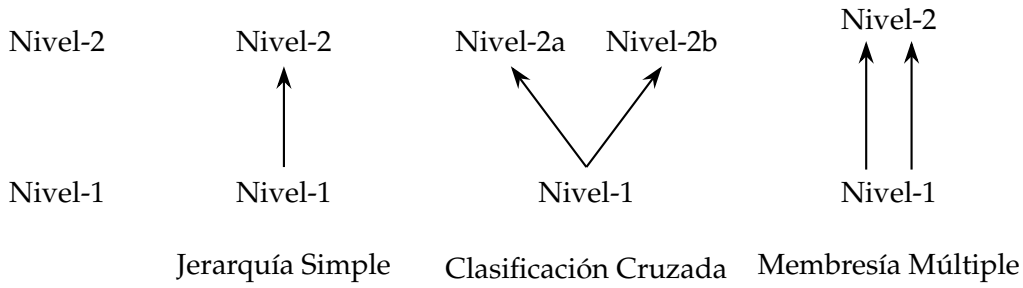

FIGURA 2.1: Estructuras de datos multinivel.

El análisis de datos multinivel se basa en que los fenómenos no pueden explicarse completamente desde un único nivel, sino que requieren considerar la interdependencia estructural entre niveles, permitiendo modelar la variabilidad tanto intra como inter niveles y representar de manera más precisa observaciones anidadas o relacionadas.

2.2.3. Análisis y modelos multinivel

El modelado de datos busca representar un fenómeno real mediante un modelo conceptual que capture sus características esenciales. En este contexto, los modelos multinivel corresponden a una familia de modelos estadísticos diseñados para analizar datos anidados o jerárquicos, en los cuales las observaciones se organizan en distintos niveles. Estos modelos pueden clasificarse según la estructura de los datos y el tipo o la distribución de la variable de respuesta. (Hritcu, R. O. S., 2015)

Los modelos multinivel permiten analizar simultáneamente variables de diferentes niveles, teniendo en cuenta la dependencia entre observaciones dentro de un mismo grupo. Al reconocer que los individuos de un mismo grupo comparten factores comunes y no son independientes, estos modelos evitan sesgos y estimaciones incorrectas que pueden surgir al aplicar métodos tradicionales de un solo nivel, mejorando la precisión de los errores estándar y la validez de las inferencias.(Hritcu, R. O. S., 2015)

2.2.4. Modelo Hurdle Multinivel

El modelo *hurdle* es una familia de modelos de dos partes diseñada para datos con una alta proporción de ceros. Para una variable de respuesta Y_{ij} observada en unidades $i = 1, \ldots, N_j$, anidadas dentro de grupos $j = 1, \ldots, J$, la especificación general separa el proceso que genera ceros del proceso que genera valores positivos, pudiendo incorporar términos jerárquicos para capturar la dependencia entre unidades pertenecientes al mismo grupo (Min, Y. and Agresti, A., 2005) .

Componente binaria

Sea $Z_{ij} = \mathbb{I}(Y_{ij} > 0)$,

el indicador de cruce del umbral. La ocurrencia de ceros se modela mediante:

$$Z_{ij} \mid \pi_{ij} \sim \text{Bernoulli}(\pi_{ij}), \qquad \text{logit}(\pi_{ij}) = \mathbf{w}_{ij}^{\top} \boldsymbol{\beta}^{(b)} + u_{j[i]} + v_i,$$

Donde $\mathbf{w}_{ij}$ es el vector de covariables, $\boldsymbol{\beta}^{(b)}$ los efectos fijos, y $(u_{j[i]}, v_i)$ representan los efectos aleatorios a nivel de grupo y unidad, respectivamente.

Componente continua

Condicional en $Y_{ij} > 0$, la magnitud del resultado se modela mediante una distribución continua truncada en cero. De manera general,

$$Y_{ij} \mid (Y_{ij} > 0) \sim F(\mu_{ij}, \sigma^2),$$

Donde F puede ser una distribución Log-normal, Gamma truncada, Weibull truncada u otra familia continua con soporte en $(0, \infty)$. El predictor lineal asociado es:

$$\mu_{ij} = \mu_0 + \mathbf{w}_{ij}^{\top} \boldsymbol{\beta}^{(g)} + u'_{j[i]} + v'_i.$$

En esta expresión, μ_0 corresponde al intercepto de la parte continua y representa el promedio del logaritmo de Y_{ij} cuando no hay covariables ni efectos aleatorios. El vector $\mathbf{w}_{ij}$ contiene las covariables del municipio i que esta dentro del departamento j, y $\boldsymbol{\beta}^{(g)}$ son los coeficientes que cuantifican su efecto sobre el valor esperado de $\log(Y_{ij})$. El término $u'_{j[i]}$ captura el efecto aleatorio del departamento j al que pertenece el municipio i, reflejando la variabilidad entre departamentos no explicada por las covariables, mientras que v'_i representa el efecto aleatorio específico del municipio, recogiendo la heterogeneidad residual a nivel municipal.

Priors jerárquicas

Los efectos aleatorios se parametrizan mediante una formulación no centrada:

$$u_{j[i]} = \sigma_u \, u_{j[i]}^{(\text{raw})}, \qquad v_i = \sigma_v \, v_i^{(\text{raw})},$$

donde

$$u_{j[i]}^{(\text{raw})}, v_i^{(\text{raw})} \sim N(0, 1)$$

Las desviaciones estándar σ_u y σ_v reciben prioris débiles y regularizadas del tipo Half-Cauchy o Half-Normal, según las recomendaciones para modelos jerárquicos.

Verosimilitud general

La verosimilitud del modelo hurdle se factoriza como:

$$L(\{\pi_i\}, \{\mu_i\}, \psi \mid \mathbf{Z}) = \prod_{i=1}^{N} (1 - \pi_i)^{1 - Z_i} \left[\pi_i \, f_Y(y_i \mid \mu_i, \psi)\right]^{Z_i}. \tag{2.2}$$

Distribución a posteriori

La distribución a posteriori se obtiene mediante:

$$p(\theta \mid \mathbf{y}) \propto L(\theta) \, p(\theta),$$

la cual, debido a la estructura de dos partes, los efectos jerárquicos y las prioris no conjugadas, carece de forma analítica cerrada. Por consiguiente, la inferencia debe realizarse mediante métodos numéricos como Hamiltonian Monte Carlo.

2.2.5. Estimación Bayesiana

La estimación de modelos jerárquicos, como los modelos *hurdle* multinivel utilizados en este trabajo para analizar la Razón de Mortalidad Materna (RMM), requiere métodos capaces de aproximar distribuciones posteriores de alta dimensión de manera eficiente y numéricamente estable. En este contexto, *Stan* se ha consolidado como una de las plataformas más robustas para la inferencia bayesiana moderna. Stan implementa el algoritmo Hamiltonian Monte Carlo (HMC) y su versión adaptativa No-U-Turn Sampler (NUTS), los cuales permiten explorar el espacio de parámetros utilizando gradientes de la log-verosimilitud, mejorando significativamente la eficiencia respecto a métodos MCMC tradicionales (Ghosal, S, et all, 2020)

En términos generales, la inferencia bayesiana consiste en obtener la distribución posterior de los parámetros del modelo θ dada la información observada y, expresada como:

$$p(\theta \mid y) \propto p(y \mid \theta)\, p(\theta) \tag{2.3}$$

donde $p(y \mid \theta)$ es la verosimilitud y $p(\theta)$ representa las distribuciones previas de los parámetros. El desafío computacional radica en que, para modelos con múltiples niveles y estructuras no lineales, esta distribución no puede obtenerse de forma analítica. Stan aborda este problema utilizando dinámica hamiltoniana. Sea

$$U(\theta) = -\log p(\theta \mid y) \tag{2.4}$$

el potencial asociado a la posterior, y sea $r \sim \mathcal{N}(0, M)$ un vector auxiliar de "momentos". El sistema hamiltoniano se define como:

$$H(\theta, r) = U(\theta) + K(r) \tag{2.5}$$

siendo $K(r)$ la energía cinética. La evolución del sistema está dada por:

$$\frac{d\theta}{dt} = \frac{\partial H}{\partial r}, \qquad \frac{dr}{dt} = -\frac{\partial H}{\partial \theta} \tag{2.6}$$

Este mecanismo produce movimientos que evitan caminatas aleatorias, permitiendo saltos amplios en el espacio paramétrico. La variante NUTS ajusta automáticamente la longitud de las trayectorias y el tamaño del paso, sin requerir calibración manual, lo que facilita la exploración de modelos de alta complejidad (Ghosal, S, et all, 2020)

El uso de Stan permite estimar modelos hurdle, que representan de manera conjunta dos procesos estocásticos: un modelo binario para la probabilidad de observar un valor cero y un modelo continuo o discreto truncado para los valores positivos. Según Min y Agresti (2005), este tipo de modelos puede presentar dificultades adicionales cuando existe correlación intra-grupo o estructuras multinivel, lo que hace relevante el uso de una plataforma como Stan para su estimación (Min, Y. and Agresti, A., 2005)

Desde R, la implementación de modelos en Stan se realiza a través de los paquetes `rstan` o `cmdstanr`. El flujo estándar consiste en: (i) preparar la lista de datos, (ii) compilar el archivo `.stan`, y (iii) ejecutar las cadenas MCMC. Posteriormente, es posible extraer distribuciones marginales, intervalos creíbles, diagnósticos de convergencia y predicciones posteriores. Esta integración facilita la incorporación de modelos jerárquicos con efectos aleatorios a distintos niveles, distribuciones truncadas

y prioris flexibles, lo cual es fundamental para el modelo *hurdle* binomial–lognormal utilizado en este estudio.

La estructura de un modelo en Stan se organiza de forma modular en bloques que definen los datos de entrada, los parámetros a estimar, las transformaciones determinísticas, la especificación del modelo mediante distribuciones previas y la verosimilitud, así como la generación de cantidades posteriores como predicciones y métricas. Esta organización permite una formulación clara y flexible de modelos jerárquicos complejos, facilitando su extensión a múltiples niveles de variación y haciendo especialmente adecuada la aplicación de Stan en el análisis bayesiano de la RMM desarrollado en este trabajo.

2.2.6. Análisis bayesiano objetivo

El análisis bayesiano objetivo corresponde a un enfoque dentro de la inferencia bayesiana orientado a la especificación de distribuciones a priori cuando la información previa es limitada o inexistente. Su propósito es definir priors basadas en criterios formales y reproducibles, de modo que la inferencia esté dominada principalmente por la información contenida en los datos (Berger, J., 2006).

En este marco, se emplean distribuciones a priori débilmente informativas o de referencia, construidas a partir de principios como invariancia o máxima entropía, las cuales permiten representar estados iniciales de conocimiento mínimo sin recurrir a procesos de elicitación subjetiva complejos. Este enfoque resulta útil en modelos jerárquicos y de alta dimensión, como los considerados en este estudio, donde el uso de priors regularizadas contribuye a estabilizar la inferencia y a garantizar la reproducibilidad del análisis.

En concordancia con Gelman, A., 2013, capitulo 2.9, en este estudio se utilizan distribuciones a priori débilmente informativas y regularizadas para asegurar que la inferencia posterior sea razonable sin imponer supuestos fuertes. Este enfoque es especialmente pertinente en municipios con baja población y pocos eventos, donde la información es limitada y puede generar estimaciones inestables. La combinación de priors débiles y una estructura jerárquica permite compartir información entre unidades, estabilizar las estimaciones y evitar valores implausibles, respetando las restricciones naturales del fenómeno.

Parametrización no centrada en modelos jerárquicos bayesianos

En los modelos jerárquicos bayesianos, la especificación de los efectos aleatorios constituye un componente central de la estructura del modelo y puede influir en las propiedades de la distribución posterior y del muestreo mediante cadenas de Markov. Una formulación común corresponde a la parametrización centrada, en la cual los efectos aleatorios se modelan directamente como

$$u_j \sim \mathcal{N}(0, \sigma_u^2),$$

donde u_j representa el efecto asociado al grupo j y σ_u^2 es el parámetro de varianza. En este caso, los efectos aleatorios y su parámetro de dispersión están definidos conjuntamente en el mismo nivel del modelo.

Una formulación alternativa es la parametrización no centrada (*non-centered parameterization*), en la cual los efectos aleatorios se expresan mediante una reparametrización del modelo jerárquico. Específicamente, se introducen variables latentes z_j

tales que

$$z_j \sim \mathcal{N}(0,1), \qquad u_j = \sigma_u \, z_j.$$

Esta representación es algebraicamente equivalente a la parametrización centrada, pero separa explícitamente la variabilidad asociada a los efectos aleatorios de su escala, permitiendo una formulación alternativa de la distribución conjunta posterior (Gelman, A., 2013).

En el contexto de la inferencia bayesiana mediante simulación MCMC, la parametrización no centrada se utiliza con el propósito de modificar la dependencia posterior entre los efectos aleatorios y los parámetros de varianza. Esta reexpresión resulta especialmente relevante en modelos jerárquicos con múltiples niveles o con información limitada a nivel de grupo, como aquellos estimados mediante algoritmos de tipo Hamiltonian Monte Carlo, en los cuales la estructura geométrica del espacio paramétrico condiciona el comportamiento del muestreo.

Fundamentos de los métodos MCMC

Los métodos de Monte Carlo vía Cadenas de Markov (MCMC) constituyen la herramienta fundamental para aproximar distribuciones posteriores cuando la inferencia analítica es inalcanzable. Su principio consiste en construir una cadena de Markov que tenga como distribución estacionaria la posterior $p(\theta \mid y)$, de modo que, tras un número suficiente de iteraciones, las muestras generadas pueden utilizarse para aproximar medias, varianzas e intervalos creíbles.

Algoritmo de Metropolis–Hastings. El algoritmo Metropolis–Hastings (MH) es uno de los métodos más generales para obtener muestras desde la posterior. A partir de un valor actual $\theta^{(t)}$, se propone un nuevo valor θ^* desde una distribución de propuesta $q(\theta^* \mid \theta^{(t)})$ y se acepta con probabilidad:

$$\alpha = \min \left\{ 1, \frac{p(\theta^* \mid y)\, q(\theta^{(t)} \mid \theta^*)}{p(\theta^{(t)} \mid y)\, q(\theta^* \mid \theta^{(t)})} \right\} \tag{2.7}$$

Este procedimiento garantiza la reversibilidad y asegura que la cadena converge hacia la distribución posterior (Ros, D. P., 2022). Sin embargo, MH es sensible a la escala y forma de la posterior, lo que puede causar problemas de mezcla lenta y alta autocorrelación, especialmente en modelos jerárquicos de alta dimensión.

Método de Gibbs Sampling. El muestreo de Gibbs es un caso particular de MCMC en el que cada parámetro se actualiza desde su distribución condicional completa. Si éstas son conocidas y fáciles de simular, Gibbs puede ser eficiente; sin embargo, su aplicabilidad depende de que exista conjugación entre prior y verosimilitud. En modelos no conjugados o con variables truncadas, las condicionales completas pueden ser intratables, lo cual motiva métodos más avanzados como HMC (Ros, D. P., 2022)

Métodos avanzados: Hamiltonian Monte Carlo y NUTS

El método *Hamiltonian Monte Carlo* (HMC) introduce un vector auxiliar de momentum $\omega \in \mathbb{R}^d$ y define el estado del sistema como (x, ω), permitiendo interpretar el muestreo como la simulación de un sistema dinámico gobernado por las ecuaciones de Hamilton (Chen, 2019; Carpenter, B, et all, 2017). Esta formulación mejora

la exploración del espacio paramétrico y reduce la autocorrelación mediante trayectorias guiadas por gradientes. El Hamiltoniano combina energía potencial $U(x)$ y cinética $V(w)$:

$$H(x,\omega) = U(x) + V(\omega) = -\log \pi(x) + \sum_{j=1}^{d} \frac{\omega_j^2}{2m_j} \tag{2.8}$$

donde $\pi(x)$ es la densidad objetivo, d es la dimensión de x, y m_j es la masa de la j-ésima coordenada. Esta elección induce que el vector de momentum se genere a partir de una distribución normal multivariante independiente, $\omega \sim N(0, M)$, con $M = \text{diag}(m_1, \ldots, m_d)$ (Chen, 2019).

La evolución del sistema sigue las ecuaciones de movimiento Hamiltonianas:

$$\frac{dx_j(t)}{dt} = \frac{\partial H(x(t),\omega(t))}{\partial \omega_j}, \qquad \frac{d\omega_j(t)}{dt} = -\frac{\partial H(x(t),\omega(t))}{\partial x_j}, \quad j = 1, \ldots, d,$$

Generan trayectorias deterministas que conservan aproximadamente la energía y permiten propuestas con alta probabilidad de aceptación .Dado que estas ecuaciones no suelen tener solución analítica, su resolución se aproxima mediante el integrador *leapfrog*, caracterizado por un tamaño de paso ϵ y L iteraciones, cuyas propiedades de reversibilidad y conservación del Hamiltoniano garantizan la validez del muestreo. (Chen, 2019).

El algoritmo HMC se resume en los siguientes pasos: (1) a partir del estado actual x_0 se genera un momentum inicial $\omega_0 \sim N(0, M)$; (2) se simula la dinámica Hamiltoniana desde (x_0, ω_0) utilizando el integrador *leapfrog* durante L pasos; (3) se obtiene un estado candidato (x^*, ω^*); (4) dicho candidato se acepta con probabilidad

$$\alpha = \text{mín}\left\{1, \frac{\exp\{-H(x^*,\omega^*)\}}{\exp\{-H(x_0,\omega_0)\}}\right\};$$

y (5) en caso de aceptación se actualiza $x = x^*$, mientras que de lo contrario se conserva $x = x_0$.

La introducción de trayectorias guiadas por gradientes permite que HMC explore de manera eficiente distribuciones posteriores de alta dimensión, superando las limitaciones de los esquemas de caminata aleatoria propios de los métodos MCMC tradicionales (Chen, 2019). Por otra parte, a diferencia de Metropolis–Hastings, donde las propuestas son locales, HMC utiliza integradores de tipo *leapfrog* para recorrer trayectorias largas y coherentes, lo que incrementa la probabilidad de aceptación y mejora la eficiencia computacional.

No-U-Turn Sampler (NUTS). El algoritmo NUTS es una extensión adaptativa de HMC que elimina la necesidad de predefinir el número de pasos de leapfrog. Según Carpenter, B, et all, 2017, NUTS construye recursivamente trayectorias en direcciones opuestas hasta detectar un "retroceso" o *U-turn*, garantizando el equilibrio detallado sin intervención manual. De esta forma, NUTS:

- Ajusta automáticamente la longitud de la trayectoria.
- Adapta el tamaño del paso durante la fase de calentamiento.
- Produce cadenas con menor autocorrelación.

- Resulta robusto para modelos con estructuras jerárquicas complejas.

2.2.7. Modelos Gaussianos Latentes (MGLs)

Desde una perspectiva bayesiana, muchos modelos jerárquicos utilizados en epidemiología y estadística espacial pueden formularse como Modelos Gaussianos Latentes (MGLs). En esta clase de modelos, la estructura de dependencia entre observaciones se introduce mediante un conjunto de variables no observadas, denominadas campo latente, que suelen incluir efectos aleatorios, componentes espaciales o temporales y términos de heterogeneidad no observada.

De manera general, un MGL se compone de tres niveles: (i) una distribución de probabilidad para las observaciones condicionada a un predictor latente, (ii) un predictor latente definido como una combinación lineal de covariables observadas y efectos latentes, y (iii) un conjunto reducido de hiperparámetros que controlan la variabilidad y dependencia de los efectos latentes. En este contexto, n denota el número de observaciones, el campo latente agrupa los efectos no observados del modelo, y el predictor latente actúa como el vínculo entre dichos efectos y la distribución de los datos.

Esta formulación es particularmente adecuada para modelos jerárquicos multinivel, como el modelo hurdle propuesto en este estudio, donde la dependencia entre unidades territoriales se introduce mediante efectos aleatorios estructurados. (Wang, X. and Yue, Y. R. and Faraway, J. J., 2018)

2.2.8. Campos aleatorios de Markov gaussianos (GMRF)

Para lograr eficiencia computacional en modelos gaussianos latentes de gran dimensión, es habitual asumir que el campo latente sigue una estructura de Campo Aleatorio de Markov Gaussiano (GMRF). Un GMRF es un vector gaussiano multivariado que satisface la propiedad de Markov, según la cual cada componente es condicionalmente independiente del resto dado un conjunto reducido de vecinos.

Esta propiedad implica que la matriz de precisión asociada al campo latente es dispersa, lo que reduce de manera sustancial los costos computacionales en la inferencia bayesiana. La formulación mediante GMRFs resulta especialmente adecuada en modelos jerárquicos y espaciales, donde la dependencia entre unidades se limita a estructuras locales o a niveles administrativos específicos. (Wang, X. and Yue, Y. R. and Faraway, J. J., 2018)

2.2.9. Evaluación del modelo y selección del modelo

Según Jabessa y Jabessa (2021), citando a Bewick, Cheek y Ball (2005), la evaluación y selección del modelo consiste en identificar el modelo que mejor se ajusta a los datos entre varias alternativas, mediante el análisis comparativo de su desempeño y capacidad predictiva. Para comparar diferentes modelos, están disponibles dos criterios en INLA: el criterio de información de desviación (DIC) y el criterio de información de Watanabe-Akaike (WAIC) (Wierzchoslawska, J., 2023)(Wang, X. and Yue, Y. R. and Faraway, J. J., 2018)(Gómez-Rubio, Virgilio, 2020)

Criterios basados en la información (DIC y WAIC)

El Criterio de Información de la Desviación (DIC), propuesto por Spiegelhalter et al. (2002), es un criterio de selección de modelos que combina la bondad de ajuste con

un término de penalización por complejidad, representado por el número efectivo de parámetros (Wang, X. and Yue, Y. R. and Faraway, J. J., 2018). Su expresión general es:

$$DIC = D(\hat{x}, \hat{\theta}) + 2p_D \tag{2.9}$$

Donde $D(\cdot)$ es la desviación, $\hat{x}$ y $\hat{\theta}$ son las expectativas posteriores de los efectos latentes y de los hiperparámetros, respectivamente, y p_D corresponde al número efectivo de parámetros, definido como:

$$p_D = E[D(\cdot)] - D(\hat{x}, \hat{\theta}) \tag{2.10}$$

Los modelos con menor DIC se consideran más adecuados, ya que logran un equilibrio entre ajuste y complejidad. El Criterio de Información de Watanabe-Akaike (WAIC) es similar al DIC, pero calcula el número efectivo de parámetros de manera diferente, considerando la variabilidad de la verosimilitud de cada observación:

Transformación integral predictiva (PIT)

La transformación integral predictiva (PIT; Marshall y Spiegelhalter, 2003) se utiliza para evaluar la calibración del modelo predictivo, es decir, verificar si las predicciones son coherentes con los datos observados (Gómez-Rubio, Virgilio, 2020). Mide, para cada observación, la probabilidad de que un nuevo valor sea menor o igual al observado:

$$PIT_{Y_o} = \pi(y_{Y_o}^{new} \leq y_{Y_o} \mid y_{-Y_O}) \tag{2.11}$$

Para datos discretos, se emplea la versión ajustada:

$$PIT_{Y_o}^{adjusted} = PIT_{Y_o} - 0.5 \times CPO_{Y_o} \tag{2.12}$$

Si los datos son discretos, puede ocurrir que $y_{Y_o}^{new}$ sea igual y_{Y_o}; en estos casos, solo se cuenta como la mitad, pues así se evita sobrestimar la probabilidad acumulada. Si el modelo se ajusta bien, los valores de PIT deben aproximarse a una distribución uniforme entre 0 y 1.

Factor de reducción de escala potencial ($\hat{R}$)

Cuando se ajustan modelos mediante métodos MCMC, es fundamental verificar la convergencia de las cadenas para asegurar que las estimaciones sean válidas. Para ello, se utiliza el factor de reducción de escala potencial, $\hat{R}$, introducido por Gelman y Rubin (1992) (Vehtari, A, et all, 2021)

Sea θ el parámetro de interés cuya distribución posterior, $p(\theta|y)$, se está estimando. Cada cadena MCMC genera un conjunto de muestras de θ, denotadas como $\theta^{(nm)}$, donde n indica la n-ésima extracción de la m-ésima cadena (Vehtari, A, et all, 2021) . A partir de estas muestras se calculan las estadísticas necesarias para evaluar la convergencia.

La varianza entre cadenas (B), mide la variabilidad entre las medias de las cadenas MCMC, es decir, cuánto difieren las medias de cada cadena respecto al promedio general de todas las cadenas.

$$B = N\frac{1}{M-1}\sum_{m=1}^{M}(\bar{\theta}^{(.m)} - \bar{\theta}^{(..)})^2, \text{ donde } \bar{\theta}^{(.m)} = \frac{1}{N}\sum_{n=1}^{N}\theta^{(nm)},\ \bar{\theta}^{(..)} = \frac{1}{M}\sum_{m=1}^{M}\bar{\theta}^{(.m)} \tag{2.13}$$

- $\bar{\theta}^{(.m)}$: El promedio de las N extracciones de la m-ésima cadena.
- $\bar{\theta}^{(..)}$: El promedio de todas las cadenas

La varianza dentro de las cadenas (W) se utiliza para evaluar la variabilidad de las muestras dentro de cada cadena MCMC. Se calcula como el promedio de las varianzas de cada cadena,

$$W = \frac{1}{M}\sum_{m=1}^{M}s_m^2, \quad s_m^2 = \frac{1}{N-1}\sum_{n=1}^{N}(\theta^{(nm)} - \bar{\theta}^{(.m)})^2 \tag{2.14}$$

- s_m^2: La varianza de las N extracciones dentro de la m-ésima cadena.

La varianza marginal posterior del estimador se calcula como un promedio ponderado de la varianza intra-cadena (W) y de la varianza inter-cadenas (B), considerando el número de extracciones por cadena (N).

$$\hat{var}^{+}(\theta|y) = \frac{N-1}{N}W + \frac{1}{N}B \tag{2.15}$$

El factor de reducción de escala potencial, $\hat{R}$, se utiliza para evaluar la convergencia de las cadenas MCMC, comparando la variabilidad intra-cadena con la variabilidad entre cadenas.

$$\hat{R} = \frac{\sqrt{\hat{var}^{+}(\theta|y)}}{\sqrt{W}} \tag{2.16}$$

Si $\hat{R}$ se aproxima a 1, indica que las cadenas han convergido y los resultados son confiables; valores mayores a 1 sugieren que la convergencia aún no se ha alcanzado.

Probability of Direction (pd)

La Probabilidad de Dirección (pd), también llamada Probabilidad Máxima del Efecto (MPE, por sus siglas en inglés), representa la probabilidad de que un parámetro, descrito por su distribución posterior, sea estrictamente positivo o negativo, según cuál sea más probable. (Makowski et al., 2019)

Este índice varía entre 50 % y 100 %: valores cercanos a 100 % indican alta certeza sobre la dirección del efecto, mientras que val ores cercanos a 50 % reflejan incertidumbre respecto a si el efecto es positivo o negativo.

El pd es independiente del modelo, ya que se calcula únicamente a partir de la distribución posterior y no requiere información adicional de los datos. Además, es robusto frente a la escala de la variable respuesta y de los predictores, y no se ve afectado por la significancia del efecto, y por lo tanto no es útil para evaluar el tamaño o la relevancia de un efecto.(Makowski et al., 2019)

Intervalo creible

Los Intervalos Creíbles Bayesianos (BCIs) constituyen el análogo bayesiano de los intervalos de confianza y se definen como rangos dentro de los cuales un parámetro se encuentra con una probabilidad posterior especificada, dados los datos

observados y la información previa. A diferencia de los intervalos de confianza, cuya interpretación se fundamenta en muestreos hipotéticos repetidos, los BCIs permiten una interpretación probabilística directa (Lovric, 2025).

Formalmente, si θ representa el parámetro de interés y $[a, b]$ es el intervalo creíble del 95 %, se cumple que:

$$Pr(\theta \in [a, b] \mid data) = 0.95$$

En el enfoque bayesiano, los parámetros del modelo se tratan como variables aleatorias, lo que permite describir la incertidumbre asociada a estos mediante distribuciones de probabilidad posteriores. En este contexto, los intervalos creíbles se definen como regiones del espacio paramétrico que contienen una proporción determinada de la distribución posterior del parámetro de interés. La construcción de estos intervalos incorpora información proveniente de una distribución a priori, la cual se combina con la información aportada por los datos observados a través del teorema de Bayes, incluso cuando dicha distribución a priori es no informativa (Lovric, 2025).

Un intervalo creíble del 95 % se interpreta como una región que, dados los datos observados y el modelo especificado, contiene al parámetro de interés con una probabilidad del 95 %. Esta interpretación probabilística directa constituye una característica central de los intervalos creíbles en el marco bayesiano.(VanderPlas, 2014).

Entre los intervalos creíbles, se encuentra el basado en cuantiles de colas iguales. Se define como

$$CR_\alpha = \left[q_{(1-\alpha)/2},\ q_{(1+\alpha)/2}\right]$$

donde $q_{(1-\alpha)/2}$ y $q_{(1+\alpha)/2}$ son los cuantiles de la distribución posterior y α es el nivel de credibilidad del intervalo.

Tamaño efectivo de muestra en simulaciones MCMC

En los modelos bayesianos estimados mediante simulación por cadenas de Markov, el número total de iteraciones generadas no coincide necesariamente con la cantidad de información independiente disponible, debido a la autocorrelación entre muestras consecutivas. Para abordar este aspecto, se emplea el tamaño efectivo de muestra (*effective sample size*, ESS), el cual mide cuántas observaciones independientes equivalentes contiene la muestra posterior generada por el algoritmo MCMC (Gelman, A., 2013).

El tamaño efectivo de muestra se calcula a partir de la autocorrelación de las cadenas MCMC. Siguiendo a Gelman, A. (2013, Sección 11.5), el ESS se define como

$$ESS = \frac{mn}{1 + 2\sum_{t=1}^{\infty} \rho_t},$$

donde m es el número de cadenas, n el número de iteraciones por cadena y ρ_t la autocorrelación de la cadena en el rezago t. Esta expresión surge del análisis asintótico de la varianza del promedio de simulaciones correlacionadas y cuantifica la pérdida de información efectiva causada por la dependencia serial entre iteraciones consecutivas.

La definición y justificación teórica se presentan explícitamente en el Capítulo 11, Sección 11.5 (*Effective number of simulation draws*) de *Bayesian Data Analysis* (Gelman, A., 2013).

Un valor elevado del ESS $(400 <)$ indica que las cadenas exploran de manera eficiente la distribución posterior y que las estimaciones obtenidas presentan menor variabilidad Monte Carlo. Por el contrario, un ESS bajo sugiere alta dependencia entre iteraciones y, por tanto, una menor precisión en la inferencia posterior, aun cuando el número total de simulaciones sea grande.

En aplicaciones modernas se distinguen dos medidas complementarias del tamaño efectivo de muestra: ESS_{bulk} y ESS_{tail}. El primero evalúa la eficiencia del muestreo en la región central de la distribución posterior, que es relevante para la estimación de medidas de tendencia central. El segundo se enfoca en la precisión del muestreo en las colas de la distribución, aspecto fundamental para la correcta estimación de cuantiles e intervalos creíbles. La consideración conjunta de ambos estadísticos resulta especialmente importante en modelos bayesianos jerárquicos complejos, donde la inferencia depende tanto del comportamiento central como de los extremos de la distribución posterior (Min, Y. and Agresti, A., 2005; Gelman, A., 2013).

2.2.10. Autocorrelación espacial

La autocorrelación espacial (AE) se concibe como un concepto central en el análisis y la comprensión de los fenómenos espaciales, derivado de la Ley de Tobler, la cual establece que "todas las cosas están relacionadas entre sí, pero las cosas más próximas en el espacio tienen una relación mayor que las distantes". En términos estadísticos, la AE proviene del concepto de correlación, que mide la relación entre dos variables aleatorias; en este caso, la correlación se calcula entre una misma variable medida en distintas localizaciones espaciales (Ramírez, M. L., 2016)

La autocorrealación espacial (AE) se divide en dos dimensiones. La primera dimensión divide la AE en vecindad y distancia. En el enfoque de vecindad, es necesario ajustar los objetos espaciales en función de su entorno, definiendo cómo se conectan entre sí (topología) y asignando ponderaciones a cada observación. En contraste, el enfoque basado en distancia calcula un indicador que refleja la proximidad entre un objeto espacial y los demás con los que se encuentra relacionado. La segunda dimensión divide la autocorrealación espacial (AE) en en asociación global y local. La autocorrelación espacial global se emplea para analizar las interacciones entre todos los elementos del conjunto de datos, mientras que la autocorrelación local se centra en examinar los patrones alrededor de cada objeto individual, evaluando hasta qué punto estos reflejan la tendencia observada a nivel global. (Ramírez, M. L., 2016)

Entre los procedimientos que pueden aplicarse para el análisis de la autocorrelación espacial se destacan: (a) el análisis de clúster y detección de valores atípicos mediante el Índice de Moran (global) y el estadístico local de Anselin, y (b) el análisis de *hotspots* de Getis & Ord, que permite identificar concentraciones significativas de valores altos o bajos. Estos métodos permiten evaluar patrones de asociación espacial y determinar la existencia de agrupamientos o disparidades en la distribución de una variable geográfica. (Ramírez, M. L., 2016)

Índice de Moran e Índice General de Getis & Ordon

El Índice de Moran (I) constituye una medida de autocorrelación espacial utilizada para analizar la intensidad y la dirección de la dependencia espacial entre las observaciones. A través de este índice es posible identificar la presencia de agrupamientos espaciales y determinar si los valores próximos tienden a asemejarse o a diferenciarse. Según Ramírez, M. L., 2016, estos índices permiten evaluar el patrón espacial y la tendencia general de los datos.

Patrick Moran fue uno de los principales exponentes en el estudio de la Autocorrelación Espacial (AE), quien, con el propósito de obtener una respuesta empírica, expuso la siguiente notación:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}\, k_i\, k_j}{\sum_{i=1}^{n} k_i^2} \tag{2.17}$$

- n: Número total de unidades espaciales consideradas.
- $k_i = (y_i - \bar{y})$: Desviación del valor observado y_i respecto a la media global $\bar{y}$.
- w_{ij}: Ponderación espacial o medida de proximidad entre las unidades i y j.
- $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$: suma total de las ponderaciones espaciales.

La matriz de ponderaciones espaciales $W = [w_{ij}]$ define la relación de proximidad entre las unidades espaciales. Toma el valor $w_{ij} = 1$ cuando las regiones i y j son vecinas o mantienen interacción espacial directa, y $w_{ij} = 0$ cuando no existe tal relación. Esta matriz permite representar la estructura de conexión entre las regiones dentro del área de estudio. Asi, el coeficiente de Moran (I) toma valores comprendidos entre -1 y $+1$. Un valor positivo indica una autocorrelación espacial directa, es decir, que las unidades con valores similares tienden a agruparse (alta con alta o baja con baja). En cambio, un valor negativo refleja una autocorrelación inversa, donde valores altos se encuentran próximos a valores bajos. Cuando I es cercano a cero, no existe una estructura espacial significativa, y los valores se distribuyen de manera aleatoria en el espacio.

Por su parte, Arthur Getis y J.K. Ord, citados en Ramírez, M. L., 2016, presentaron su índice general, el cual mide la concentración de valores altos o bajos en una determinada área de estudio, de la siguiente manera:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}\, y_i\, y_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} y_i\, y_j}, \quad \forall j \neq i \tag{2.18}$$

- x_i y x_j: Valores del atributo para las unidades espaciales i y j
- w_{ij}: Ponderación espacial entre las unidades espaciales i y j
- n: Número total de unidades espaciales del conjunto de datos

Al obtener estos índices globales se generan un conjunto de cinco valores: el índice observado, el índice esperado, la varianza, la puntuación z y el valor p. Estos valores permiten evaluar si el patrón espacial detectado se debe al azar o si refleja una distribución significativa de los datos en el espacio.

El índice observado es el valor calculado a partir de los datos reales, que refleja el patrón espacial en el área de estudio. El índice esperado es el valor teórico que se

obtendría si el patrón fuera aleatorio, sin ninguna estructura espacial significativa. La varianza mide la dispersión de los datos, indicando cuánto varían los valores observados respecto al índice esperado. Juntos,

El valor p representa la probabilidad de que el patrón observado pueda explicarse únicamente por procesos aleatorios. Por lo tanto, un valor p muy pequeño indica que es poco probable que el patrón se deba al azar, lo que permite rechazar la hipótesis nula, la cual plantea que existe aleatoriedad espacial completa, tanto en la distribución de las entidades como en los valores asociados a ellas. Una vez rechazada la hipótesis nula, el signo de la puntuación z se vuelve relevante: una puntuación z positiva indica concentración de valores altos en el área de estudio, mientras que una puntuación z negativa señala concentración de valores bajos. Por su parte, una puntuación z cercana a cero sugiere que no existe un patrón de concentración claro en los valores del área analizada (Ramírez, M. L., 2016)

También se puede aplicar índices de autocorrelación espacial a nivel local. Uno de los más conocidos es el Índice Local de Anselin (LISA), propuesto por Luc Anselin en 1995, basado en el índice global de Moran que se encarga de identificar los agrupamientos (clusters) espaciales de entidades que poseen valores similares. Por su parte, el índice local de Getis & Ord (Gi), mide el grado de agrupamiento o clustering para valores altos o bajos, es decir dado un conjunto de entidades ponderadas, identifica puntos calientes (valores altos) y puntos fríos (valores bajos) estadísticamente significativos (Ramírez, M. L., 2016)

2.2.11. Inflación de ceros

En datos de conteo asociados a eventos poco frecuentes es habitual observar una proporción de valores nulos superior a la esperada bajo modelos clásicos, fenómeno conocido como inflación de ceros. Esta característica motiva el uso de modelos de dos partes que separan el proceso que determina la ocurrencia del evento del que describe la magnitud del conteo positivo.

En este estudio se adopta el modelo *hurdle*, el cual asume que los ceros se generan a partir de un proceso binario que determina si el evento ocurre o no, mientras que las observaciones positivas se modelan mediante una distribución de conteo truncada en cero (Min, Y. and Agresti, A., 2005; Zeileis, A. and Kleiber, C. and Jackman, S., 2008). Existen enfoques alternativos para el manejo de inflación de ceros, como los modelos inflados de ceros, los cuales no se consideran en este trabajo dado el objetivo metodológico planteado.

2.2.12. Análisis de Componentes Principales (ACP)

El Análisis de Componentes Principales (ACP) es una de las técnicas más importantes y ampliamente utilizadas del análisis multivariado aplicado a matrices de datos bidimensionales (dos vías). Fue introducido por Karl Pearson (1901) y formalizado posteriormente por Harold Hotelling (1933) como una técnica estadística matemática (Montes Mora, C. L., 2024)

El Análisis de Componentes Principales (ACP) es una técnica estadística cuyo objetivo es reducir la dimensionalidad de un conjunto de datos, conservando la mayor parte de la variabilidad original. Para ello, transforma un conjunto de variables cuantitativas posiblemente correlacionadas en un nuevo conjunto de variables no correlacionadas, denominadas componentes principales, ordenadas de manera que

las primeras concentran la mayor proporción de la varianza total. El ACP se aplica a una matriz de datos de dimensión $n \times p$, compuesta por n individuos y p variables continuas, y permite obtener una representación de los datos en un espacio de menor dimensión que facilita la interpretación de las relaciones entre individuos y variables.

Dado que las variables pueden estar en distintas escalas, se estandarizan para que cada una tenga media cero y desviación estándar uno, evitando que alguna domine el análisis. La estandarización se realiza por columnas mediante:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \; Donde: \;\; \bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \;\; s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2, \;\; s_j = \sqrt{s_j^2} \quad (2.19)$$

La matriz resultante Z contiene las puntuaciones estandarizadas de los datos. Para la correcta aplicación del ACP se requiere que las variables sean continuas, que la matriz X no presente datos faltantes y que exista una relación lineal entre las variables.

Según el criterio de Kaiser, el número de componentes principales se determina conservando únicamente aquellos con autovalores mayores que 1, dado que estos explican una proporción de variabilidad superior a la de una variable estandarizada individual.

Análisis de la nube de individuos $(\mathbb{R}^p)$

La nube de individuos en el ACP representa a los n individuos en el espacio definido por las variables medidas, proyectándolos en un subespacio de dos dimensiones. Su análisis se basa en medir la proximidad entre individuos mediante la distancia euclidiana:

$$d^2(i, i') = \sum_{j=1}^{p} m_j \, (x_{ij} - x_{i'j})^2$$

Donde m_j es la métrica que pondera la influencia de cada variable j.Esta distancia permite cuantificar la similitud o diferencia entre los perfiles de los individuos en las p variables consideradas.

Las coordenadas de los individuos en el espacio factorial se obtienen mediante $(\psi_\alpha = Z, u_\alpha)$, donde (Z) es la matriz de datos estandarizados y (u_α) corresponde al eje ortogonal asociado al componente (α). Cada componente se obtiene como una combinación lineal de las variables estandarizadas (Z_j), dada por

$$\begin{aligned}
\psi_1 &= u_{11}Z_1 + u_{12}Z_2 + \cdots + u_{1p}Z_p \; (Primer\ componente) \\
\psi_2 &= u_{21}Z_1 + u_{22}Z_2 + \cdots + u_{2p}Z_p \; (Segundo\ componente) \\
&\vdots \\
\psi_p &= u_{p1}Z_1 + u_{p2}Z_2 + \cdots + u_{pp}Z_p \; (P - esimo\ componente)
\end{aligned}$$

De esta forma, la nube de individuos queda representada en un sistema de ejes que maximiza la variabilidad explicada, preservando las distancias originales y permitiendo interpretar la proximidad entre individuos según sus perfiles en las (p) variables.

Análisis de la nube de variables $(\mathbb{R}^n)$

En la nube de variables, cada variable se representa en el espacio $(\mathbb{R}^n)$ y, al estar estandarizadas, todas tienen distancia 1 al origen. Esto hace que se distribuyan sobre una hiperesfera de radio 1. El objetivo es encontrar el plano factorial que concentre la mayor parte de la información sobre las correlaciones entre variables; este plano se conoce como el círculo de correlaciones.

La distancia entre dos variables se expresa como:

$$d^2(j, j') = \sum_{i=1}^{p} (z_{ij} - z_{ij'})^2$$

De esta expresión, según Lebart et al. (1995), se obtiene que

$$\sum_{i=1}^{p} z_{ij}^2 = \sum_{i=1}^{p} z_{ij'}^2 = 1, \quad \text{y} \quad \sum_{i=1}^{p} z_{ij} z_{ij'} = c_{jj'}$$

Donde $c_{jj'}$ es el coeficiente de correlación entre j y j'. Esta relación permite interpretar la proximidad entre variables: cuando $c_{jj'} \approx 1$, las variables están fuertemente correlacionadas y en el círculo de correlaciones se representan con vectores casi alineados; si $c_{jj'} = 0$, no presentan correlación y sus vectores forman ángulos cercanos a 90°; y cuando $c_{jj'} = -1$, la correlación es negativa y los vectores se ubican en direcciones opuestas dentro del círculo.

Representación Simultánea

En el ACP, las nubes de individuos y de variables se definen en espacios distintos: la de individuos en $(\mathbb{R}^p)$ con ejes (u), y la de variables en $(\mathbb{R}^n)$ con ejes (v). Por ello no pueden representarse directamente en un mismo espacio. Sin embargo, es posible una representación simultánea proyectando las direcciones de las variables del gráfico de correlaciones sobre el plano factorial de los individuos dentro del espacio $(\mathbb{R}^p)$.

2.3. Antecedentes

La mortalidad materna, en el ámbito internacional, ha sido abordada mediante diferentes metodologías. Entre los estudios analizados se encuentra uno realizado en Etiopía, donde se empleó un modelo logístico multinivel bayesiano con el propósito de explorar los factores demográficos y socioeconómicos determinantes en la mortalidad materna. Este enfoque permitio modelar las variaciones existentes a nivel individual y regional, destacando factores significativos como la edad materna, el estado civil, el número de hijos vivos y el nivel educativo. Además, se hace uso de las técnicas de Monte Carlo en cadenas de Markov que permitieron optimizar el ajuste del modelo, identificando diferencias significativas entre regiones y destacando la relevancia de los determinantes sociodemográficos y económicos (Jabessa y Jabessa, 2021)

Otro estudio relevante se llevó a cabo en Brasil, donde se realizó un análisis espacial y temporal para identificar patrones y áreas de alta incidencia en la mortalidad materna. Hace uso de métodos como el índice global de Moran y el análisis local LISA (Local Indicators of Spatial Association) para establecer conglomerados espaciales por medio de indicadores como el PIB, el Índice de Gini, el Índice de Desarrollo

Humano Municipal (IDHM) y la esperanza de vida al nacer, puesto que están asociadas significativamente con mayores niveles de mortalidad (Oliveira, I. V. G, et all, 2023)

De manera complementaria, otro estudio realizado en Brasil también emplea un análisis espacial, incluyendo estadísticas bayesianas empíricas locales para identificar clusters espaciales de alta mortalidad. Este agrupamiento reveló que algunos conglomerados (regiones) presentaban un mayor riesgo relativo de mortalidad materna, provocado por los factores socioeconómicos como el ingreso per cápita y el nivel educativo.

Un estudio reciente desarrollado en Kenia analizó las desigualdades territoriales en indicadores reproductivos, maternos y de salud infantil empleando un modelo jerárquico bayesiano espaciotemporal. Utilizando información mensual del sistema rutinario DHIS2 entre 2018 y 2021 para 47 condados y 290 sub-condados, los autores modelaron simultáneamente cinco indicadores, incluida la razón de mortalidad materna institucional, mediante efectos espaciales estructurados y no estructurados que capturan variación geográfica y temporal. El análisis mostró disparidades marcadas entre regiones, con resultados consistentemente desfavorables en territorios del noreste y la costa, y destacó el papel de determinantes como la densidad de establecimientos de salud, la educación materna y el nivel de riqueza. Adicionalmente, el estudio construyó un índice compuesto de desempeño en salud materno-infantil que permitió identificar desigualdades subnacionales no visibles en los promedios nacionales. Este abordaje bayesiano resalta la utilidad de los modelos jerárquicos para analizar la variabilidad territorial en salud materna utilizando datos rutinarios y ofrece un referente metodológico comparable al análisis multinivel propuesto en el presente estudio (Karimi et al., 2025).

En el ámbito de los modelos jerárquicos bayesianos aplicados a datos de conteo con exceso de ceros y dependencia espacial-temporal, Ghosal et al. desarrollan un modelo *hurdle* mixto que incorpora componentes aleatorios estructurados y no estructurados para capturar heterogeneidad espacial y temporal. La formulación separa la probabilidad de ocurrencia del evento y la distribución de los valores positivos mediante una parte binaria y una distribución negativa binomial truncada. El modelo incluye efectos aleatorios a nivel estatal y de condado, así como términos de suavizamiento espacial mediante *thin plate splines*. La estimación se realiza en Stan utilizando el algoritmo NUTS, empleando aprioris débiles para los efectos fijos y componentes de varianza (Ghosal, S, et all, 2020)

Asimismo, Monnahan (2024) presenta una síntesis del flujo de trabajo bayesiano moderno aplicado a modelos jerárquicos complejos, destacando componentes fundamentales para la formulación, validación y comparación de modelos en contextos de datos con alta variabilidad. El autor expone el uso del algoritmo NUTS para la obtención de muestras eficientes, así como la necesidad de evaluar la convergencia de cadenas y el tamaño efectivo de muestra. El trabajo describe herramientas para la construcción y evaluación de distribuciones a priori mediante *prior predictive checks* y resalta el uso de *posterior predictive checks* como mecanismo de verificación de la concordancia del modelo con los datos observados. También se abordan métodos de comparación de modelos como WAIC y PSIS-LOO, junto con consideraciones sobre estructuras jerárquicas y la incorporación de efectos aleatorios en un enfoque plenamente bayesiano (Monnahan, C. C., 2024)

Los antecedentes mencionados muestran la diversidad de metodologías empleadas para estudiar la mortalidad materna y los fenómenos asociados, incluyendo enfoques espaciales, temporales y multinivel, así como aplicaciones bayesianas para datos con variabilidad marcada y presencia de ceros.

Para el caso planteado en este estudio, se aplicará un modelo multinivel bayesiano en Stan para analizar la razón de mortalidad materna, con el fin de evaluar la disparidad a nivel conglomerado, departamental e individual considerando los factores determinantes.

Capítulo 3

Metodología

En este capítulo se presenta la metodología propuesta para alcanzar los objetivos del estudio. En primer lugar,se detalla la fuente de los datos, se ofrece una descripción de los datos, se definen los niveles a considerar en el modelo multinivel y se especifican las posibles variables del estudio. Finalmente, se describe el análisis estadístico del estudio, que incluye el análisis exploratorio de datos, la selección de las variables predictoras, la selección y comparación de los modelos, y el diagnóstico del modelo.

3.1. Descripción de los datos

Ante la falta de una base de datos consolidada que incluya factores asociados a la mortalidad materna, se elaboró una nueva base a partir de variables recopiladas en dos sitios web oficiales, siendo estos "TerriData" y "DANE", correspondiente al año 2023 o al más proximo posible, al tratarse de variables sociodemográficas que suelen mantenerse relativamente estables en el tiempo, su inclusión representa una aproximación válida y confiable.

Para la construcción de la base utilizada en este estudio, primero se descargaron desde TerriData las variables sociodemográficas disponibles para cada municipio, las cuales se encuentran organizadas por código departamento, departamento, código municipio, municipio e indicador. Estas variables se integraron en una única estructura mediante un proceso de unificación basado en el nombre del municipio y el departamento, asegurando que cada registro quedara asociado correctamente a su territorio. La razón de mortalidad materna, considerada como la variable respuesta, fue obtenida de TerriData para los municipios, mientras que para los departamentos se calculó a partir de los registros de estadísticas vitales del Departamento Administrativo Nacional de Estadística (DANE) y del sistema de vigilancia en salud pública (SIVIGILA).

La base resultante contiene 1.130 observaciones, correspondientes a los 32 departamentos, el distrito capital Bogotá y 1.098 municipios del país. Incluye 19 variables de tipo socioeconómico, demográfico y sanitario, además de 2 variables de localización. La Tabla 3.1 presenta un resumen detallado con el nombre de cada variable, el año al que corresponde la medición y el código asociado.

i	Variable	Año	Código
1	Código Municipio	-	Código Municipio
2	Departamento	-	Departamento
3	Municipio	-	Municipio
4	RMM	2023	RMM
5	Cobertura bruta en educación - Total	2022	Cob Edu Tot
6	Cobertura de acueducto rural (Censo)	2018	Cob Acu Rural
7	Cobertura de acueducto urbana (Censo)	2018	Cob Acu Urb
8	Cobertura de Energía Eléctrica Rural (Censo)	2018	Cob EE Rural
9	Cobertura de Energía Eléctrica Urbana (Censo)	2018	Cob EE Urb
10	Cobertura neta en educación total - Mujeres	2022	Cob Edu Muj
11	Índice de Incidencia del Conflicto Armado - IICA	2021	IICA
12	Índice de Necesidades Básicas Insatisfechas - NBI - en el área rural	2018	NBI Rural
13	Índice de Necesidades Básicas Insatisfechas - NBI - en el área urbana	2018	NBI Urb
14	Índice de pobreza multidimensional - IPM	2018	IPM
15	Promedio de controles prenatales	2020	Prom Ctrl Prenat
16	Porcentaje de partos atendidos por personal calificado	2021	Porc Partos Per
17	PIB	2023	PIB
18	Prop de Personas en NBI (%)	2021	Prop Pers NBI
19	IRCA	2023	IRCA
20	Longitud	-	Long
21	Latitud	-	Lat

TABLA 3.1: Diccionario de datos.

Los datos pueden entenderse con una estructura multinivel: el nivel 1 corresponde a los municipios y el nivel 2 a los departamentos, de manera que cada municipio se encuentra anidado dentro de su respectivo departamento.

3.1.1. Análisis estadístico

El análisis estadístico se divide en cuatro secciones: Análisis exploratorio de datos, composición de modelo, la selección de las variables predictoras, la selección y comparación de modelos, y el diagnóstico del modelo.

Análisis exploratorio de datos

El análisis exploratorio de datos (EDA), es considerado un instrumento indispensable al momento de realizar las aproximaciones iniciales al estudio de la estructura de la información socio-espacial en una determinada área de estudio.

Por medio de diferentes indicadores estadísticos, se generan conglomerados para analizar el comportamiento de la razón de mortalidad materna, la cual se ve influenciada por factores determinantes, que podrían variar entre los departamentos, municipios y/o agrupaciones de los mismos. También se lleva a cabo la creación de representaciones espaciales en mapas para visualizar la distribución geográfica de la razón de mortalidad materna, facilitando la identificación de áreas de concentración y posibles patrones según los factores determinantes.

Selección de las variables predictoras

La selección de las variables se realizó mediante un proceso combinado que incluyó la revisión de antecedentes y la consulta estructurada con expertos. En primer lugar, se llevó a cabo una revisión de literatura centrada en estudios que analizan la mortalidad materna desde enfoques espaciales, multinivel y bayesianos. Entre ellos se consideraron trabajos que emplean modelos multinivel bayesianos en contextos territoriales Jabessa y Jabessa, 2021, investigaciones sobre tendencias temporales y patrones de agrupación espacial Oliveira, I. V. G, et all, 2024, y análisis regionales de la mortalidad materna a lo largo del tiempo Oliveira, I. V. G, et all, 2023. Esta revisión permitió identificar factores sociodemográficos, territoriales y sanitarios recurrentes en la literatura especializada.

Complementariamente, se aplicó la metodología Delphi para obtener consenso entre expertos sobre los factores que debían ser incorporados en el modelo. Este procedimiento consistió en realizar rondas sucesivas de consulta anónima, retroalimentación controlada y refinamiento de criterios, con el fin de establecer cuáles variables eran consideradas indispensables y cuáles podían clasificarse como potencialmente relevantes para la modelación de la Razón de Mortalidad Materna (RMM). La aplicación de esta metodología permitió integrar el juicio experto de manera sistemática y reproducible, garantizando que la selección de variables respondiera tanto a la evidencia empírica como a la experiencia acumulada en el área.

Análisis via ACP

Una vez construida y depurada la base de datos, se aplicó un Análisis de Componentes Principales (ACP) como procedimiento exploratorio para examinar la estructura multivariada de las variables sociodemográficas incluidas en el estudio. Este análisis permitió evaluar las correlaciones entre los indicadores, resumir la información en un conjunto reducido de componentes y reconocer patrones de variabilidad conjunta entre las variables. (Montes Mora, C. L., 2024)

El ACP facilitó identificar aquellas variables con mayor contribución a los componentes principales, así como unidades territoriales (Departamentos y Municipios) que compartían perfiles similares o que se diferenciaban del resto. Esta exploración inicial aportó elementos para orientar la selección de factores asociados a la Razón de Mortalidad Materna (RMM) y para caracterizar la heterogeneidad territorial presente en los datos.

3.2. Modelo

3.2.1. Definición del modelo estadístico para la RMM: un enfoque bayesiano

Este trabajo se desarrolló bajo un enfoque bayesiano con el propósito de incorporar, de manera explícita, la información disponible sobre la Razón de Mortalidad Materna (RMM) y permitir que dicha evidencia se combine con la verosimilitud proveniente de los datos observados. Desde esta perspectiva, los parámetros del modelo se tratan como cantidades aleatorias, caracterizadas mediante una distribución a priori que refleja el conocimiento previo, formal o empírico sobre el fenómeno de estudio.

Al integrar la información previa con los datos, se obtiene una distribución a posteriori que resume de forma coherente la incertidumbre asociada a los parámetros y posibilita la cuantificación probabilística de los efectos. Esta formulación resulta especialmente útil en escenarios donde los recuentos son bajos, la estructura del modelo es jerárquica o existe heterogeneidad entre unidades territoriales, condiciones habituales en el análisis de la mortalidad materna a nivel municipal y departamental. Asimismo, el enfoque bayesiano facilita la estimación de modelos multinivel y la generación de intervalos de credibilidad, herramientas adecuadas para describir la variabilidad entre territorios y para incorporar múltiples fuentes de incertidumbre (Berger, J., 2006).

3.2.2. Uso del análisis bayesiano objetivo en la especificación del modelo

En este estudio se adoptó un análisis bayesiano objetivo con el fin de emplear distribuciones a priori no informativas que permitieran que la estimación estuviera guiada principalmente por los datos observados, en coherencia con la disponibilidad limitada de conocimiento previo específico sobre la Razón de Mortalidad Materna (RMM) en niveles municipales y departamentales.

Bajo este enfoque, las distribuciones a priori se seleccionaron según criterios propuestos por Berger, J., 2006, quienes recomiendan utilizar priors mínimamente informativas cuando no existe información previa confiable o cuando se busca reducir la influencia de supuestos subjetivos en modelos con estructuras jerárquicas. Estas priors actúan como una convención metodológica que permite comunicar de manera transparente la incertidumbre y facilita la reproducibilidad del análisis.

En el modelo hurdle multinivel empleado en este trabajo, se asignaron priors débiles o no informativas a los coeficientes de regresión y a las varianzas de los efectos aleatorios, con el objetivo de dejar que la estructura de los datos determinara la forma de la distribución a posteriori. Este tipo de especificación es adecuado en contextos donde se analizan recuentos bajos, variabilidad territorial significativa y modelos jerárquicos con múltiples parámetros, condiciones propias del estudio de la RMM en Colombia.

La utilización de priors no informativas también favoreció la implementación computacional en `Stan`, dado que este enfoque proporciona estabilidad en la exploración del espacio de parámetros y permite obtener distribuciones posteriores consistentes sin requerir procesos extensos de elicitación previa. De este modo, las decisiones metodológicas se centraron en la estructura del modelo, la definición de los niveles jerárquicos y los criterios de convergencia.

3.2.3. Modelo hurdle para RMM con inflación de ceros

Dado que en la base de datos la variable razón de mortalidad materna presenta una proporción elevada de ceros a nivel municipal (inflación de ceros), como se observa en la Figura 2.11 (Capítulo 2), se propone por utilizar un modelo hurdle, el cual está diseñado para manejar adecuadamente esta inflación de ceros en la RMM a nivel municipal.

El modelo hurdle, propuesto por Mullahy (1986), utiliza un proceso de modelado en dos etapas. La primera etapa estima una variable binaria que determina si la respuesta se encuentra por debajo o por encima del umbral (el cual, en este caso, es cero). La segunda etapa aplica un modelo truncado para explicar únicamente las

observaciones que superan dicho umbral, es decir, aquellas con valores positivos (Min, Y. and Agresti, A., 2005)

A continuación se plantea el modelo para la variable respuesta, razón de mortalidad materna del siguiente modo:

Y_{ij} : la razón de mortalidad materna (RMM) observada en el municipio i del departamento j, que puede ser cero o positiva.

El modelo hurdle parte de la premisa de que los ceros y los valores positivos observados en la variable de interés son generados por dos procesos distintos. Un proceso binario que se encargar de modelar el evento de que RMM materna tome el valor de cero, es decir no hayan ocurrido muertes maternas en el municipio. Y un proceso continuo, por medio del cual se modela el evento de que $Y_{ij} \in (0, \infty)$

Proceso binario: Presencia o ausencia de mortalidad materna

En la primera etapa del modelo, se aborda la ocurrencia o no de muertes maternas en cada municipio, lo cual constituye un proceso de decisión binaria. Es decir, se modela la probabilidad de que haya mortalidad materna (es decir, que $Y_{ij} > 0$). Para esto se define un variable indicadora Z_{ij}

$$Z_{ij} = \begin{cases} 0, & \text{si } Y_{ij} = 0 \\ 1, & \text{si } Y_{ij} > 0 \end{cases} \tag{3.1}$$

Este proceso se modela mediante un modelo logístico (logit):

$$P(Z_{ij} = 1) = \pi_{ij}, \quad \text{donde} \quad \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha_0 + \mathbf{w}_{ij}^{\top}\boldsymbol{\beta}^{(b)} + u_{j[i]} + v_i \tag{3.2}$$

donde

- α_0: Intercepto que fija la probabilidad promedio de registrar un caso cuando las covariables y efectos aleatorios son cero.
- $\mathbf{w}_{ij}$: Vector de covariables para el municipio i en el departamento j. Considerando las covariables presentes en la tabla 3.1
- $\boldsymbol{\beta}^{(b)}$: Vector de coeficientes para las covariables en la parte binaria.
- $u_{j[i]}$: Efecto aleatorio del departamento j al que pertenece el municipio i.
- v_i: Efecto aleatorio específico del municipio i.

Esta componente modela el "umbral" (hurdle) que determina la probabilidad de que Y_{ij} sea positivo, es decir, la probabilidad de que se observe al menos un caso en el municipio i del departamento j. Un caso se define como que la RMM en el municipio i sea mayor que cero, lo cual es equivalente a registrar al menos una muerte materna.

Parte continua: Proceso para los valores positivos $(Y_{ij} > 0)$

Condicional a que $Y_{ij} > 0$, es decir, dado que en el municipio i se registra al menos un caso de mortalidad materna, se modela la distribución de la RMM como una variable continua, positiva y sin cota superior. En este contexto, distribuciones como la log-normal o la gamma resultan adecuadas para capturar su comportamiento.

$$Y_{ij} \mid (Y_{ij} > 0) \sim \text{Lognormal}(\mu_{ij}, \sigma^2) \quad \Longleftrightarrow \quad \log(Y_{ij}) \mid (Y_{ij} > 0) \sim N(\mu_{ij}, \sigma^2) \tag{3.3}$$

En este caso, $\mu_{ij} \in \mathbb{R}$ corresponde al predictor lineal del modelo, mientras que $\sigma \in \mathbb{R}_{>0}$ representa la desviación estándar en la escala logarítmica. El predictor lineal, que define la función de enlace, se especifica como:

$$\mu_{ij} = \mu_0 + \mathbf{w}_{ij}^{\top} \boldsymbol{\beta}^{(g)} + u'_{j[i]} + v'_i \tag{3.4}$$

Donde:

- μ_0: Intercepto de la parte continua y corresponde al promedio de $\log(Y_{ij})$ especificamente al logaritmo de la RMM cuando no hay presencia de las covariables y los efectos aleatorios.
- $\mathbf{w}_{ij}$: Vector de covariables asociadas al municipio i en el departamento j, utilizadas para explicar la magnitud de la RMM cuando esta es positiva. Considerando las covariables presentes en la tabla 3.1
- $\boldsymbol{\beta}^{(g)}$: Vector de coeficientes asociados a las covariables $\mathbf{w}_{ij}$, que cuantifican su efecto sobre el valor esperado de $\log(Y_{ij})$.
- $u'_{j[i]}$: Representa el efecto aleatorio correspondiente al departamento j al que pertenece el municipio i, capturando la variabilidad entre departamentos no explicada por las covariables.
- v'_i: Efecto aleatorio específico del municipio i, que recoge la heterogeneidad residual entre municipios

3.2.4. Efectos aleatorios con dependencia espacial

En el modelo hurdle propuesto, se incorporan efectos aleatorios a nivel departamental y municipal para capturar la variabilidad que no se explica con las covariables.En ambos procesos del modelo, la parte binaria y la parte continua, se incluyen efectos aleatorios departamentales y municipales, que se representan como:

$$(u_{j[i]},\, v_i) \text{ (parte binaria)} \qquad (u'_{j[i]},\, v'_i) \text{ (parte continua)}$$

En el contexto territorial, además de la variabilidad entre unidades, puede existir dependencia entre áreas vecinas. Por esta razón, los efectos aleatorios pueden especificarse bajo dos posibles estructuras: una independiente, donde cada unidad territorial se trata como autónoma, y otra espacialmente estructurada, en la cual se permite que unidades vecinas presenten patrones de similitud. Bajo dependencia espacial, los efectos se modelan mediante distribuciones Gaussianas siguiendo la estructura de los Campos Aleatorios Gaussianos (GMRF) (Lee, 2013)

En general, para un efecto aleatorio cualquiera con estructura espacial se modela como un vector Gaussiano con media cero y matriz de precisión τQ. Donde τ es el parámetro de precisión (inverso de la varianza σ^2) y la matriz$Q$ de estructura espacial que codifica la posible dependencia entre unidades vecinas en cada nivel. Los elementos fuera de la diagonal de Q son distintos de cero únicamente cuando dos unidades territoriales son vecinas (Lee, 2013)

De este modo, los efectos aleatorios para cada parte del modelo hurdle se definen de la siguiente manera:

Efectos para la parte binaria:

$$u_{j[i]} \sim N\big(0, \tau_u^{(b)\,-1} Q_u^{-1}\big), \quad v_i \sim N\big(0, \tau_v^{(b)\,-1} Q_v^{-1}\big) \tag{3.5}$$

Efectos para la parte continua:

$$u'_{j[i]} \sim N\big(0, \tau_u^{(g)\,-1} Q_u^{-1}\big), \quad v'_i \sim N\big(0, \tau_v^{(g)\,-1} Q_v^{-1}\big) \tag{3.6}$$

donde:

- $u_{j[i]}(\cdot)$ representa el efecto aleatorio departamental, en la parte binaria (b) o parte continua (g)
- $v_i(\cdot)$ representa el efecto aleatorio municipal, en la parte binaria (b) o parte continua (g)
- Q_u y Q_v son las matrices de precisión espacial para departamentos y municipios, respectivamente. Cuando no hay dependencia espacial $Q = I$
- $\tau_u(\cdot)$ y $\tau_v(\cdot)$ son las precisiones específicas para cada parte del modelo.

Cuando no existe evidencia de dependencia espacial ($Q = I$), los efectos aleatorios se tratan como variables aleatorias, cada uno con distribución normal y varianza específica para la parte binaria o continua del modelo.

$$u_{j[i]} \sim N\big(0, (\tau_u^{(b)})^{-1}\big), \quad v_i \sim N\big(0, (\tau_v^{(b)})^{-1}\big) \tag{3.7}$$

$$u'_{j[i]} \sim N\big(0, (\tau_u^{(g)})^{-1}\big), \quad v'_i \sim N\big(0, (\tau_v^{(g)})^{-1}\big) \tag{3.8}$$

Cada u_j y v_i corresponde a una variable aleatoria específica para cada unidad territorial, mientras que en conjunto conforman los vectores aleatorios de efectos departamentales y municipales del modelo.

$$u = (u_1, \ldots, u_J), \quad v = (v_1, \ldots, v_N)$$

$$u' = (u'_1, \ldots, u_J), \quad v' = (v'_1, \ldots, v'_N)$$

Donde J es el número de departamentos y N_j el número de municipios en el departamento.Estos vectores siguen conjuntamente una distribución Gaussiana multivariada.

3.3. Modelo completo

El modelo estadístico que describe el comportamiento de la RMM para el municipio i (Y_{ij}), considera la combinación de las dos componentes antes definidas, una para los ceros y otro para los valores positivos. Esta se expresa de la siguiente manera:

$$f_{Y_{ij}}(y) = \begin{cases} 1 - \pi_{ij}, & y = 0 \\ \pi_{ij} \cdot f_{Y_{ij}}(y \mid y > 0), & y > 0 \end{cases} \tag{3.9}$$

donde π_{ij} representa la probabilidad de que Y_{ij} supere el umbral y se define como:

$$\pi_{ij} = \frac{\exp(\mathbf{w}_{ij}^{\top} \boldsymbol{\beta}^{(b)} + u_{j[i]} + v_i)}{1 + \exp(\mathbf{w}_{ij}^{\top} \boldsymbol{\beta}^{(b)} + u_{j[i]} + v_i)} \tag{3.10}$$

Para los casos en los que $Y_{ij} > 0$, se modela la razón de mortalidad materna mediante una distribución Log-normal, cuya función de densidad es:

$$f_{Y_{ij}}(y \mid y > 0) = \frac{1}{y\,\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log y - \mu_{ij})^2}{2\sigma^2} \right), \qquad y > 0. \tag{3.11}$$

Como alternativa, la parte continua también puede modelarse mediante una distribución Gamma si se considera que esta ofrece un ajuste más apropiado para la dispersión de los datos. En ese caso, se asume:

$$Y_{ij} \mid (Y_{ij} > 0) \sim \text{Gamma}(\mu_{ij}, \kappa), \qquad \log(\mu_{ij}) = \mathbf{w}_{ij}^{\top} \boldsymbol{\beta}^{(g)} + u'_{j[i]} + v'_i, \tag{3.12}$$

La expresión $\log(\mu_{ij})$ corresponde a la función de enlace empleada en los modelos lineales generalizados con distribución Gamma. Este enlace asegura que la media sea estrictamente positiva y permite vincularla con el predictor lineal que incorpora covariables y efectos aleatorios.

En esta parametrización, μ_{ij} representa la media del término Gamma y κ su parámetro de forma. Para obtener la forma explícita de la densidad, se parte de la formulación clásica de la distribución Gamma con parámetro de forma $\kappa > 0$ y parámetro de escala $\theta > 0$, cuya densidad se define como:

$$f(y \mid \kappa, \theta) = \frac{1}{\Gamma(\kappa)\,\theta^{\kappa}}\, y^{\kappa - 1} \exp\left(-\frac{y}{\theta} \right), \qquad y > 0. \tag{3.13}$$

En esta parametrización clásica, la media está dada por:

$$\mu = \kappa\,\theta.$$

Si la distribución se parametriza mediante la media μ_{ij} y la forma κ, entonces:

$$\theta = \frac{\mu_{ij}}{\kappa}.$$

Sustituyendo $\theta = \mu_{ij}/\kappa$ en la densidad clásica, se obtiene la expresión equivalente utilizada en el modelo:

$$f_{Y_{ij}}(y \mid y > 0) = \frac{\kappa^{\kappa}}{\Gamma(\kappa)\,\mu_{ij}^{\kappa}}\, y^{\kappa-1} \exp\left(-\kappa\,\frac{y}{\mu_{ij}}\right), \qquad y > 0,\ \kappa > 0,\ \mu_{ij} > 0. \tag{3.14}$$

De este modo, la parte binaria identifica qué municipios presentan mortalidad materna distinta de cero y cuáles no, mientras que la parte continua modela la variación en la razón de mortalidad materna entre aquellos municipios con valores positivos. Ambos componentes incorporan efectos aleatorios específicos para departamentos y municipios, lo que permite capturar la estructura jerárquica y la posible correlación espacial en los datos.

3.3.1. Verosimilitud del modelo Hurdle

Sea Y_{ij} la Razón de Mortalidad Materna (RMM) observada en el municipio i, con $i = 1, \ldots, N_j$, donde N_j corresponde al número total de municipios dentro del departamento j. Recordando la ecuacion 3.1 , Z se define como:

$$Z_{ij} = \begin{cases} 0, & \text{si } Y_{ij} = 0 \\ 1, & \text{si } Y_{ij} > 0 \end{cases}$$

de modo que $Z_{ij} = 0$ cuando no se registra mortalidad materna en el municipio i, y $Z_{ij} = 1$ cuando la RMM es positiva.

En el modelo hurdle, la probabilidad de que el municipio i no registre muertes maternas es $(1 - \pi_{ij})$, mientras que para los casos con $Y_{ij} > 0$, la contribución del dato continuo está dada por la densidad $f_Y(y_{ij} \mid \mu_i, \psi)$, donde $f_Y(\cdot)$ representa la distribución continua utilizada en la segunda parte del modelo (por ejemplo, Log-normal o Gamma),y ψ representa el parámetro de dispersión correspondiente, que puede ser σ en el caso de la distribución Log-normal o κ si se emplea la distribución Gamma.

La contribución de cada observación a la verosimilitud puede expresarse de forma unificada como:

$$(1 - \pi_{ij})^{1-Z_{ij}} \left[\pi_{ij}\, f_Y(y_{ij} \mid \mu_{ij}, \psi)\right]^{Z_{ij}}.$$

Asumiendo independencia condicional entre municipios, la verosimilitud total del modelo es:

$$L(\{\pi_{ij}\}, \{\mu_{ij}\}, \psi \mid \mathbf{Z}) = \prod_{i=1}^{N} (1 - \pi_{ij})^{1-Z_{ij}} \left[\pi_{ij}\, f_Y(y_{ij} \mid \mu_i, \psi)\right]^{Z_{ij}}. \tag{3.15}$$

En esta expresión:

- $(1 - \pi_{ij})^{1-Z_{ij}}$ corresponde a los municipios con $Y_i = 0$,
- $\left[\pi_{ij}\, f_Y(y_{ij} \mid \mu_{ij}, \psi)\right]^{Z_{ij}}$ corresponde a los municipios con $Y_{ij} > 0$,
- f_Y queda especificada posteriormente según se implemente la distribución Log-normal o Gamma.

3.3.2. Distribuciones a priori

La especificación de las distribuciones a priori siguió un enfoque jerárquico bayesiano, utilizando prioris no informativas. Para los efectos aleatorios departamentales

y municipales se empleó una parametrización no centrada, que consiste en representar cada efecto como el producto entre una desviación estándar y un término estandarizado. En esta formulación, los efectos "raw" corresponden a variables auxiliares que no contienen información territorial, sino que actúan como base estocástica sobre la cual se escala la variabilidad real.

Efectos aleatorios y parametrización no centrada

Los efectos aleatorios para la parte Bernoulli del modelo se expresan como:

$$u_{j[i]} = \sigma_{\text{dep},b}\, u_{j[i]}^{(\text{raw})}, \qquad v_i = \sigma_{\text{mpio},b}\, v_i^{(\text{raw})}, \tag{3.16}$$

y para la parte continua:

$$u'_{j[i]} = \sigma_{\text{dep},g}\, u_{j[i]}^{(\text{raw})'}, \qquad v'_i = \sigma_{\text{mpio},g}\, v_i^{(\text{raw})'}. \tag{3.17}$$

Los términos estandarizados cumplen:

$$u_{j[i]}^{(\text{raw})},\ v_i^{(\text{raw})},\ u_{j[i]}^{(\text{raw})'},\ v_i^{(\text{raw})'} \sim N(0,1). \tag{3.18}$$

Esta representación separa la estructura de variación ($u^{(\text{raw})}$) de su magnitud (σ), lo que mejora la identificación del modelo y favorece la estabilidad del muestreo Hamiltoniano. Gelman, A., 2013 señalan que la parametrización no centrada reduce divergencias y mejora la mezcla en modelos jerárquicos con información limitada.

Además, aunque las desviaciones estándar asociadas a los efectos aleatorios reciben prioris truncadas como Half-Cauchy o Half-Normal, esta especificación no altera la forma condicional de los efectos aleatorios, que permanece como una distribución Normal con media cero y varianza σ^2. Lo que cambia es su distribución marginal, que presenta colas más pesadas debido a la incertidumbre asociada a σ.

Prioris para las desviaciones estándar

Las desviaciones estándar de los efectos aleatorios reciben prioris Half-Cauchy o Half-Normal, seleccionadas por su capacidad para regularizar (restricciones) sin imponer restricciones excesivas. Las Half-Cauchy presentan colas pesadas y permiten acomodar variabilidad amplia cuando la información empírica es escasa (Gelman, 2006), como se presenta en el caso de la RMM, mientras que las Half-Normal (N^+) generan mayor concentración y se emplean cuando se espera menor dispersión.

$$\sigma_{\text{dep},b} \sim \text{Cauchy}^+(0,1), \qquad \sigma_{\text{mpio},b} \sim \text{Cauchy}^+(0,0.5), \tag{3.19}$$

$$\sigma_{\text{dep},g} \sim \text{Cauchy}^+(0,0.25), \qquad \sigma_{\text{mpio},g} \sim N^+(0,0.05^2), \qquad \sigma \sim N^+(0,1.5^2). \tag{3.20}$$

Los valores de escala reflejan diferencias esperadas en la variabilidad departamental y municipal, así como entre la parte Bernoulli y la parte continua del modelo.

En el contexto de la RMM, los efectos aleatorios corresponden a varianzas en distintos niveles jerárquicos y en las dos partes del modelo. Como se muestra en Gelman, 2006 y en los ejemplos de 3 y 8 escuelas, las distribuciones Half-Cauchy actúan como priors débilmente informativas: permiten que los datos dominen la estimación de la varianza, pero evitan colas extremadamente largas y posteriors impropias que pueden ocurrir con priors uniformes o inverse-gamma poco informativas.

Dado que las desviaciones estándar son siempre positivas, se utilizan distribuciones Half-Cauchy o Half-Normal, que están truncadas en $[0, \infty)$, asegurando coherencia con su interpretación como magnitudes no negativas. La Half-Cauchy, con colas pesadas, permite acomodar una amplia variabilidad cuando la información empírica es escasa, evitando que la posterior se vuelva impropia. La Half-Normal, en cambio, concentra mayor probabilidad cerca de cero, siendo útil cuando se espera menor dispersión y se desea un efecto de regularización más fuerte sin imponer información demasiado estricta.

En el modelo RMM, los valores de escala de las desviaciones estándar reflejan la variabilidad esperada en cada nivel jerárquico y en cada parte del modelo. Para la parte Bernoulli, la desviación estándar departamental ($\sigma_{\text{dep},b}$) se establece con una escala más amplia, permitiendo reflejar la mayor variabilidad entre departamentos aun con pocos datos por grupo. La desviación estándar municipal ($\sigma_{\text{mpio},b}$)se define con una escala menor para mantener la posterior más controlada, dado que la variabilidad a nivel municipal suele ser menor y los datos proporcionan más información.

Al igual que en la parte binaria, en la parte continua Log-Normal la desviación estándar departamental $(\sigma_{\text{dep},g})$ tiene una escala mayor que la desviación estándar municipal $(\sigma_{\text{mpio},g})$, reflejando que se espera mayor variabilidad entre departamentos que entre municipios, especialmente entre aquellos que pertenecen al mismo departamento. Al usar una Half-Cauchy para $\sigma_{\text{dep},g}$ se permite que la posterior considere valores altos de variabilidad departamental si los datos lo sugieren, mientras que la Half-Normal para $\sigma_{\text{mpio},g}$ mantiene la dispersión municipal más concentrada cerca de cero, acorde con la menor variabilidad esperada a ese nivel.

La desviación estándar residual (σ) captura la variabilidad no explicada por los efectos fijos ni por los efectos aleatorios departamentales o municipales en la parte continua del modelo. Se asigna una prior con una escala relativamente amplia respecto a las demás, para que la estimación sea dominada por los datos y la dispersión individual observada pueda ser adecuadamente representada.

Se debe mencionar que, en el modelo RMM, las covariables X están estandarizadas y las respuestas de la parte continua tienen una escala conocida. Por eso, se usan escalas pequeñas, cercanas a 1 o menores, para las priors de las desviaciones estándar, de modo que sean consistentes con los efectos esperados y no asignen probabilidad a valores demasiado grandes o poco plausibles.

Prioris para los efectos fijos

Los efectos fijos se modelaron mediante prioris Normales:

$$\alpha_0,\, \mu_0 \sim N(0,1), \qquad \beta_{bk},\, \beta_{gk} \sim N(0,1). \tag{3.21}$$

Estas distribuciones son débiles pero regularizadoras, permitiendo que los datos guíen la estimación sin imponer supuestos fuertes, lo cual es apropiado en modelos hurdle con baja frecuencia de eventos (Min y Agresti, 2005). En conjunto, el uso de prioris Half-Cauchy, Half-Normal y Normales estandarizadas proporciona un esquema bayesiano regularizado, coherente con modelos jerárquicos complejos y con el enfoque bayesiano objetivo adoptado en este estudio.

3.3.3. Distribución a posteriori

La distribución a posteriori del modelo se construye combinando la verosimilitud del modelo hurdle con las distribuciones a priori asignadas a cada uno de los parámetros.

Parámetros del modelo

El conjunto de parámetros está dado por:

- α_0: intercepto de la parte Bernoulli.
- μ_0: intercepto de la parte continua.
- $\beta_{b1}, \ldots, \beta_{bm_b}$: efectos fijos de la parte Bernoulli.
- $\beta_{g1}, \ldots, \beta_{gm_g}$: efectos fijos de la parte continua.
- $\sigma_{\text{dep},b}$ y $\sigma_{\text{mpio},b}$: desviaciones estándar departamental y municipal para la parte Bernoulli.
- $\sigma_{\text{dep},g}$ y $\sigma_{\text{mpio},g}$: desviaciones estándar departamental y municipal para la parte continua.
- σ: desviación estándar de la parte continua (si se utiliza la distribución Log-normal).
- $u_{j[i]}^{(\text{raw})}$, $u_{j[i]}^{(\text{raw})'}$ para $j = 1, \ldots, J$: efectos aleatorios departamentales estandarizados.
- $v_i^{(\text{raw})}$, $v_i^{(\text{raw})'}$ para $i = 1, \ldots, N_j$: efectos aleatorios municipales estandarizados.

Nota: m_b y m_g representan el número de covariables incluidas en la parte Bernoulli y en la parte continua, respectivamente. En consecuencia, $\boldsymbol{\beta}_b = (\beta_{b1}, \ldots, \beta_{bm_b})$ y $\boldsymbol{\beta}_g = (\beta_{g1}, \ldots, \beta_{gm_g})$.

Los efectos aleatorios reales se obtienen mediante la parametrización no centrada:

$$u_{j[i]} = \sigma_{\text{dep},b}\, u_{j[i]}^{(\text{raw})}, \qquad v_i = \sigma_{\text{mpio},b}\, v_i^{(\text{raw})},$$
$$u'_{j[i]} = \sigma_{\text{dep},g}\, u_{j[i]}^{(\text{raw})'}, \qquad v'_i = \sigma_{\text{mpio},g}\, v_i^{(\text{raw})'}.$$

Verosimilitud

La verosimilitud del modelo hurdle se expresa como:

$$L = \prod_{i=1}^{N} (1 - \pi_{ij})^{1 - Z_{ij}} \left[\pi_{ij}\, f_Y(y_{ij} \mid \mu_{ij}, \psi) \right]^{Z_{ij}},$$

donde $f_Y(y_{ij} \mid \mu_{ij}, \psi)$ es la densidad continua (Log-normal o Gamma) y ψ representa el parámetro de dispersión correspondiente.

Distribuciones a priori

Los interceptos reciben prioris normales:

$$\alpha_0 \sim N(0,1), \qquad \mu_0 \sim N(0,1).$$

Los efectos fijos se especifican mediante:

$$\prod_{k=1}^{m_b}(\beta_{bk} \mid \cdots)N \sim (0,1), \qquad \prod_{k=1}^{m_g}(\beta_{gk} \mid \cdots)N \sim (0,1)$$

Las desviaciones estándar se modelan como:

$$\sigma_{\text{dep},b} = \text{Cauchy}^{+}(0,1), \qquad \sigma_{\text{mpio},b} = \text{Cauchy}^{+}(0,0.5),$$

$$\sigma_{\text{dep},g} \sim \text{Cauchy}^{+}(0,0.25), \qquad \sigma_{\text{mpio},g} \sim N^{+}(0,0.05^2), \qquad \sigma \sim N^{+}(0,1.5^2).$$

Los efectos aleatorios estandarizados tienen prioris normales:

$$\prod_{j=1}^{J} N(u_{j[i]}^{(\text{raw})} \mid 0,1), \qquad \prod_{i=1}^{N} N(v_i^{(\text{raw})} \mid 0,1),$$

$$\prod_{j=1}^{J} N(u_{j[i]}^{(\text{raw})'} \mid 0,1), \qquad \prod_{i=1}^{N} N(v_i^{(\text{raw})'} \mid 0,1).$$

Distribución a posteriori

Combinando verosimilitud y prioris, la distribución posterior toma la forma:

$$p(\text{parámetros} \mid \mathbf{y}, \mathbf{z}) \propto \left[\prod_{i=1}^{N}(1-\pi_{ij})^{1-Z_{ij}} \left(\pi_{ij}\, f_Y(y_{ij} \mid \mu_{ij}, \psi)\right)^{Z_{ij}}\right] \times (\text{producto de todas las prioris}).$$

Esta es la distribución de la cual se obtienen muestras mediante Hamiltonian Monte Carlo (HMC) con el adaptador No-U-Turn Sampler (NUTS) implementado en `Stan`, dada la ausencia de una forma cerrada para la posterior.

La distribución a posteriori del modelo no puede obtenerse en forma cerrada debido a la estructura jerárquica multinivel, la presencia de dos componentes con naturalezas distintas (Bernoulli y Log-normal), y el uso de distribuciones a priori no conjugadas, como las half-Cauchy asignadas a las desviaciones estándar jerárquicas. Estas características hacen que la posterior involucre integrales de alta dimensión que no admiten solución analítica.

Situaciones de este tipo son comunes en modelos bayesianos complejos, tal como se discute en Gelman, A., 2013 y en la formulación clásica de modelos *hurdle* con efectos aleatorios propuesta por Min, Y. and Agresti, A., 2005.

3.3.4. Diagnóstico del modelo

Para un modelo ajustado mediante `Stan`, el diagnóstico de convergencia se realiza a partir de los estadísticos y herramientas específicas del algoritmo NUTS/HMC. En este tipo de estimación es necesario verificar el estadístico $\widehat{R}$, que permite evaluar

si las cadenas se han mezclado adecuadamente y si están muestreando de la misma distribución posterior. Asimismo, se revisan los tamaños efectivos de muestra, tanto ESS_{bulk} como ESS_{tail}, con el propósito de valorar la precisión de las estimaciones en las regiones centrales y en las colas de las distribuciones posteriores.

De forma complementaria, se examinan los gráficos de trazas (*trace plots*), los cuales permiten observar visualmente la mezcla, estabilidad y ausencia de patrones sistemáticos en las cadenas simuladas. Estas verificaciones son fundamentales en la validación de modelos en `Stan`, pues permiten establecer si el algoritmo NUTS/HMC exploró de manera adecuada el espacio paramétrico y si las simulaciones alcanzaron un comportamiento estacionario. A partir de estos elementos se confirma la convergencia del modelo y la confiabilidad de las distribuciones posteriores empleadas en su interpretación.

3.3.5. Estimación e interpretación de los coeficientes del modelo

Una vez verificada la convergencia del modelo y la adecuada exploración del espacio paramétrico, la inferencia sobre los efectos de las covariables se realiza a partir de las distribuciones posteriores de los coeficientes de regresión asociados a los efectos fijos. En el modelo *hurdle* multinivel, estos coeficientes se estiman de manera diferenciada para cada uno de los componentes del modelo. En la parte Bernoulli, los coeficientes corresponden a los efectos de las covariables sobre el logaritmo de las *odds* de que la Razón de Mortalidad Materna sea mayor que cero, mientras que en la parte lognormal los coeficientes representan efectos sobre el valor esperado del logaritmo de la RMM condicionada a su ocurrencia. En ambos casos, los coeficientes se obtienen directamente de las simulaciones posteriores generadas mediante el algoritmo HMC/NUTS, lo que permite caracterizar su incertidumbre de forma completa.

La evaluación de la relevancia estadística de las covariables se fundamenta en criterios propios de la inferencia bayesiana, basados en las distribuciones posteriores de los coeficientes estimados. Para cada covariable se consideran la media posterior como medida puntual del efecto, los intervalos creíbles al 95 % como medida de incertidumbre y la probabilidad posterior de dirección como indicador de la consistencia del signo del efecto. En el componente Bernoulli, los coeficientes pueden transformarse mediante la función exponencial para su interpretación en términos de razones de *odds*, mientras que en el componente lognormal esta transformación permite interpretar los coeficientes en la escala original de la Razón de Mortalidad Materna, como cambios relativos asociados a variaciones en las covariables, condicionados a la ocurrencia del evento. Este procedimiento proporciona un marco sistemático y coherente para evaluar la relevancia estadística de las covariables en cada componente del modelo, manteniendo una interpretación diferenciada entre los procesos de ocurrencia y magnitud.

Capítulo 4

Resultados y análisis

4.1. Resultado 1: Análisis descriptivo, espacial y de agrupamiento territorial de la Razón de Mortalidad Materna en Colombia

4.1.1. Introducción

El primer objetivo de este estudio consiste en analizar el comportamiento geográfico de la Razón de Mortalidad Materna (RMM) en Colombia, con el fin de identificar posibles agrupaciones o patrones territoriales que permitan reconocer áreas de mayor riesgo y orientar intervenciones prioritarias. Este análisis espacial y descriptivo también sirve para comprender las desigualdades territoriales y socioeconómicas que influyen en la mortalidad materna, y proporciona la base empírica necesaria para la posterior modelación jerárquica mediante métodos bayesianos.

Diversos estudios han mostrado que en Colombia persisten brechas estructurales entre territorios en materia de pobreza, acceso a servicios públicos, provisión estatal y capacidad institucional. Cortés, D. and Vargas, J. F., 2012 señalan que estas desigualdades regionales son profundas, persistentes y vinculadas al rezago histórico en bienes públicos e infraestructura. Estos contrastes impactan directamente los determinantes de la salud materna: condiciones de ruralidad, carencias en infraestructura básica, limitaciones educativas y deficiencias en los servicios de salud generan riesgos diferenciados entre departamentos y municipios.

El análisis espacial y de agrupamiento permite examinar si estas disparidades se reflejan en patrones territoriales de la RMM. Para ello, se evalúa la presencia de autocorrelación espacial y se identifican territorios con niveles similares de RMM y características sociodemográficas, económicas y sanitarias. Sin embargo, como advierten Siabato, W. and Guzmán-Manrique, J., 2019, la autocorrelación espacial no debe asumirse de antemano, pues depende de la escala y del fenómeno analizado.

En caso de no presentarse correlación espacial, se recurre a las agrupaciones territoriales obtenidas mediante análisis de conglomerados jerárquicos, las cuales reflejan similitudes estructurales entre departamentos. Estos perfiles sintetizan patrones latentes de desigualdad y se contextualizan a partir de la literatura sobre inequidad regional en Colombia, en particular la persistencia de territorios con rezagos profundos en provisión estatal y capacidad institucional.

Estadística	Descripción
n	Número de observaciones disponibles.
media	Promedio aritmético de la variable.
desviación estándar (sd)	Medida de dispersión alrededor de la media.
coeficiente de variación (CV %)	Variabilidad relativa respecto al promedio.
mediana	Valor central de los datos ordenados.
mínimo (min)	Valor mínimo observado.
máximo (max)	Valor máximo observado.
primer cuartil (Q1)	Valor por debajo del cual se encuentra el 25 % de los datos.
tercer cuartil (Q3)	Valor por debajo del cual se encuentra el 75 % de los datos.

TABLA 4.1: Estadísticas descriptivas utilizadas

4.1.2. Análisis descriptivo a nivel departamental

Variable	n	Media	SD	CV (%)	Mediana	Min	Max	Q1	Q3
RMM	33	98.28	69.86	71.08	85.51	0.00	297.18	56.80	117.90
Cob Edu Tot	33	99.55	11.08	11.13	102.06	66.01	118.49	94.41	105.21
Cob Acu Rural	33	43.31	23.28	53.75	47.64	3.79	84.47	26.89	62.92
Cob Acu Urb	33	85.73	19.50	22.75	94.19	31.24	99.53	84.62	97.80
Cob EE Rural	33	75.73	25.58	33.78	83.39	11.66	99.16	56.99	94.24
Cob EE Urb	33	97.58	3.18	3.26	98.79	84.69	99.76	97.08	99.32
Cob Edu Muj	33	89.02	10.50	11.80	94.05	55.68	101.08	83.71	95.91
NBI Rural	33	35.02	22.05	62.98	27.11	5.88	86.17	19.21	45.32
NBI Urb	33	16.99	13.18	77.58	13.28	3.34	68.29	6.92	21.69
IPM	33	37.24	18.27	49.06	37.83	9.00	78.82	22.72	44.56
Prom Ctrl Prenat	33	5.08	1.33	26.18	5.46	1.87	6.66	4.79	5.96
Porc Partos Per	33	92.97	11.23	12.08	97.06	53.66	100.00	93.29	99.30
PIB	32	47963.05	73376.95	152.99	25756.74	426.94	347978.95	8280.10	45157.09
Prop Pers NBI	31	26.09	19.32	74.07	18.96	6.25	68.94	11.30	33.32

TABLA 4.2: Estadísticas descriptivas por variable - Departamentos

Las estadísticas descriptivas presentadas en la Tabla 4.2 muestran una variabilidad entre departamentos en diferentes indicadores demográficos, socioeconómicos y sanitarios, salvo en el caso del IRCA y el IICA, que solo están disponibles para tres y un departamento, respectivamente. . La RMM presenta un coeficiente de variación superior al 70 %, lo que evidencia diferencias en el riesgo de mortalidad materna entre regiones. Los departamentos con mayores niveles de pobreza multidimensional, altos índices de necesidades básicas insatisfechas (NBI) y coberturas limitadas de servicios públicos tienden a exhibir valores más altos de RMM, coherente con patrones de desigualdad territorial documentados en el país Cortés, D. and Vargas, J. F., 2012. Se puede corroborar mediante las correlaciones observadas en la figura 4.4

Por otro lado, variables asociadas a acceso a servicios de salud materna como controles prenatales o atención del parto por personal calificado, muestran niveles altos en la mayoría de departamentos, aunque con excepciones marcadas en territorios históricamente rezagados (Amazonas, Vichada, Vaupés, Chocó). Las coberturas rurales tienen mayor variabilidad que las urbanas, lo que también coincide con las brechas urbano–rurales reportadas en diversos estudios nacionales.

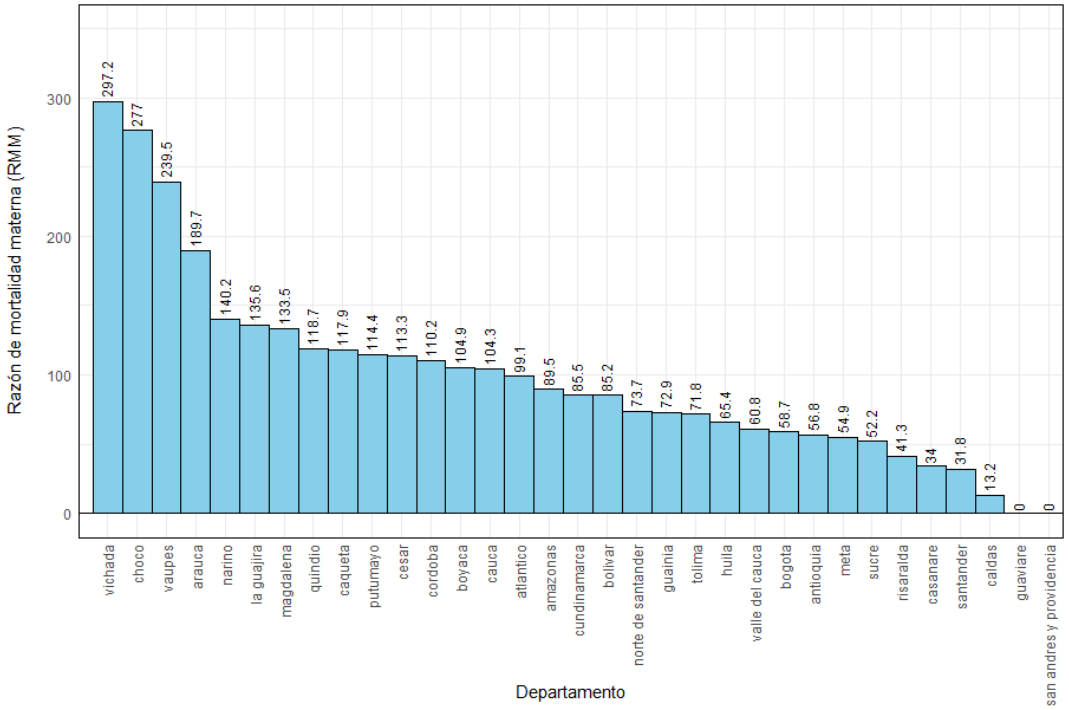

FIGURA 4.1: Razón de Mortalidad Materna(RMM) por Departamento

La figura 4.1 muestra una marcada heterogeneidad en la Razón de Mortalidad Materna (RMM) entre los departamentos de Colombia. Se observa que varios territorios superan de manera considerable el promedio nacional, destacándose especialmente Vaupés, Vichada, Chocó y Arauca. En contraste, también existen departamentos con niveles notablemente inferiores al promedio. Estas diferencias sugieren desigualdades significativas en el acceso a servicios de salud materna a nivel territorial, las cuales impactan directamente la atención durante el embarazo, el parto y el puerperio. Esta variabilidad territorial subraya la necesidad de priorizar intervenciones en las regiones con mayores niveles de RMM.

Asimismo, se identifican departamentos como Guaviare y San Andrés que no reportan valores de RMM. Esto podria deberse a problemas de subregistro y no a la inexistencia de mortalidad materna.

Se elaboró un mapa de calor (heatmap) de la Razón de Mortalidad Materna (RMM) por departamentos de Colombia (Figura 4.2), para representar espacialmente su comportamiento. Esta representación permite distinguir contrastes regionales marcados y reconocer agrupaciones de departamentos con niveles particularmente altos o bajos de mortalidad materna.

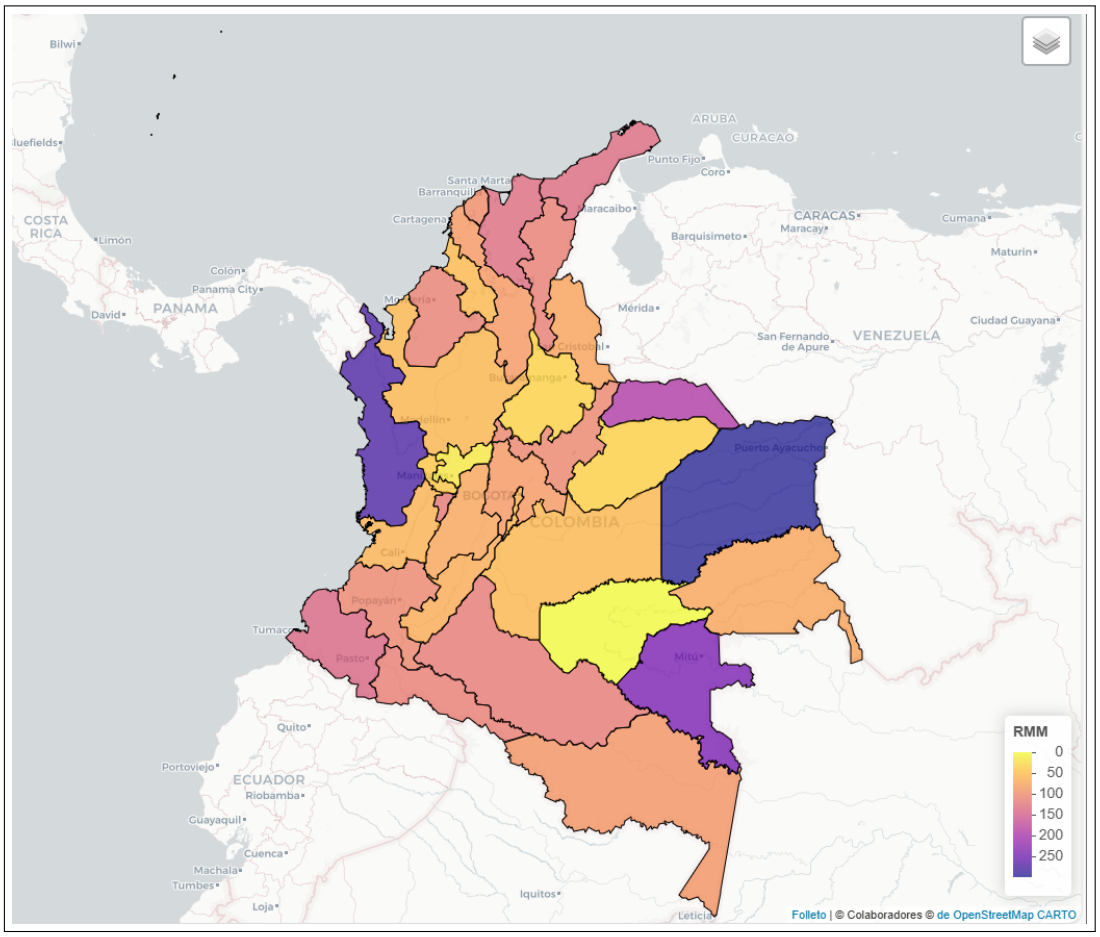

FIGURA 4.2: Mapa de calor (heatmap) de la RMM registrada por Departamento

Se destacan departamentos como: Vichada, Vaupes y chocó con indices más altos. No obstante, no se evidencia un patrón claro de autocorrelación geográfica en los valores de la RMM; es decir, los departamentos contiguos no necesariamente presentan niveles similares de mortalidad materna. En caso de que esta especulación sea cierta, las posibles agrupaciones no estarian determinadas por la proximidad territorial, sino más bien por la presencia de características socioeconómicas, demográficas y de acceso a servicios que son compartidas entre departamentos incluso distantes entre sí (Figura 4.2).

Análisis espacial exploratorio

Con el propósito de evaluar si la distribución geográfica de la RMM presenta patrones espaciales significativos, se aplicaron los estadísticos de autocorrelación espacial global de Moran y Geary. Según Siabato, W. and Guzmán-Manrique, J., 2019, la autocorrelación espacial permite identificar si valores similares de una variable tienden a agruparse espacialmente o si, por el contrario, están distribuidos de manera aleatoria.

Los resultados obtenidos para la RMM a nivel departamental sugieren que no existe autocorrelación espacial global significativa. Esto implica que los departamentos contiguos no presentan necesariamente valores de RMM similares, y por lo tanto la variación territorial no sigue un patrón estructurado por proximidad geográfica.

Es fundamental aclarar dos puntos conceptuales:

- **Ausencia de autocorrelación espacial no implica homogeneidad territorial.** La falta de dependencia espacial indica que los valores altos o bajos de RMM no se agrupan geográficamente, pero sí pueden reflejar desigualdades estructurales asociadas a pobreza, ruralidad, conflicto o capacidad institucional.
- **Autocorrelación espacial no es equivalente a dependencia jerárquica.** Según Sánchez-Cantalejo, E. and Ocaña-Riola, R., 1999 , la estructura jerárquica emerge porque las unidades de análisis (municipios) están anidadas dentro de unidades mayores (departamentos). Esto genera dependencia intragrupo, incluso cuando no hay dependencia espacial entre unidades vecinas.

En otras palabras, aunque no exista un patrón espacial continuo, sí puede existir una estructura jerárquica real que debe ser modelada, ya que los municipios dentro de un mismo departamento comparten condiciones institucionales, administrativas y de provisión de servicios que los hacen más similares entre sí que respecto a municipios de otros departamentos. Esta distinción conceptual es clave para justificar el uso posterior de modelos multinivel bayesianos: la decisión de emplearlos se sustenta en la estructura administrativa y de gobernanza del país y en la variabilidad intra e interdepartamental, no en la presencia de autocorrelación espacial.

Análisis univariado

Con el fin de explorar el comportamiento individual de cada una de las variables cuantitativas del estudio, se realizó un análisis univariado empleando diagramas de caja (boxplots). Esta representacón permite identificar la mediana, la dispersión, la simetría y la presencia de posibles valores atípicos

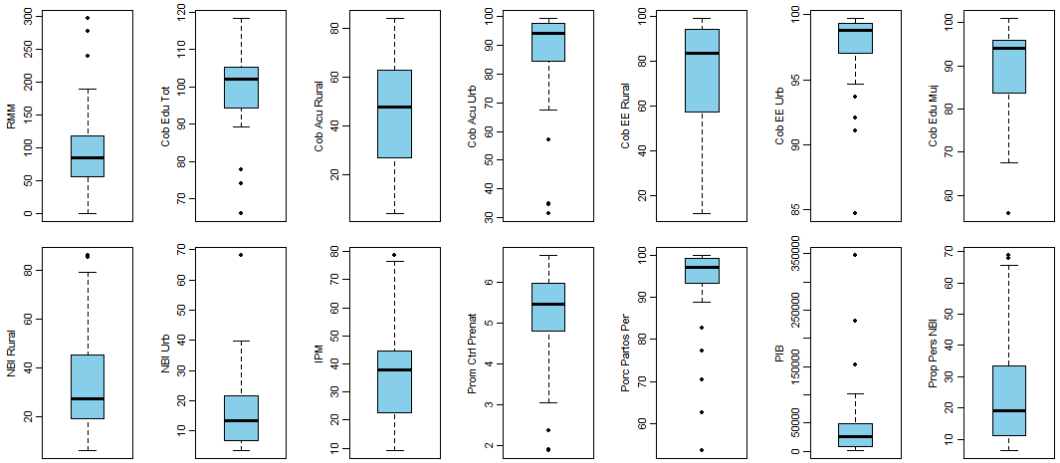

FIGURA 4.3: Boxplots de las variables sociodemográficas asociadas a la RMM

RMM	**Cob Edu Tot**	**Cob Acu Urb**	**Cob EE Urb**
vichada, vaupes, bogota	vichada, vaupes, amazonas	amazonas, san andres y providencia, putumayo, bogota	vichada, vaupes, amazonas, san andres y providencia
Cob Edu Muj	**NBI Rural**	**NBI Urb**	**IPM**
vaupes	vichada, vaupes	bogota	vaupes
Prom Ctrl Prenat	**Porc Partos Per**	**PIB**	**Prop Pers NBI**
vichada, vaupes, amazonas	vichada, vaupes, amazonas, san andres y providencia, bogota	san andres y providencia, atlantico, antioquia	vichada, vaupes

TABLA 4.4: Departamentos identificados como valores atípicos por variable

Hay una presencia marcada de valores atípicos en la mayoría de las variables, muchos de los cuales se repiten de manera consistente entre indicadores asociados. No obstante, las coberturas rurales de acueducto y educación constituyen una excepción clara: en estas se observa que las zonas rurales presentan niveles de cobertura inferiores y con mayor variabilidad en comparación con sus contrapartes urbanas. Aunque el indicador de Necesidades Básicas Insatisfechas (NBI) muestra un comportamiento inverso, al registrar valores más altos en el ámbito rural mantiene la misma tendencia general: mayor homogeneidad en las áreas urbanas y disparidades más pronunciadas en las zonas rurales.

Análisis de correlación entre variables

Con el fin de explorar la estructura relacional entre las variables sociodemográficas, económicas y sanitarias, se construyó una matriz de correlaciones a nivel departamental. Este análisis permite identificar patrones lineales que pueden influir en la variabilidad de la RMM y facilita la selección preliminar de covariables relevantes para modelos posteriores.

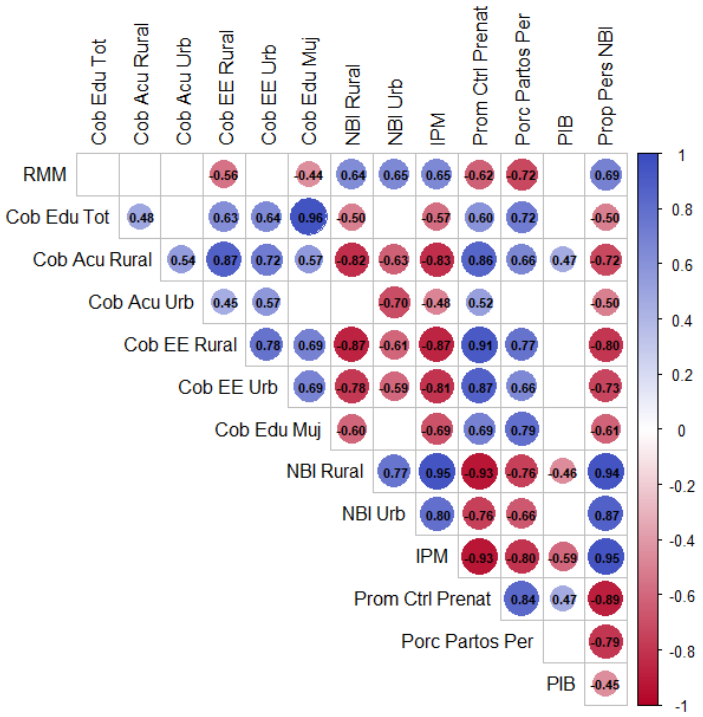

FIGURA 4.4: Correlograma de variables sociodemográficas y RMM

El correlograma muestra que la Razón de Mortalidad Materna (RMM) mantiene correlaciones moderadas, tanto positivas como negativas, con diversas variables socioeconómicas. En particular, se observa una correlación directa con el porcentaje de personas con Necesidades Básicas Insatisfechas (Prop Pers NBI), el NBI urbano y rural, y el Índice de Pobreza Multidimensional (IPM). Por el contrario, la RMM presenta correlaciones inversas con las coberturas educativas (urbana y rural), el promedio de controles prenatales y el porcentaje de partos atendidos por personal calificado. Estos patrones sugieren que mayores niveles de pobreza o carencias básicas se asocian con una mayor mortalidad materna, mientras que mejores niveles educativos y mayor acceso a servicios de salud se relacionan con una reducción de la RMM.

Asimismo, se destaca que todas las variables de cobertura muestran una correlación positiva superior al 50 % con el promedio de controles prenatales y con la atención del parto por personal calificado. Esto indica que mejores niveles de cobertura en servicios públicos y educativos suelen acompañarse de una mayor utilización de servicios prenatales y de una mayor institucionalización del parto. No obstante, estas coberturas no presentan una correlación lineal fuerte con la RMM, lo que sugiere que su efecto sobre la mortalidad materna podría depender de otros factores contextuales.

Análisis de Componentes Principales (ACP)

El Análisis de Componentes Principales (ACP) se aplicó con el propósito de reducir la dimensionalidad del conjunto de datos y explorar la estructura subyacente entre las variables sociodemográficas. Para ello, se seleccionaron aquellas variables relevantes con un porcentaje de datos faltantes considerado bajo según la literatura, donde se considera que valores menores al 5 % son bajos (Masconi et al., 2015), y se recomienda no utilizar variables con más del 50 %, y a veces ni más del 20 % de valores ausentes (Centraal Bureau voor de Statistiek, s.f.). En proceso de imputación para variables con pocos datos faltantes ($< 5\,\%$) se hace mediante el método de k vecinos más cercano.

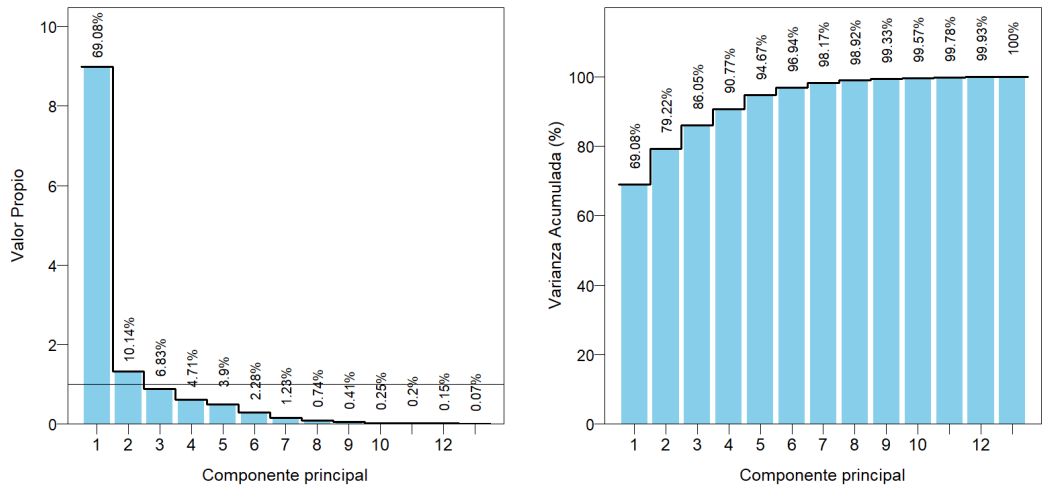

FIGURA 4.5: ACP: valores propios de cada componente (izquierda), varianza acumulada (%) explicada por los componentes (derecha)

Con base en la proporción de varianza explicada (Figura 4.5), se seleccionaron los dos primeros componentes principales. Esta selección garantiza una representación

adecuada del conjunto de datos al concentrar la mayor parte de la inercia total, es decir, la variabilidad presente en las variables originales, maximizando así la retención de información relevante y reduciendo al mínimo la pérdida de interpretabilidad.

El primer componente principal ($PC1$), presenta cargas positivas elevadas en las variables asociadas a cobertura educativa y de salud, mientras que muestra cargas negativas importantes en variables relacionadas con pobreza y carencias. Por ello, $PC1$ puede interpretarse como un eje de *bienestar y desarrollo social*, donde valores altos indican mejores condiciones educativas y sanitarias, así como menores niveles de pobreza. Así, este componente permite diferenciar con claridad a los territorios con mayor cobertura de aquellos que enfrentan mayores necesidades básicas insatisfechas.

Por su parte, el segundo componente principal ($PC2$) muestra una influencia positiva de variables como la Cobertura Educativa de Mujeres y la Cobertura Educativa Total, mientras que presenta cargas negativas en el PIB y en la Cobertura de Acueducto Urbano.Esto sugiere un eje asociado a la equidad educativa, en contraste con el nivel de desarrollo económico. En otras palabras, algunas regiones presentan una cobertura educativa femenina relativamente alta a pesar de no estar entre las más favorecidas económicamente.

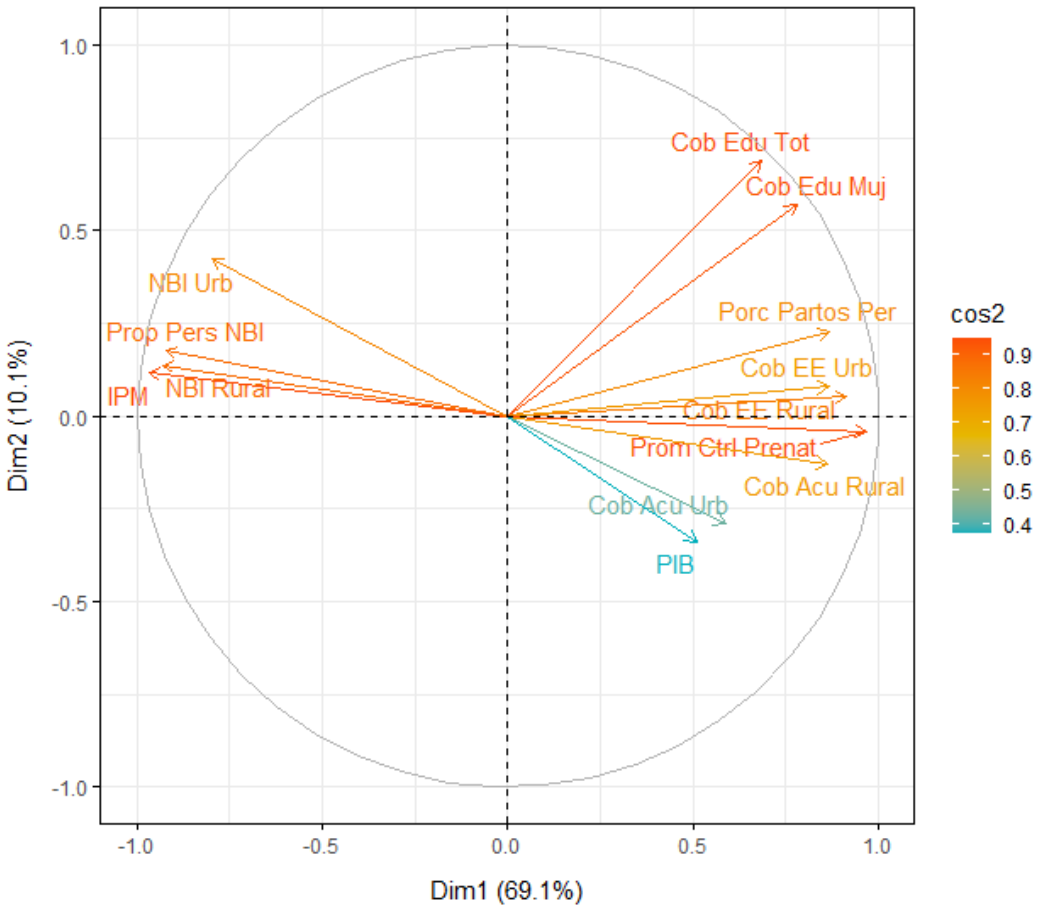

FIGURA 4.6: Circulo de correlación

El ciculo de correlación (Figura 4.6) facilita la interpretación de las relaciones entre variables y componentes, mostrando visualmente la contribución y dirección de cada variable dentro del espacio factorial.

En los dos primeros componentes principales, que explican el 79.2 % de la variabilidad total, las variables Cobertura Educativa Total, Cobertura Educativa de Mujeres, Cobertura de Acueducto Urbano y Cobertura de Educación Urbana y Rural se ubican próximas al eje positivo del Componente 1 (PC1). En contraparte, las variables NBI Rural, NBI Urbano, Proporción de Personas con NBI e IPM se proyectan en el lado negativo de PC1, en clara oposición a las coberturas, lo que refleja la contraposición esperada entre altos niveles de cobertura y mayores carencias. Esta configuración indica una alta correlación dentro de cada grupo y una contribución importante al eje asociado al bienestar y desarrollo social, distinguiendo territorios con mayor infraestructura educativa, mejores servicios públicos y menores niveles de pobreza. Ademas de estar bien representadas en los dos primeros componentes principales

Cobertura Educativa Total y de Mujeres contribuyen a ambos ejes, mientras que PIB y Cobertura de Acueducto Urbano tienen menor peso en el plano de los dos primeros componentes; como consecuencia, aportan menos a diferenciar los territorios en este plano.

En el análisis de clústeres se identifican tres grupos bien diferenciados de departamentos. Este número de clústeres fue seleccionado automáticamente por el método HCPC (Hierarchical Clustering on Principal Components), que combina clustering jerárquico y análisis de componentes principales; HCPC determina el número óptimo de grupos analizando los saltos en las alturas de fusión del dendrograma, buscando maximizar la homogeneidad dentro de los clústeres y la separación entre ellos. La elección de tres clústeres se respalda además mediante dos métodos complementarios: el gráfico de codo (WSS), que muestra un cambio notable en la pendiente al seleccionar tres grupos, y la estadística Gap, que indica que tres clústeres maximizan la separación entre grupos de manera significativa. Ambos criterios confirman que este número de clusters representa de manera eficiente y coherente la estructura del conjunto de datos (Figura 4.7)

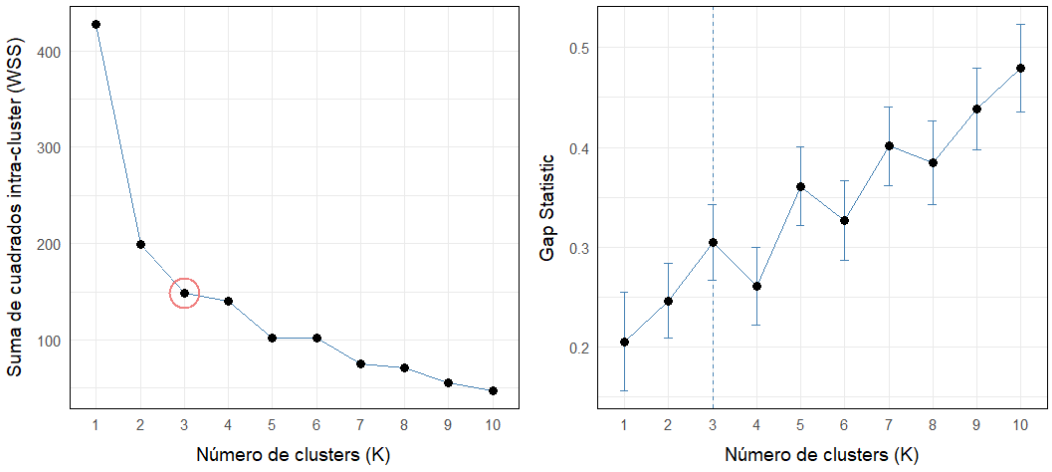

FIGURA 4.7: número óptimo de clusters mediante WSS (Izquierda) y Gap Statistic (Derecha)

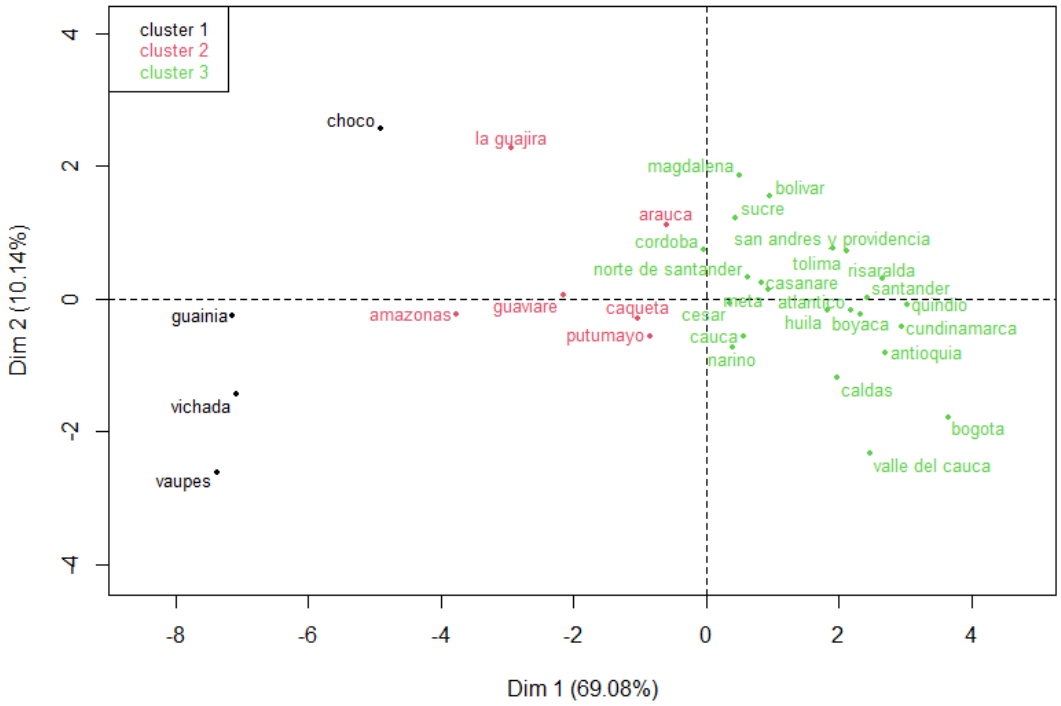

FIGURA 4.8: Mapa factorial de los departamentos según las variables sociodemográficas

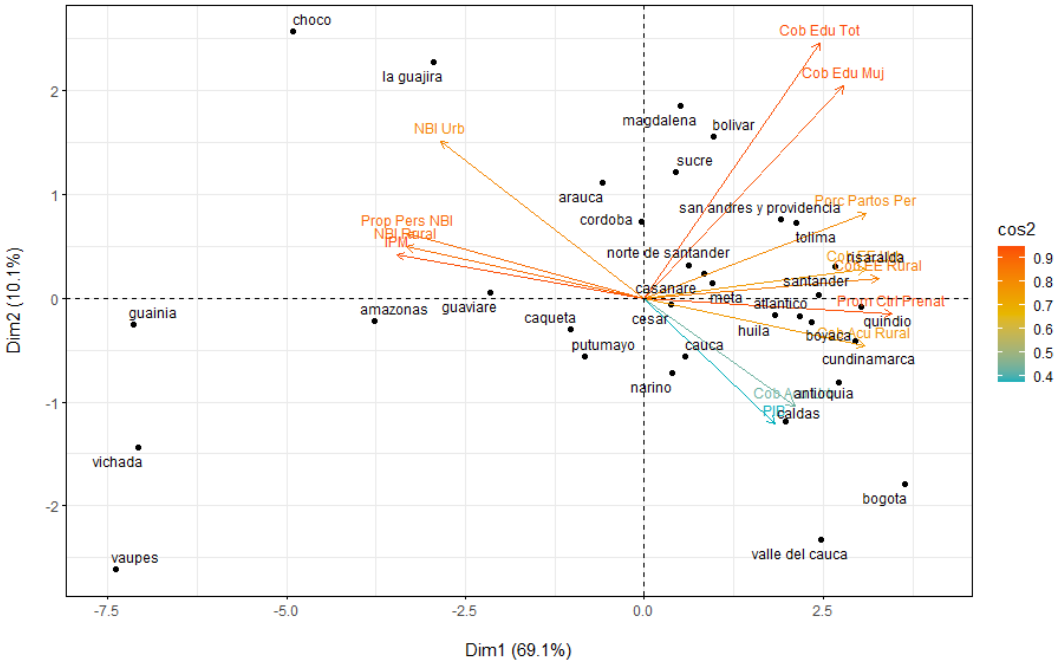

FIGURA 4.9: Biplot simultáneo de individuos (departamentos) y variables sociodemográficas

Los resultados agrupan a los departamentos en tres categorías:

- **Clúster 1 (Negro):** Vaupés, Vichada, Guainía y Chocó. territorios con rezagos extremos en acceso, cobertura y condiciones socioeconómicas, coincidiendo con las zonas más desfavorecidas señaladas por Cortés, D. and Vargas, J. F., 2012

- **Clúster 2 (Rojo):** La Guajira, Amazonas, Guaviare, Caquetá, Putumayo y Arauca. Departamentos con alta vulnerabilidad social, pobreza multidimensional significativa y limitaciones en infraestructura.
- **Clúster 3 (Verde):** Resto de departamentos, caracterizados por mejores niveles socioeconómicos, mejores coberturas educativas y sanitarias, y mayor capacidad institucional.

Este análisis multivariado refuerza la idea de que la desigualdad territorial en Colombia no es aleatoria, sino estructural y persistente, y está asociada con las condiciones que determinan los desenlaces en salud materna. Cabe mencionar que los resultados obtenidos en cuanto a los clouster planteados son cosistentes con el estudio de Cortés, D. and Vargas, J. F., 2012

4.1.3. Análisis descriptivo a nivel departamental

Tras el análisis departamental, se examina la información a nivel municipal, lo que permite capturar variabilidad intra-departamental y precisar focos territoriales de mayor o menor riesgo. Este nivel de desagregación es fundamental, dado que los departamentos presentan una marcada heterogeneidad interna en condiciones socioeconómicas, acceso a servicios de salud y variables relacionadas con infraestructura básica.

La tabla cruzada que resume, para cada departamento, el número de municipios, la media, la desviación estándar y el coeficiente de variación de cada variable de estudio se encuentra en los anexos. Este tipo de análisis es útil porque facilita comparaciones entre territorios con diferente tamaño poblacional y estructura administrativa, permitiendo interpretar las variables de manera más homogénea.

En esta caracterización destacan varias condiciones territoriales relevantes. Amazonas y San Andrés se identifican como los departamentos con menos municipios (3 y 2, respectivamente), mientras que Antioquia y Boyacá son los más grandes en términos administrativos. La RMM presenta una variabilidad particularmente alta dentro de casi todos los departamentos, evidenciada por coeficientes de variación elevados; esto sugiere desigualdades internas incluso en regiones donde la media departamental es baja.

En contraposición, las coberturas de servicios, especialmente en áreas urbanas muestran mayor homogeneidad intradepartamental, mientras que las zonas rurales evidencian una dispersión mucho mayor, lo cual coincide con las profundas brechas urbano–rurales documentadas en el país (Cortés, D. and Vargas, J. F., 2012)

A continuación, se presentan los valores faltantes de la base de datos a nivel municipal.

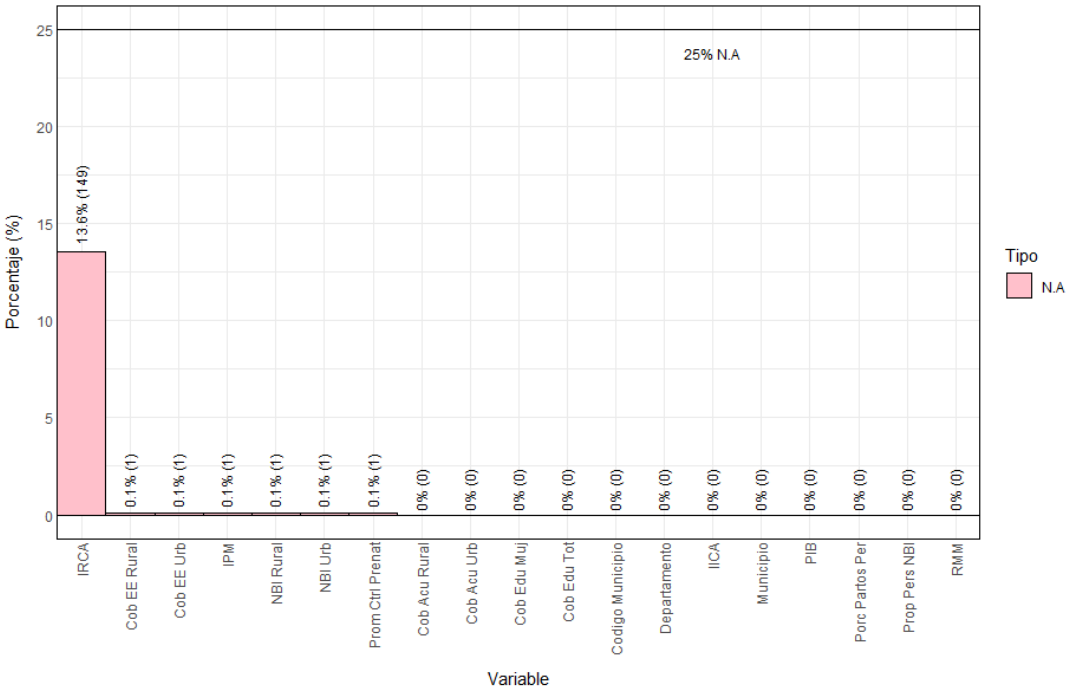

FIGURA 4.10: Porcentaje de datos faltantes a nivel departamental

En la Figura 4.10 se observa que las variables con mayor proporción de datos faltantes son IRCA (13.6 %) y, en menor medida, Longitud y Latitud (2.7 % cada una). El resto de variables presenta porcentajes inferiores al 1 %. La ausencia de información en IRCA puede influir en los análisis relacionados con la calidad del agua. La base de datos mantiene un buen nivel de completitud, lo que permite continuar con los análisis propuestos sin mayores restricciones.

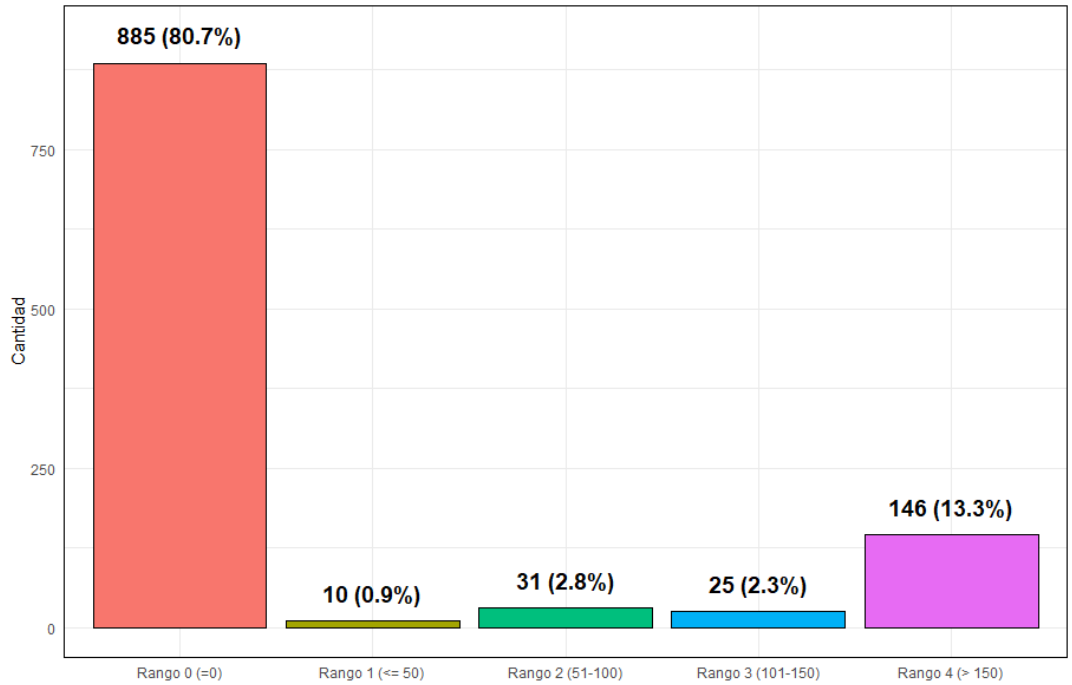

FIGURA 4.11: RMM categorizado por rangos

La Figura 4.11 muestra una marcada concentración de municipios (80.7 %) con

RMM igual a cero, lo que podría reflejar tanto una baja incidencia real de muertes maternas como posibles problemas de subregistro, especialmente en territorios con limitaciones en la vigilancia epidemiológica. En contraste, un 13.3 % presenta valores superiores a 150, evidenciando zonas de alto riesgo. Los rangos intermedios agrupan porcentajes muy bajos (0.9 %, 2.8 % y 2.3 %), lo que confirma la fuerte asimetría en la distribución de la RMM y la necesidad de enfoques focalizados en las regiones más afectadas.

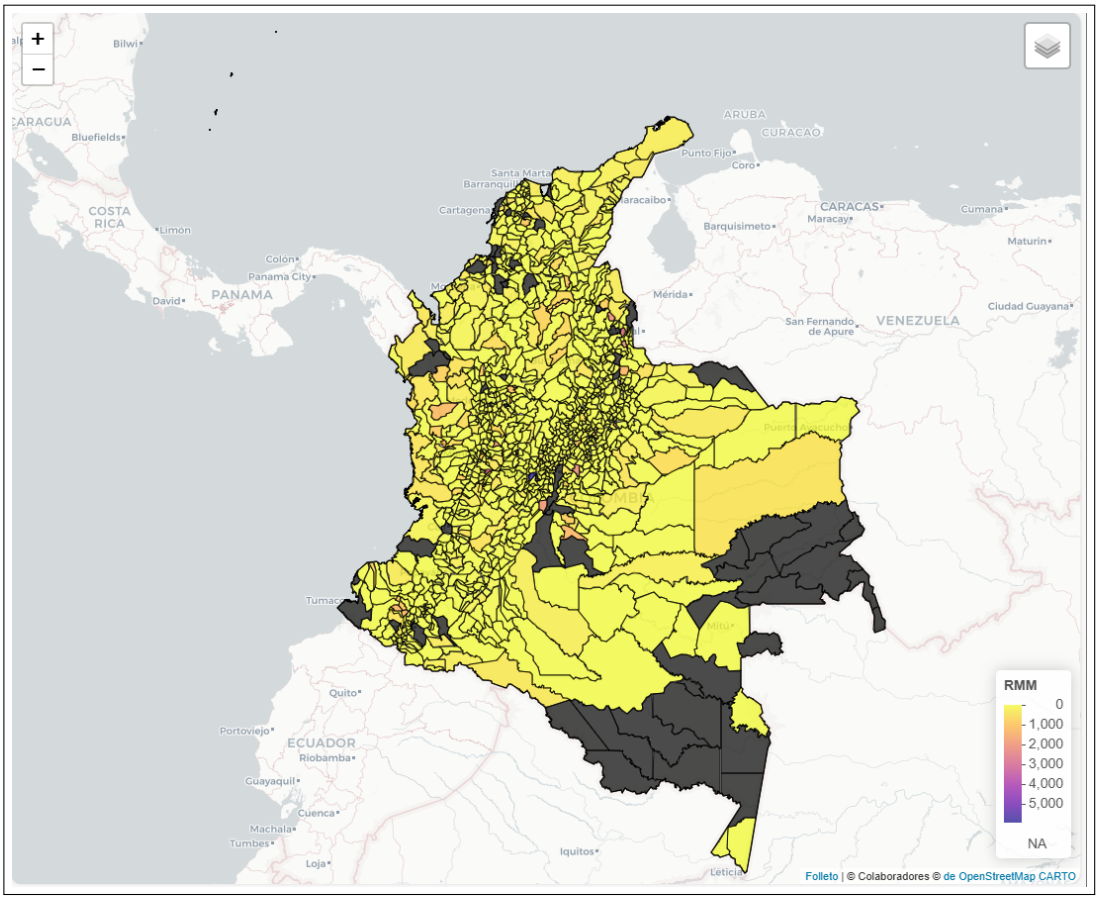

FIGURA 4.12: Mapa de calor (heatmap) de la RMM registrada por municipio

La Figura 4.12 presenta la distribución espacial de la Razón de Mortalidad Materna a nivel municipal en Colombia, evidenciando un patrón territorial caracterizado por marcadas diferencias entre regiones. En la mayoría del territorio se cuenta con información completa, lo que permite identificar valores relativamente bajos en municipios con mayor presencia urbana y mejor infraestructura institucional, mientras que los incrementos más notorios se concentran en zonas periféricas, extensas y con menor densidad poblacional. Estas variaciones son coherentes con lo documentado por el Departamento Nacional de Planeación, el cual señala la existencia de brechas persistentes entre regiones núcleo y regiones periféricas en aspectos como provisión estatal, acceso a servicios y capacidades institucionales locales (Cortés, D. and Vargas, J. F., 2012).

La presencia de valores faltantes se limita a un conjunto reducido de municipios ubicados principalmente en los departamentos amazónicos y de la Orinoquía. De acuerdo con el DNP, estas áreas se caracterizan por su baja presencia institucional,

limitaciones históricas en infraestructura administrativa y dificultades de conectividad, factores que afectan tanto la provisión de servicios públicos como la calidad y continuidad del registro estadístico (Cortés, D. and Vargas, J. F., 2012). Por ello, los municipios sin reporte deben entenderse como expresión de condiciones estructurales del territorio y no como eventos aislados asociados al fenómeno de la mortalidad materna.

En conjunto, el mapa evidencia la heterogeneidad territorial de la RMM y refuerza la importancia de incorporar estructuras jerárquicas en el análisis posterior. Las diferencias entre municipios no obedecen únicamente a la proximidad geográfica, sino a desigualdades históricas y capacidades institucionales divergentes que han configurado un patrón territorial complejo, ampliamente reconocido en los estudios sobre disparidades regionales en Colombia. Estos elementos justifican la necesidad de enfoques multinivel que permitan capturar la variabilidad intra e interdepartamental en la comprensión de la mortalidad materna.

4.1.4. Evaluación de agrupamientos territoriales y análisis comparativo

El presente apartado tiene como propósito identificar y analizar patrones de agrupamiento territorial de la Razón de Mortalidad Materna (RMM) a nivel departamental y municipal. Aunque la exploración visual mediante mapas temáticos puede sugerir posibles indicios de ausencia de patrones espaciales continuos, esta apreciación es solo preliminar. Como señalan Siabato, W. and Guzmán-Manrique, J., 2019, la interpretación visual puede indicar agrupamientos o dispersión, pero no permite establecer de manera formal la presencia o ausencia de autocorrelación espacial. Para ello, se emplean técnicas de correlación espacial y métodos de agrupamiento que permiten reconocer conjuntos de territorios con niveles similares de RMM.

En caso de no encontrarse correlación espacial, se recurrirá a las agrupaciones de departamentos obtenidas en el apartado Análisis de Componentes Principales (ACP) mediante el análisis de conglomerados jerárquicos. Estas agrupaciones reflejan la similitud estructural entre los territorios, considerando dimensiones sociodemográficas, económicas y de provisión de servicios.Esta metodologia permite definir perfiles territoriales basados en patrones latentes de desigualdad.Esta metodología permite definir perfiles territoriales basados en patrones latentes de desigualdad, los cuales se contextualizan a partir de la literatura sobre inequidad regional en Colombia, especialmente el trabajo de Cortés, D. and Vargas, J. F., 2012, quienes documentan la persistencia de regiones con rezagos estructurales profundos en acceso a bienes públicos, infraestructura y capacidad institucional.

Además, la interpretación de estos agrupamientos es fundamental para el modelo jerárquico bayesiano, ya que la variabilidad entre departamentos, incluso sin un patrón espacial, justifica el uso de interceptos aleatorios y de una estructura multinivel. Como señalan Sánchez-Cantalejo, E. and Ocaña-Riola, R., 1999, la pertinencia de los modelos jerárquicos se debe a la organización administrativa y territorial de los datos y no a su dependencia espacial.

Correlación espacial

Para identificar patrones geográficos de la RMM y posibles áreas prioritarias, se aplicaron métodos de autocorrelación espacial global y local utilizando matrices de

vecindad basadas en los cinco vecinos más cercanos. Con estas matrices se calcularon los índices globales de Moran y Geary, que permiten evaluar la presencia de asociación espacial en el conjunto de datos.

Según la definición clásica de Siabato, W. and Guzmán-Manrique, J., 2019, la correlación espacial describe la relación sistemática entre los valores de una variable y la ubicación geográfica de las unidades territoriales . Esta dependencia solo aparece cuando unidades vecinas tienden a exhibir valores similares o distintos con una frecuencia mayor a la esperada por azar.

A nivel municipal, los valores de Moran ($I = -0.01$, $p = 0.706$) y Geary ($C = 1.03$, $p = 0.699$) indican que la RMM se distribuye de manera aleatoria entre los municipios. El índice observado, cercano al índice esperado, junto con la varianza, muestra que la dispersión de los valores observados no se aleja de lo que sería esperado por azar; la puntuación z, próxima a cero, confirma la ausencia de un patrón de concentración, y el valor p evidencia que no se puede rechazar la hipótesis nula de aleatoriedad espacial completa. Esto sugiere que, aunque existen desigualdades territoriales en la RMM, estas no siguen un patrón de contigüidad geográfica, sino que dependen de factores estructurales o institucionales independientes de la vecindad física (Tabla 4.5).

Estadístico	**Municipios**		**Departamentos**	
	Moran's I	**Geary's C**	**Moran's I**	**Geary's C**
Indice observado	-0.0101	1.0338	-0.0627	0.8765
Indice esperado	-0.00094	1.0000	-0.03125	1.0000
Varianza	0.000287	0.004230	0.0076269	0.0128116
Puntuación Z	-0.5429	-0.5201	-0.3606	1.0911
Valor-p	0.7064	0.6985	0.6408	0.1376

TABLA 4.5: Pruebas globales de autocorrelación espacial para Municipios y Departamentos

Para identificar patrones locales, se aplicó el índice local de Moran (LISA), que permite detectar conglomerados estadísticamente significativos de valores altos (Alto–Alto), bajos (Bajo–Bajo) o áreas atípicas (Alto–Bajo, Bajo–Alto). Si bien se observaron algunos municipios clasificados como Alto–Alto o Alto–Bajo, la mayoría se ubicó en la categoría "No significativo", lo que coincide con la ausencia de autocorrelación espacial a nivel global (Tabla 4.6)

Alto-Alto	**Bajo-Alto**	**No significativo**
8	54	1006

TABLA 4.6: Clasificación de municipios según el estadístico Local Moran (LISA)

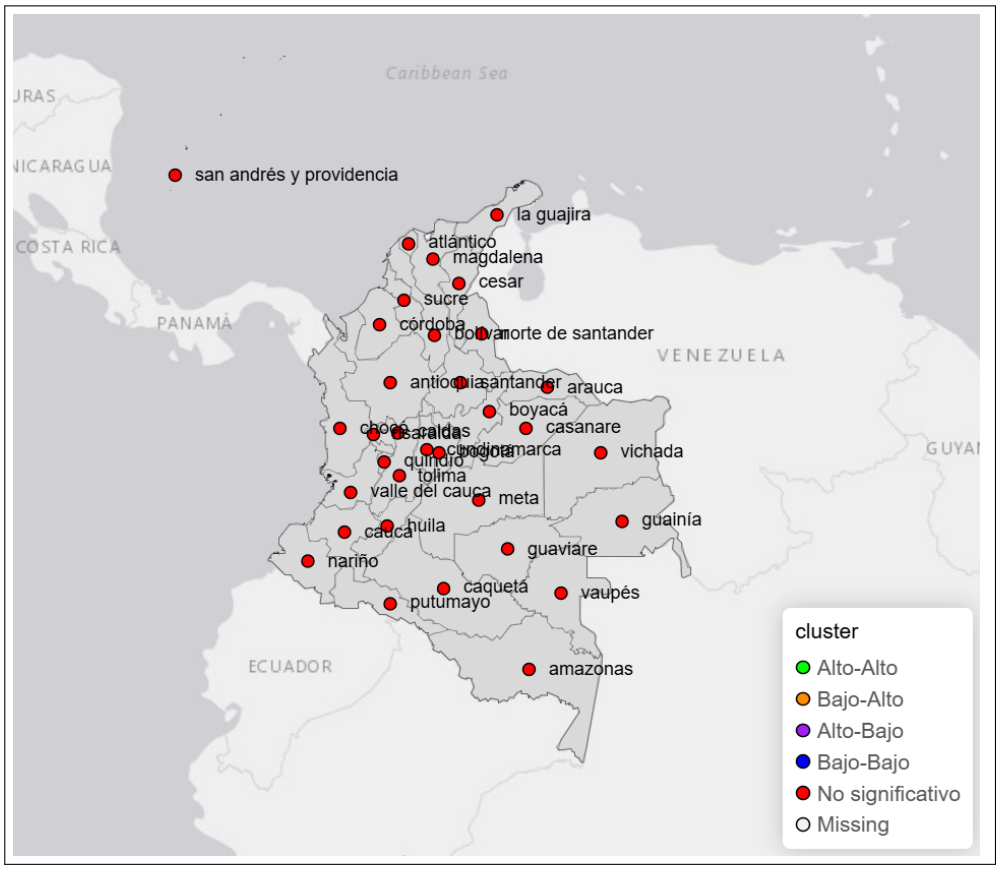

FIGURA 4.13: Mapa coroplético (choropleth map) - Departamentos

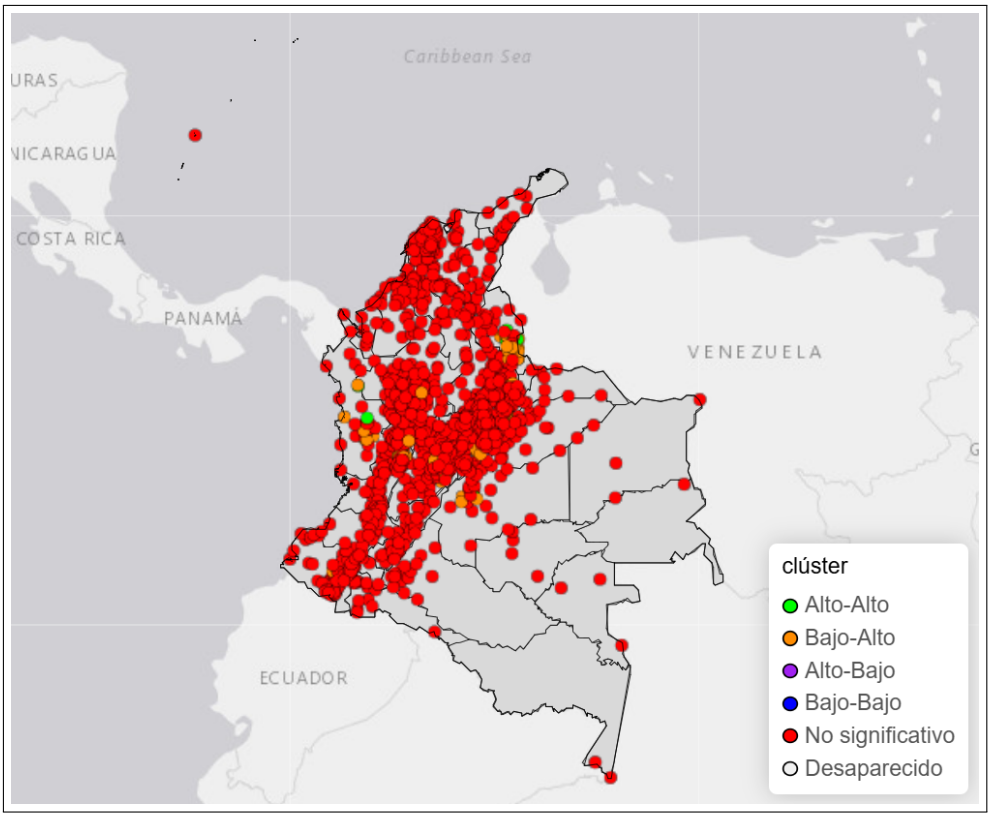

FIGURA 4.14: Mapa coroplético (choropleth map) - Municipios

Implicaciones de la independencia espacial en la estructura del modelo

Desde el punto de vista metodológico y epidemiológico, la independencia espacial tiene tres implicaciones fundamentales:

1. **Los clusters no implican autocorrelación espacial.** La ausencia de autocorrelación global indica que la similitud entre departamentos no depende de su proximidad geográfica. Por lo tanto, los clústeres identificados reflejan similitudes estructurales y no dependencia espacial. Como señalan Siabato, W. and Guzmán-Manrique, J., 2019, la autocorrelación espacial implica una relación entre los valores y su localización contigua, algo que los métodos de clustering no consideran.

2. **La variabilidad territorial es real, aunque no espacialmente dependiente.** Los departamentos pueden diferir significativamente en pobreza, infraestructura, acceso educativo o capacidad institucional, pero estas diferencias no siguen necesariamente la proximidad geográfica. Esto concuerda con las desigualdades regionales reportadas por Cortés, D. and Vargas, J. F., 2012

3. **La ausencia de autocorrelación no afecta la justificación del modelo jerárquico.**La estructura multinivel se basa en la organización territorial (municipios dentro de departamentos) y no en la proximidad geográfica.

Como se estableció en la Sección 3.3.2, los efectos aleatorios se modelaron mediante distribuciones a priori Gaussianas, y según lo comprobado en la Sección 4.1.4, no existe dependencia espacial ni a nivel departamental ni municipal. Por tanto, el modelo bayesiano no requiere incluir un término de dependencia espacial, lo que permite capturar la variabilidad interdepartamental e intradepartamental sin imponer una estructura espacial no respaldada por los datos.

De este modo, las desigualdades regionales persisten, aunque no siguen una lógica estrictamente espacial, sino que responden a factores institucionales, socioeconómicos y de capacidad territorial, lo que genera una heterogeneidad territorial que justifica la implementación de un modelo multinivel. Según Sánchez-Cantalejo, E. and Ocaña-Riola, R., 1999, los datos en salud pública suelen presentar una estructura jerárquica, en la que los municipios se organizan dentro de departamentos, lo que hace recomendable modelar la variabilidad inter e intradepartamental mediante efectos aleatorios.

Aunque la independencia espacial ya fue discutida en la sección anterior, estos hallazgos permiten precisar la jerarquía y los términos a incluir:

- La independencia espacial observada confirma que no es necesario incluir efectos de autocorrelación espacial como CAR (Conditional Autoregressive) o BYM (Besag-York-Mollié), que se aplican cuando existe dependencia geográfica entre áreas vecinas.

- La variabilidad significativa entre departamentos respalda la incorporación de interceptos aleatorios a nivel departamental.

- La heterogeneidad dentro de cada departamento justifica la inclusión del nivel municipal en la jerarquía del modelo.

Caracterización de los grupos de departamentos

Las agrupaciones departamentales se definieron a partir de los patrones de similitud observados en los factores sociodemográficos, identificados mediante el análisis de clúster sobre componentes principales (HCPC), que combina reducción de dimensionalidad (PCA) con agrupamiento jerárquico (Figura 4.8). Este enfoque permite formar grupos de departamentos con perfiles similares en múltiples variables, facilitando la identificación de áreas con características sociodemográficas y sanitarias comunes.

Los clústeres (Figura 4.8,Tabla 4.7) reflejan diferencias en los niveles de RMM y en factores estructurales como cobertura de servicios públicos (acueducto, energía), acceso a educación, pobreza multidimensional (NBI, IPM), condiciones ambientales (IRCA) y atención en salud materna (control prenatal y partos atendidos por personal capacitado), sintetizando cómo estos elementos caracterizan a cada departamento.

Variable	Cluster 1		Cluster 2		Cluster 3	
	Media (sd)	Mediana (p25,p75)	Media (sd)	Mediana (p25,p75)	Media (sd)	Mediana (p25,p75)
RMM	222 (102)	258 (198,282)	108 (62.6)	116 (95.8,131)	65.2 (30.1)	63.5 (50,90.2)
Cob Edu Tot	80.0 (15.5)	75.9 (72,83.8)	97.9 (8.16)	95.1 (91.9,103)	101 (7.2)	102 (98,105)
Cob Acu Rural	9.70 (8.28)	6.52 (5.49,10.7)	18.2 (8.74)	14.9 (12.7,24.4)	58.0 (14.5)	57.8 (45,70)
Cob Acu Urb	59.3 (28.6)	58.7 (34.6,83.4)	77.3 (15.1)	75.2 (68.2,89.5)	91.0 (13.5)	95.0 (90,98)
Cob EE Rural	29.8 (18.4)	27.2 (17.2,39.9)	51.3 (18.8)	52.6 (41.2,56.0)	88.5 (8.5)	90.0 (82,95)
Cob EE Urb	91.6 (5.23)	92.4 (89.5,94.5)	95.7 (1.98)	96.5 (95.1,96.9)	98.9 (0.5)	99.0 (98.5,99.5)
Cob Edu Muj	69.0 (12.0)	67.6 (64.6,72.0)	86.7 (8.61)	84.0 (81.5,92.2)	92.0 (6.0)	93.0 (90,95)
NBI Rural	78.3 (10.9)	82.2 (75.1,85.4)	46.6 (15.8)	45.5 (40.6,47.3)	26.0 (12.0)	23.5 (17,36)
NBI Urb	42.8 (17.5)	36.1 (32.0,46.9)	20.7 (6.96)	21.1 (15.8,25.7)	12.0 (7.0)	10.0 (6,18)
IPM	72.9 (7.28)	75.2 (70.9,77.2)	48.6 (8.97)	45.3 (42.7,56.1)	30.0 (11.0)	27.0 (21,40)
Prom Ctrl Prenat	2.29 (0.552)	2.13 (1.89,2.53)	4.21 (0.710)	4.35 (3.67,4.73)	5.75 (0.50)	5.70 (5.40,6.10)
Porc Partos Per	67.4 (12.3)	66.6 (60.5,73.5)	91.0 (6.72)	93.4 (93.3,94.0)	96.5 (3.0)	97.5 (96,99)
PIB	2232 (3118)	806 (523,2515)	7275 (7057)	6042 (2408,8120)	68000 (80000)	33000 (25000,85000)
Prop Pers NBI	65.4 (4.19)	66.6 (64.0,68.1)	31.8 (12.0)	29.7 (24.7,34.3)	17.0 (11.0)	13.0 (9,22)
RMM plot	222 (102)	258 (198,282)	108 (61.5)	116 (95.8,131)	66.0 (31.0)	63.5 (50,90)

TABLA 4.7: Estadísticas descriptivas por cluster (Clusters 1, 2 y 3)

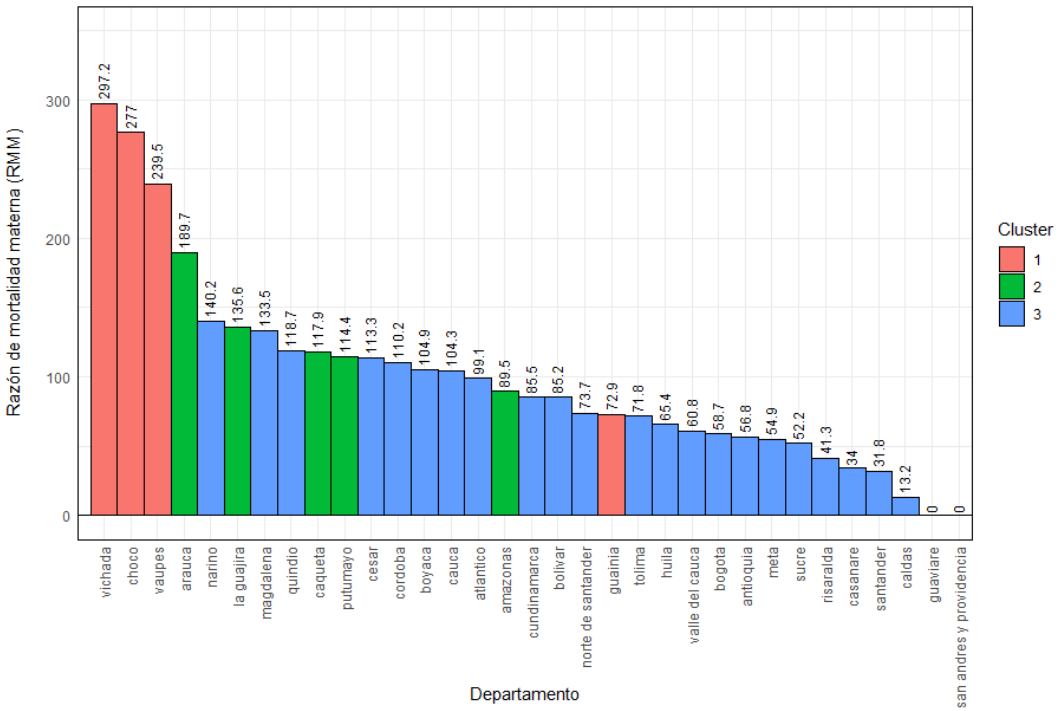

FIGURA 4.15: RMM por Departamento con Identificación de Clústeres

El primer clúster se distingue por condiciones socioeconómicas menos favorables en comparación con los otros grupos. Los departamentos incluidos presentan bajas coberturas de educación en todos los niveles considerados (total, urbana y rural), así como menor acceso a servicios públicos esenciales, como acueducto y energía eléctrica.Asimismo, muestran altos niveles de Necesidades Básicas Insatisfechas (NBI) tanto en áreas urbanas como rurales, junto con un elevado Índice de Pobreza Multidimensional (IPM), lo que evidencia carencias estructurales significativas en salud, educación y vivienda. El acceso a educación femenina también es reducido y la proporción de partos atendidos por personal capacitado es menor, lo que indica limitaciones en la atención de salud materna. En términos económicos, estos departamentos presentan un PIB per cápita relativamente bajo y, de manera consistente con las demás variables de vulnerabilidad, registran los valores más altos de NBI y de IPM entre los tres clústeres.

En el gráfico de clústeres (Figura 4.15) se observa que los departamentos del segundo y tercer clúster presentan una cierta superposición visual, lo que indica que algunos de sus valores en las variables socioeconómicas y de servicios son cercanos o similares.

El segundo clúster agrupa departamentos con condiciones socioeconómicas intermedias, consistentes en todas las variables consideradas. Aunque en algunos casos la diferencia entre los clústeres 2 y 3 es mínima, lo que hace que ambos grupos no sean tan diferentes en ciertos indicadores, Aunque no destacan por condiciones particularmente desfavorables ni especialmente favorables, este clúster sirve como punto intermedio que permite diferenciar claramente los departamentos con mayores y menores ventajas socioeconómicas

El tercer clúster agrupa departamentos con condiciones socioeconómicas y de servicios notablemente más favorables. Estos departamentos presentan altas coberturas de educación y servicios públicos en áreas rurales y urbanas, bajos niveles de NBI e IPM, y un acceso significativamente mejor a la educación femenina y a la atención de salud materna, reflejado en promedios elevados de controles prenatales y en una proporción cercana al 100 % de partos atendidos por personal calificado. En comparación con el primer clúster, la brecha en términos de vulnerabilidad social y económica es clara: mientras que el primer clúster concentra las mayores carencias, el tercer clúster representa regiones con capacidades más altas de acceso a servicios y mejores condiciones de vida.

Al analizar la Razón de Mortalidad Materna (RMM) por clúster, se observa que, en términos generales, los departamentos con mejores condiciones socioeconómicas tienden a presentar valores más bajos de mortalidad materna, mientras que los departamentos más vulnerables muestran valores más altos y dispersos, aunque existe una variabilidad relativa dentro de cada grupo.

4.1.5. Síntesis del Resultado Uno

Los resultados muestran que no se identifican patrones espaciales evidentes en la distribución geográfica de la RMM en Colombia, según la inspección de los mapas temáticos y, principalmente, los resultados de las pruebas estadísticas aplicadas. En cambio, el análisis descriptivo, territorial y multivariado evidencia desigualdades marcadas, con mayores niveles en departamentos históricamente rezagados como Vaupés, Vichada, Chocó y Guainía, en línea con lo documentado por Cortés, D. and

Vargas, J. F., 2012 sobre las brechas regionales en acceso, pobreza y capacidad institucional. Esta heterogeneidad territorial indica que los departamentos vecinos no necesariamente presentan niveles similares de RMM.

Finalmente, es importante resaltar que la exploración espacial realizada no determina por sí misma la elección de un modelo multinivel. Como señalan Sánchez-Cantalejo, E. and Ocaña-Riola, R., 1999, la dependencia jerárquica proviene de la propia organización administrativa del país: los municipios están anidados dentro de los departamentos y comparten estructuras institucionales, redes de servicios, sistemas de gobernanza y condiciones de provisión en salud. Esta estructura genera variabilidad tanto dentro como entre departamentos, lo que hace necesario modelarla explícitamente.

4.2. Resultado 2: Modelado bayesiano de la Razón de Mortalidad Materna

En este apartado aborda el Objetivo 3 del estudio: modelar la Razón de Mortalidad Materna (RMM) mediante métodos bayesianos, considerando la estructura multinivel de los datos e identificando patrones y diferencias significativas tanto dentro como entre los distintos niveles territoriales.

Los análisis descriptivos, espaciales y multivariados desarrollados en los Resultados 1 y 2 evidenciaron tres elementos fundamentales para la elección del modelo:

1. **Exceso de ceros en la RMM**: La distribución de la RMM presenta una asimetría marcada, con muchos municipios con valor cero. Esto motiva el uso de un esquema de modelación tipo *hurdle*, que permite separar la ausencia de eventos de la magnitud de la RMM cuando ocurren.

2. **Heterogeneidad territorial y ausencia de autocorrelación espacial:** Existe heterogeneidad territorial consistente con la estructura jerárquica de los datos (municipios anidados dentro de departamentos). Esto respalda la necesidad de un modelo multinivel

3. **Diferencias estructurales entre territorios:** Los análisis de componentes principales (ACP) y clústeres mostraron diferencias estructurales entre territorios que justifican la inclusión de efectos aleatorios a nivel departamental y municipal.

De acuerdo con los criterios de información (DIC) y WAIC, el modelo con distribución Log-normal presentó un mejor ajuste en comparación con el modelo Gamma, por lo que se seleccionó como la especificación final para la parte continua del modelo.

Modelo	**DIC**	**WAIC**
Modelo log-Normal	606.85	610.28
Modelo Gamma	3157.73	3171.88

TABLA 4.8: Indicadores de información para los modelos ajustados (Parte continua)

A partir de estos hallazgos, se ajustó un modelo hurdle multinivel bayesiano empleando muestreo MCMC mediante el algoritmo No-U-Turn Sampler (NUTS), una extensión del método Hamiltonian Monte Carlo (HMC) implementado en el lenguaje probabilístico. El modelo se compone de dos partes: (i) un modelo logístico para modelar la probabilidad de registrar una RMM igual a cero, y (ii) un modelo lognormal para los valores positivos de la RMM. Ambas ecuaciones (3.2, 3.4) incorporan efectos aleatorios independientes para departamentos y municipios, coherentes con la ausencia de dependencia espacial detectada en los Resultados 1 y 2, pero acordes con la heterogeneidad jerárquica entre y dentro de los territorios. `Stan`.

El presente resultado se organiza en cuatro bloques:

1. **Diagnóstico y convergencia del muestreo MCMC:** Se presentan los indicadores *Rhat*, *ESS* y la inspección visual mediante traceplots y densidades posteriores.

2. **Estimación de los parámetros principales del model:** incluye los interceptos, las desviaciones estándar de los efectos aleatorios y la varianza residual.

3. **Efectos aleatorios jerárquicos:** Permiten identificar diferencias intra- e inter-departamentales en la distribución de la RMM.

4. **Predicciones posteriores:** Se presentan los valores esperados, los intervalos creíbles y la comparación con los valores observados.

Este enfoque integral permite caracterizar de manera simultánea la presencia o ausencia de mortalidad materna, la magnitud de la RMM cuando esta ocurre y las desigualdades territoriales subyacentes, ofreciendo un marco sólido para la interpretación epidemiológica y la toma de decisiones en salud pública.

4.2.1. Diagnóstico de supuestos: Indenpendencia condicional y linealidad

En la parte discreta del modelo hurdle multinivel, la ocurrencia de al menos una muerte materna sigue una distribución Bernoulli (3.1), y su probabilidad se vincula con las covariables mediante un enlace logit (3.2) y efectos aleatorios departamentales y municipales.

Linealidad en el logit (parte Bernoulli)

El supuesto de linealidad en el logit implica que la relación entre las covariables y el logaritmo de los odds ratio (OR) se representa adecuadamente mediante una combinación lineal en la escala del predictor. Para evaluar la adecuación de esta especificación, se construyó un gráfico de residuos agrupados (binned residual plot), en el que se representan los residuos $Z_i - \hat{p}_i$ frente al predictor lineal posterior medio. Bajo una especificación funcional correcta, los residuos promedio deben oscilar alrededor de cero sin mostrar patrones sistemáticos ni curvaturas evidentes.

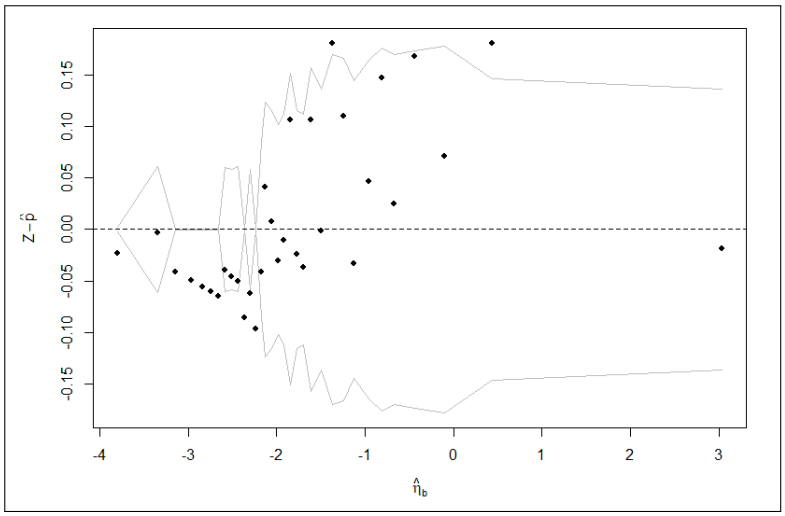

FIGURA 4.16: Linealidad en el logit: Residuos agrupados (Binned Residuals)

La figura 4.16 mostra que los puntos se distribuyen de manera aproximadamente simétrica alrededor de la línea horizontal en cero, sin evidenciar tendencias monótonas ni estructuras no lineales pronunciadas a lo largo del rango del predictor.

Aunque se observa cierta dispersión en algunas regiones, no se identifican patrones consistentes que sugieran una forma funcional alternativa.En consecuencia, el diagnóstico indica que la linealidad en el predictor se cumple para la parte Bernoulli del modelo.

Independencia condicional (parte Bernoulli)

Bajo la estructura jerárquica especificada, se asume que las observaciones son independientes condicionalmente a las covariables y a los efectos aleatorios departamentales y municipales, es decir, $Z_i \perp Z_j \mid X, u_{dep}, u_{mun}$. En el marco bayesiano, este supuesto no se interpreta como una condición para la validez asintótica, sino como una coherencia del modelo probabilístico especificado. Para evaluar su plausibilidad, se calcularon los residuos promedio por departamento y se representaron frente al promedio de las probabilidades predichas.

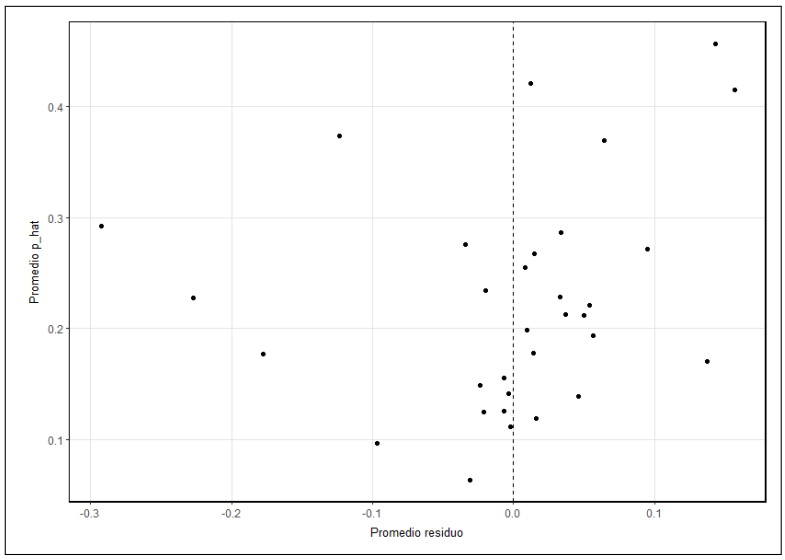

FIGURA 4.17: Independencia condicional por departamento

La distribución de los residuos agregados no evidencia estructuras sistemáticas ni tendencias funcionales respecto al nivel promedio de riesgo estimado, lo que es consistente con la hipótesis de independencia condicional una vez incorporados los efectos jerárquicos. (4.17)

Linealidad en el logit (parte Lognormal)

Para los municipios con razón de mortalidad materna positiva se asumió que la variable transformada en escala logarítmica sigue una distribución normal condicional (3.3), con un predictor lineal definido en (3.4). Este supuesto implica linealidad en el predictor continuo y normalidad condicional en la escala logarítmica, y se evaluó mediante un gráfico cuantílico–cuantílico de los residuos estandarizados.

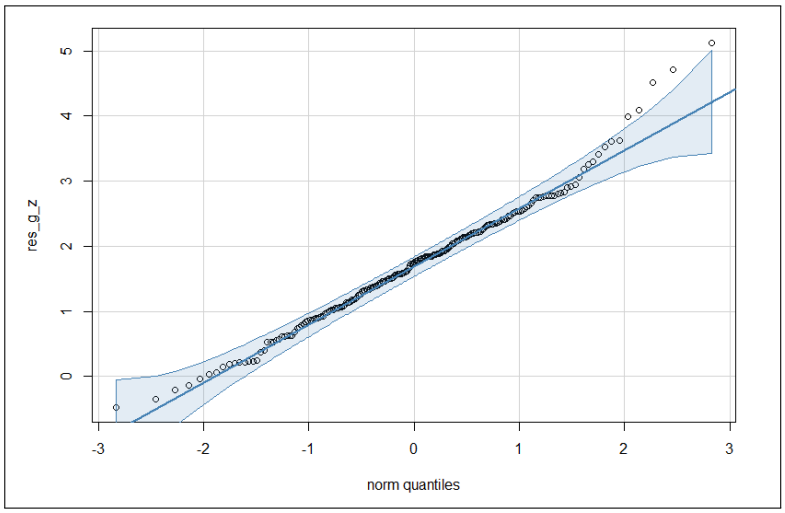

FIGURA 4.18: Linealidad del logit: QQ-plot

El gráfico Q–Q de los residuos estandarizados muestra un buen ajuste en la región central de la distribución, con desviaciones leves en las colas, lo que indica que la suposición de normalidad condicional en la escala logarítmica se cumple de manera razonable. (4.18)

Independencia condicional (parte Lognormal)

En la componente positiva del modelo se asume que, condicionando en las covariables y en los efectos aleatorios, las observaciones son independientes entre departamentos, es decir, $y_i \perp y_j \mid X, u_{dep}, u_{mun}$. Para ello, se calcularon los residuos estandarizados y se representó su promedio por departamento frente al promedio de las predicciones.

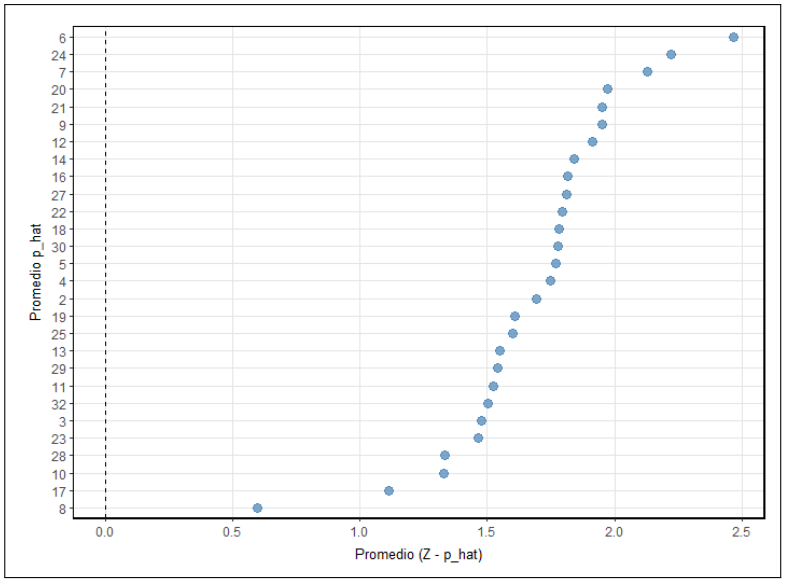

FIGURA 4.19: Independencia condicional por departamento

Los residuos promedio presentan un desplazamiento positivo respecto a cero, lo que indica que la media condicional estimada tiende a subestimar los valores observados en la componente positiva de la razón de mortalidad materna. Dado que

el intercepto presenta adecuada convergencia y precisión posterior, este comportamiento se interpreta como posible limitación en la especificación funcional o en la forma distributiva adoptada para la componente continua. Este comportamiento no implica necesariamente dependencia estructural entre las observaciones, sino que podría reflejar limitaciones en la especificación funcional o en la forma distributiva adoptada para la parte continua. (4.19)

En el marco bayesiano, este resultado sugiere que podría ser conveniente explorar especificaciones más flexibles, como distribuciones con colas más pesadas o estructuras de varianza alternativas, como posibles mejoras del modelo en futuras extensiones del análisis.

4.2.2. Diagnóstico y convergencia del muestreo MCMC

Previo a la interpretación de los parámetros del modelo hurdle multinivel, es fundamental evaluar la calidad del muestreo MCMC y verificar que las cadenas generadas por `Stan` hayan alcanzado convergencia, es decir, que las muestras representen de manera adecuada la distribución posterior de los parámetros. Para este propósito se emplearon los indicadores estándar de la literatura bayesiana, incluyendo el factor de Gelman–Rubin (*Rhat*), el tamaño efectivo de muestra (*ESS*) y la inspección visual mediante traceplots y densidades posteriores.

Para la estimación posterior se utilizó el algoritmo Hamiltonian Monte Carlo con el adaptador No-U-Turn Sampler (NUTS), implementado en `Stan`. El muestreo se llevó a cabo mediante cuatro cadenas independientes ejecutadas en paralelo. Cada cadena se corrió con un total de 15 000 iteraciones, de las cuales las primeras 5 000 correspondieron al periodo de calentamiento (*warm-up* o *burn-in*), destinado a la adaptación automática de los parámetros del muestreador, tales como el tamaño de paso y la matriz de masa. Las 10 000 iteraciones restantes fueron conservadas para la inferencia, dando lugar a un total de 40 000 muestras posteriores combinadas. Adicionalmente, se empleó un valor de `adapt_delta` igual a 0.995 y una profundidad máxima del árbol de 15, con el fin de reducir la probabilidad de divergencias y mejorar la exploración del espacio paramétrico en un modelo jerárquico, conforme a las recomendaciones metodológicas para HMC en modelos bayesianos multinivel (Gelman, A., 2013).

La Tabla 4.9 presenta las estimaciones posteriores de los parámetros principales del modelo hurdle multinivel, que consta de una parte binaria (b) y una parte gaussiana (g). Se incluyen los interceptos (α_0, μ_0), las desviaciones estándar de los efectos aleatorios a nivel de departamento ($\sigma^{(b)}_{\text{dep}}$, $\sigma^{(g)}_{\text{dep}}$) y a nivel de municipio ($\sigma^{(b)}_{\text{mpio}}$, $\sigma^{(g)}_{\text{mpio}}$), así como la desviación estándar residual de la parte gaussiana (σ_y), junto con los indicadores de convergencia de las cadenas MCMC.

En términos de autocorrelación, la eficiencia del muestreo puede evaluarse indirectamente a través del tamaño efectivo de muestra. Los valores elevados de ESS_{bulk} y ESS_{tail} observados para todos los parámetros (Tabla 4.9) indican una baja dependencia serial entre iteraciones consecutivas y una exploración eficiente tanto de la región central como de las colas de la distribución posterior. Esto sugiere que el algoritmo HMC–NUTS logró una adecuada mezcla de las cadenas, incluso en presencia de una estructura jerárquica compleja y de un modelo de dos componentes.

Asimismo, la concordancia entre las estimaciones obtenidas a partir de múltiples cadenas independientes, reflejada en valores de $\widehat{R}$ cercanos a la unidad, respalda

la estabilidad del muestreo y la ausencia de problemas de convergencia local. La combinación de altos tamaños efectivos de muestra y factores de Gelman-Rubin ($\widehat{R}$) satisfactorios proporciona evidencia consistente de que la inferencia posterior no está dominada por la autocorrelación ni por un muestreo ineficiente. (Tabla 4.9)

Finalmente, aunque no se realiza un análisis formal de sensibilidad, la estabilidad de los diagnósticos MCMC y la ausencia de patologías numéricas sugieren que los resultados del modelo no son altamente sensibles a la especificación adoptada. En particular, el uso de prioris débiles y regularizadas en los parámetros jerárquicos, junto con una parametrización no centrada, contribuye a una inferencia robusta y coherente con la estructura de los datos. En este sentido, los diagnósticos presentados respaldan la fiabilidad de los resultados obtenidos bajo la especificación del modelo considerada.

Parámetro	Media posterior	SD	q5	Mediana	q95	Rhat	ESS_bulk	ESS_tail
α_0	-1.280	0.176	-1.580	-1.270	-1.020	1.00	10752	8894
μ_0	5.650	0.103	5.470	5.650	5.810	1.00	28920	29885
$\sigma^{(b)}_{\text{dep}}$	0.400	0.163	0.131	0.395	0.673	1.00	10347	10778
$\sigma^{(b)}_{\text{mpio}}$	0.380	0.320	0.0273	0.298	1.010	1.00	4222	5178
$\sigma^{(g)}_{\text{dep}}$	0.365	0.109	0.192	0.362	0.548	1.00	9991	10243
$\sigma^{(g)}_{\text{mpio}}$	0.0400	0.0305	0.00302	0.0336	0.0993	1.00	27419	20631
σ_y	0.754	0.0427	0.688	0.752	0.828	1.00	30027	26809

TABLA 4.9: Diagnóstico de convergencia para los parámetros principales del modelo hurdle multinivel.

Los valores de *Rhat* fueron exactamente 1.00 para todos los parámetros, lo cual indica una perfecta convergencia entre cadenas, sin señales de divergencia ni mezcla deficiente. Asimismo, los valores de *ESS_bulk* y *ESS_tail* superan ampliamente el umbral recomendado de 400, reflejando un muestreo eficiente y una exploración adecuada de la distribución posterior, incluso en componentes jerárquicos más complejos como los efectos municipales.

Inspección visual del muestreo

Además de los indicadores numéricos de convergencia, se evaluó la estabilidad de las cadenas MCMC mediante inspección visual de los traceplots, las densidades posteriores y los intervalos creíbles al 90 %. Las Figuras 4.20, 4.21 y 4.22 muestran que:

- Las cadenas MCMC presentan trayectorias estables a lo largo de las iteraciones, explorando de manera repetida la misma región del espacio paramétrico y sin mostrar tendencias sistemáticas, cambios abruptos o dependencia persistente de los valores iniciales. Este comportamiento indica una adecuada mezcla entre cadenas y sugiere que el muestreo ha alcanzado un régimen estacionario, condición necesaria para que las muestras sean representativas de la distribución posterior.
- Las densidades posteriores son suaves, unimodales y consistentes con los intervalos de credibilidad obtenidos numéricamente.

- Los intervalos creíbles al 90 % muestran distribuciones posteriores regulares, sin multimodalidad ni colas atípicas.

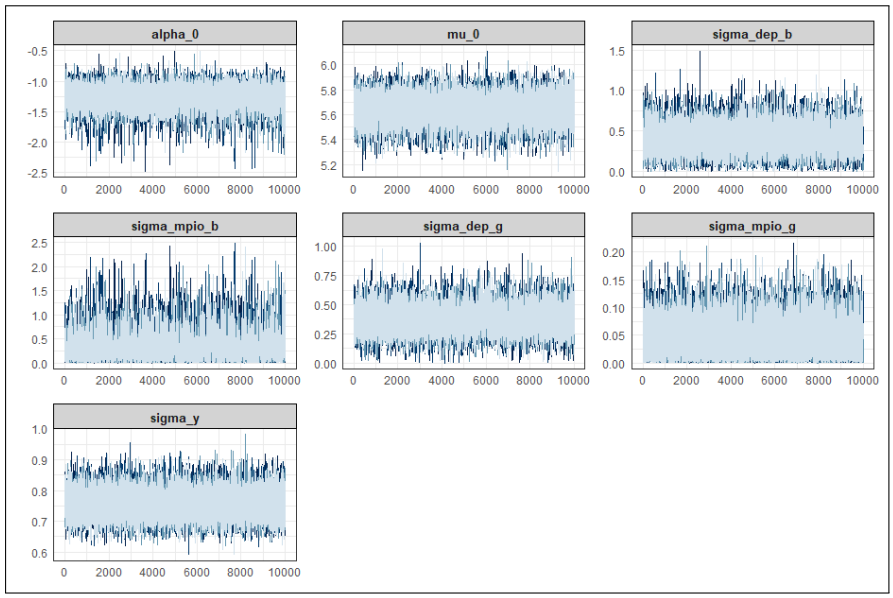

FIGURA 4.20: Traceplots de los parámetros principales del modelo hurdle multinivel.

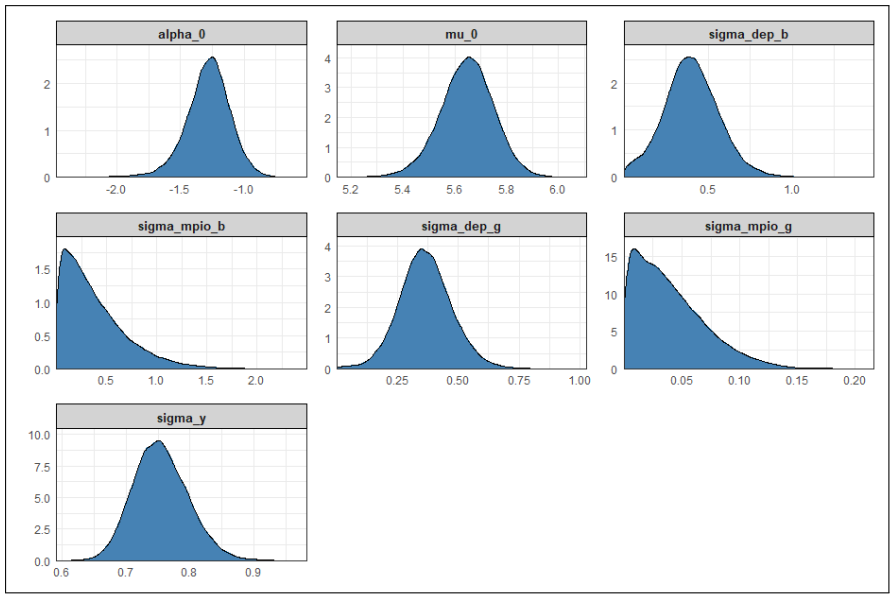

FIGURA 4.21: Densidades posteriores de los parámetros principales del modelo hurdle multinivel

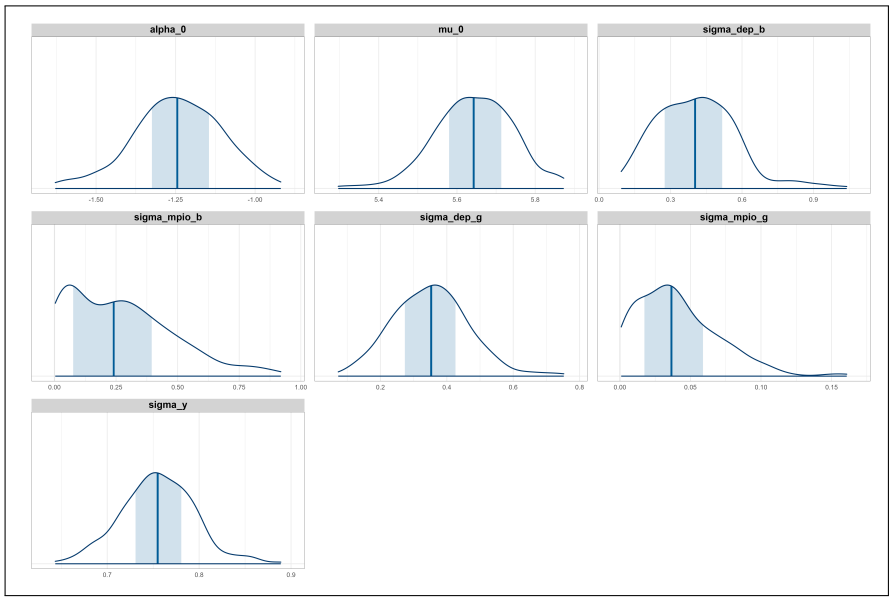

FIGURA 4.22: Áreas de credibilidad al 90 % de los parámetros principales del modelo hurdle multinivel

En conjunto, los resultados del diagnóstico MCMC indican que el proceso de muestreo alcanzó un comportamiento estable y bien mezclado, lo que permite sustentar que las inferencias presentadas en las secciones siguientes se basan en muestras representativas de la distribución posterior del modelo hurdle multinivel.

Parámetros de referencia del modelo hurdle

Una vez verificada la convergencia y estabilidad del muestreo MCMC, se procede a la interpretación de los parámetros de referencia del modelo hurdle multinivel. En particular, se analizan los parámetros α_0 y μ_0, correspondientes a las componentes Bernoulli y lognormal, respectivamente. Estos parámetros representan el nivel medio del predictor lineal condicionado a la especificación del modelo y a las covariables tal como fueron incorporadas, con los efectos aleatorios jerárquicos centrados en cero.

El parámetro α_0 corresponde al término de referencia de la componente binaria (Bernoulli) del modelo y controla la probabilidad de que un municipio registre una Razón de Mortalidad Materna mayor que cero. La estimación posterior $\hat{\alpha}_0 = -1.28$ se traduce en una probabilidad promedio aproximada del 22 %, calculada mediante la función logística. Este resultado es consistente con la elevada proporción de municipios que no reportan muertes maternas durante el período de estudio, sin que ello implique la ausencia de riesgo subyacente.

$$\frac{1}{1+\exp(-\alpha_0)} \tag{4.1}$$

Por su parte, el parámetro μ_0 corresponde al término de referencia de la componente continua del modelo, asociada a la distribución lognormal de la RMM condicionada a ser positiva. La estimación posterior $\hat{\mu}_0 = 5.65$ indica que, entre los municipios con mortalidad materna reportada, el valor esperado de la RMM se sitúa alrededor de $\exp(\mu_0) \approx 280$–290 muertes por cada 100 000 nacidos vivos, una vez consideradas las covariables incluidas y la estructura jerárquica del modelo.

Variabilidad jerárquica y residual

La Tabla 4.10 resume la variabilidad entre y dentro de los departamentos. Estas desviaciones estándar describen cuánto difieren los territorios entre sí en ambas partes del modelo hurdle.

Parámetro	Media posterior	SD	q5	q95
$\sigma^{(b)}_{\text{dep}}$	0.400	0.163	0.131	0.673
$\sigma^{(b)}_{\text{mpio}}$	0.380	0.320	0.027	1.010
$\sigma^{(g)}_{\text{dep}}$	0.365	0.109	0.192	0.548
$\sigma^{(g)}_{\text{mpio}}$	0.040	0.031	0.003	0.099
σ_y	0.754	0.043	0.688	0.828

TABLA 4.10: Estimaciones posteriores de la variabilidad jerárquica y residual del modelo hurdle.

En la parte Bernoulli del modelo, las desviaciones estándar asociadas a los efectos aleatorios departamentales y municipales,($\sigma^{(b)}_{\text{dep}}$) y ($\sigma^{(b)}_{\text{mpio}}$), presentan magnitudes comparables. Este resultado indica que la variabilidad en la probabilidad de registrar al menos una muerte materna no se concentra en un único nivel administrativo, sino que se distribuye tanto entre departamentos como entre municipios dentro de cada departamento. En consecuencia, ambos niveles contribuyen de manera relevante a explicar la heterogeneidad observada en el componente binario del modelo, lo que justifica la inclusión simultánea de efectos aleatorios departamentales y municipales.

En la parte continua del modelo, la variabilidad a nivel departamental es considerablemente mayor que la municipal ($\sigma^{(g)}_{\text{dep}} \gg \sigma^{(g)}_{\text{mpio}}$). Esto indica que, cuando ocurre mortalidad materna, la magnitud de la RMM depende principalmente de diferencias estructurales entre departamentos, como la capacidad hospitalaria, la disponibilidad de personal calificado y las brechas de acceso a los servicios de salud.

Por último, la desviación estándar residual $\sigma_y = 0.75$ cuantifica la dispersión de la Razón de Mortalidad Materna en la parte continua del modelo, una vez incorporados los efectos jerárquicos departamentales y municipales. Su estimación posterior refleja que persiste variabilidad en los valores positivos de la RMM que no es capturada por la estructura multinivel especificada, en concordancia con la heterogeneidad observada previamente en la distribución de este indicador.

4.2.3. Efectos aleatorios jerárquicos

El modelo hurdle multinivel incorpora efectos aleatorios a nivel departamental y municipal para capturar la variabilidad que no es explicada por las covariables ni por los interceptos fijos. Estos efectos permiten analizar las diferencias territoriales: por un lado, en la probabilidad de que un municipio registre al menos un caso de mortalidad materna (parte binaria del modelo) y, por otro, en la magnitud de la RMM cuando se presenta mortalidad (parte continua del modelo).

Dado que los efectos aleatorios se modelan con distribuciones a priori centradas en cero, los valores posteriores positivos indican territorios con mayor riesgo relativo (en la parte Bernoulli) o con niveles esperados de RMM más altos (en la parte log-normal), manteniendo constantes las demás covariables.Por el contrario, los valores negativos señalan territorios donde, una vez ajustado el modelo, la probabilidad o la magnitud de la RMM es relativamente menor. Considerando que el modelo incluye 32 departamentos y 1.019 municipios, se presentan únicamente los resultados resumidos y los territorios más representativos, mientras que las tablas completas se incluyen en los Anexos.

Efectos aleatorios departamentales

En la parte continua del modelo (Tabla 4.12), los efectos aleatorios departamentales describen cómo varía la magnitud de la Razón de Mortalidad Materna entre departamentos cuando la mortalidad materna ocurre. Estos efectos se interpretan de manera relativa al comportamiento promedio nacional: valores positivos indican que, una vez ocurrido el evento, la RMM tiende a ser mayor en comparación con el promedio, mientras que valores negativos sugieren una magnitud menor. Esta interpretación es siempre condicional a la ocurrencia del evento y no corresponde a niveles absolutos de la RMM.

La RMM estimada reportada en la tabla se obtiene a partir del predictor lineal del componente continuo del modelo hurdle, el cual se encuentra definido en escala logarítmica. En particular, para cada departamento j, el predictor lineal se expresa como la suma del intercepto del modelo y el efecto aleatorio departamental correspondiente, es decir,

$$\eta_j = \mu_0 + u_j^{(g)}.$$

Para interpretar los resultados en la escala original de la Razón de Mortalidad Materna, se aplica la transformación exponencial, de modo que la RMM estimada se calcula como:

$$\widehat{\text{RMM}}_j = \exp\left(\mu_0 + u_j^{(g)}\right). \tag{4.2}$$

Esta cantidad corresponde al valor esperado de la Razón de Mortalidad Materna condicionado a que el evento ocurra ($Y > 0$) y representa la magnitud promedio de la RMM que se espera observar una vez se ha registrado al menos una muerte materna. En consecuencia, estos valores resumen la intensidad del indicador bajo la ocurrencia del evento e integran tanto los efectos fijos como la variabilidad departamental capturada por el modelo.

Departamento	Media posterior	q95
La Guajira	0.387	1.09
Huila	0.341	1.04
Chocó	0.263	0.866
Norte de Santander	0.243	0.936
Tolima	0.231	0.790

TABLA 4.11: Media posterior para la probabilidad de presentar un mayor riesgo de mortalidad materna en los departamentos (efectos aleatorios, componente Bernoulli)

En la parte Bernoulli del modelo (Tabla 4.11), los efectos aleatorios departamentales capturan diferencias sistemáticas en la probabilidad de registrar al menos una muerte materna entre los departamentos, una vez controladas las covariables sociodemográficas incluidas en el modelo. Los valores positivos indican una mayor propensión relativa a la ocurrencia del evento en comparación con el promedio nacional, mientras que valores cercanos a cero reflejan comportamientos similares al nivel de referencia. En este sentido, La Guajira, Huila, Chocó, Norte de Santander y Tolima presentan una mayor probabilidad relativa de registrar mortalidad materna, lo que evidencia la existencia de heterogeneidad departamental en la ocurrencia del evento.

Departamento	**Efecto aleatorio (escala log)**		**RMM estimada**	
	Media posterior	**q95**	**Media posterior**	**q95**
La Guajira	-0.159	0.224	331	474
Huila	0.116	0.540	437	644
Chocó	-0.287	0.053	290	403
Norte de Santander	-0.154	0.302	337	513
Tolima	0.140	0.516	443	612

TABLA 4.12: Media posterior de los efectos aleatorios departamentales en la parte continua (lognormal) del modelo hurdle, para los departamentos con mayor probabilidad de ocurrencia de mortalidad materna.

En la parte continua del modelo (Tabla 4.12), los efectos aleatorios departamentales describen cómo varía la magnitud de la Razón de Mortalidad Materna entre departamentos cuando la mortalidad materna ocurre. Estos efectos se interpretan de manera relativa al comportamiento promedio nacional: valores positivos indican que, una vez ocurrido el evento, la RMM tiende a ser mayor en comparación con el promedio, mientras que valores negativos sugieren una magnitud menor. Esta interpretación es siempre condicional a la ocurrencia del evento y no corresponde a niveles absolutos de la RMM.

La RMM estimada reportada en la tabla se obtiene aplicando una transformación a partir del predictor lineal del modelo continuo y corresponde al valor esperado de la Razón de Mortalidad Materna condicionado a que el evento ocurra. Este valor resume la magnitud promedio de la RMM que se espera observar cuando se registra al menos una muerte materna, integrando los efectos fijos y aleatorios incluidos en el modelo.

Es importante señalar que los resultados de la parte continua no necesariamente coinciden con los de la parte Bernoulli, dado que el modelo hurdle separa explícitamente los procesos que determinan la ocurrencia del evento de aquellos que explican su intensidad. Así, un departamento puede presentar una alta probabilidad de registrar mortalidad materna sin exhibir simultáneamente la mayor magnitud condicional de la RMM, lo cual refleja la complejidad y heterogeneidad territorial del fenómeno.

Efectos aleatorios municipales

Los efectos aleatorios a nivel municipal muestran un comportamiento diferenciado entre los dos componentes del modelo hurdle. En la parte Bernoulli, varios

municipios presentan efectos aleatorios positivos superiores a 0.20 (Tabla 4.13), lo que indica una mayor probabilidad relativa de registrar al menos un evento de mortalidad materna, una vez ajustado el modelo por covariables y por los efectos departamentales. Estos resultados evidencian heterogeneidad municipal en la ocurrencia del evento, asociada a condiciones locales no observadas por las variables incluidas.

En contraste, en la parte lognormal del modelo, los efectos aleatorios municipales se concentran en valores cercanos a cero (Tabla 4.14). Esto indica que, una vez ocurre la mortalidad materna ($Y > 0$), la magnitud esperada de la Razón de Mortalidad Materna no varía sustancialmente entre municipios, y se mantiene cercana al nivel promedio definido por el intercepto del modelo. En consecuencia, la contribución municipal en este componente es limitada y actúa como un ajuste fino sobre la intensidad del indicador.

Las Tablas 4.13 y 4.14 presentan los municipios con mayores efectos aleatorios en cada componente.

Departamento	**Municipio (Bernoulli)**	**Media posterior**	**q95**
Antioquia	Campamento	0.299	1.52
Huila	Isnos	0.230	1.37
Santander	San Benito	0.215	1.33
Cauca	Silvia	0.211	1.32
Meta	El Dorado	0.207	1.29
Chocó	Medio Baudó	0.205	1.27

TABLA 4.13: Media posterior para la probabilidad de presentar un mayor riesgo de mortalidad materna en los municipios (efectos aleatorios, componente Bernoulli).

Departamento	**Municipio**	**Efecto aleatorio (escala log)**		**RMM estimada**	
		Media posterior	**q95**	**Media posterior**	**q95**
Antioquia	Campamento	0.0000434	0.0807	285	340
Huila	Isnos	0.00166	0.0837	286	341
Santander	San Benito	0.00758	0.0982	287	344
Cauca	Silvia	0.00312	0.0858	286	341
Meta	El Dorado	0.000337	0.0815	285	340
Chocó	Medio Baudó	0.00814	0.0980	287	344

TABLA 4.14: Media posterior de los efectos aleatorios municipales en la parte continua (lognormal) del modelo hurdle, para los municipios con mayor probabilidad de ocurrencia de mortalidad materna.

La RMM estimada reportada en la Tabla 4.14 se obtiene mediante la transformación exponencial del predictor lineal del componente lognormal, que combina el intercepto del modelo y el efecto aleatorio municipal. Estos valores representan el valor esperado de la RMM condicionado a la ocurrencia del evento y no un promedio poblacional. La similitud entre las estimaciones municipales refleja que los efectos aleatorios en este nivel son de pequeña magnitud en comparación con el intercepto, por lo que la variación municipal en la intensidad de la RMM es limitada una vez se ha producido la mortalidad materna.

Los resultados permiten concluir que:

- Se observa heterogeneidad territorial en la probabilidad de registrar mortalidad materna, con efectos relevantes tanto a nivel departamental como municipal en la parte Bernoulli del modelo.
- La variabilidad en la magnitud de la RMM, condicionada a la ocurrencia del evento, se explica principalmente por diferencias departamentales, mientras que los efectos municipales en la parte continua son reducidos.
- El modelo hurdle multinivel permite distinguir entre los factores asociados a la ocurrencia de la mortalidad materna y aquellos que influyen en su intensidad, capturando de manera adecuada la estructura jerárquica del territorio colombiano.

4.2.4. Predicciones posteriores y evaluación del ajuste

A partir de los parámetros estimados, se generaron predicciones posteriores de la Razón de Mortalidad Materna (RMM) para cada municipio. Para cada unidad territorial se calcularon la media posterior, la desviación estándar y los cuantiles 5 %, 50 % y 95 %, lo que permite evaluar tanto el nivel esperado de RMM como la incertidumbre asociada a cada estimación. La Tabla 4.15 presenta, de manera resumida, los diez municipios con los valores esperados más altos y más bajos de RMM; el detalle completo para los 1.100 municipios se encuentra en el Anexo correspondiente.

Departamento	Municipio	Media	Desv. Est.	q_5	Mediana	q_{95}
Municipios con mayor RMM esperada						
Chocó	bagadó	600.0	334.0	209.0	531.0	1228.0
Nariño	el charco	419.0	236.0	128.0	374.0	863.0
Norte de Santander	hacarí	404.0	214.0	146.0	361.0	809.0
Chocó	bojayá	402.0	174.0	177.0	373.0	726.0
Chocó	medio san juan	394.0	250.0	109.0	338.0	871.0
Nariño	roberto payán	393.0	509.0	29.1	232.0	1281.0
Chocó	medio baudó	371.0	229.0	99.6	324.0	803.0
Chocó	río quito	346.0	121.0	180.0	330.0	567.0
Antioquia	murindó	340.0	179.0	125.0	304.0	678.0
Chocó	bajo baudó	317.0	158.0	115.0	288.0	615.0
Municipios con menor RMM esperada						
Santander	encino	14.3	10.1	3.44	12.2	31.8
Santander	guapotá	14.9	12.1	3.10	12.2	35.5
Santander	san sebastián	15.1	11.6	3.19	12.4	35.5
Cauca	san francisco	15.8	12.6	3.30	12.7	38.7
Cauca	guachené	16.3	10.7	4.26	14.1	35.7
Huila	altamira	16.5	11.7	4.13	13.8	37.5
Boyacá	busbanzá	16.7	13.9	3.42	13.5	40.4
Cauca	padilla	17.0	11.3	4.40	14.6	37.3
Boyacá	firavitoba	18.4	13.9	4.09	15.2	43.2
Boyacá	viracachá	18.7	16.9	3.41	14.5	47.1

TABLA 4.15: Resumen de la distribución posterior de la RMM en municipios seleccionados.

La Figura 4.23 compara la RMM observada con la media posterior de la predicción para cada municipio, incluyendo intervalos creíbles al 90 %. La línea diagonal punteada representa el ajuste perfecto (*predicción = observación*). La mayoría de los puntos se concentran en valores bajos de RMM, donde las predicciones se ajustan razonablemente a los datos observados, aunque el modelo tiende a suavizar los valores extremos. En los municipios con RMM muy alta, la media posterior suele situarse por debajo del valor observado, reflejando el efecto de *shrinkage* característico

del enfoque jerárquico: las predicciones extremas se retraen hacia la media global al combinar información de otros municipios y departamentos.

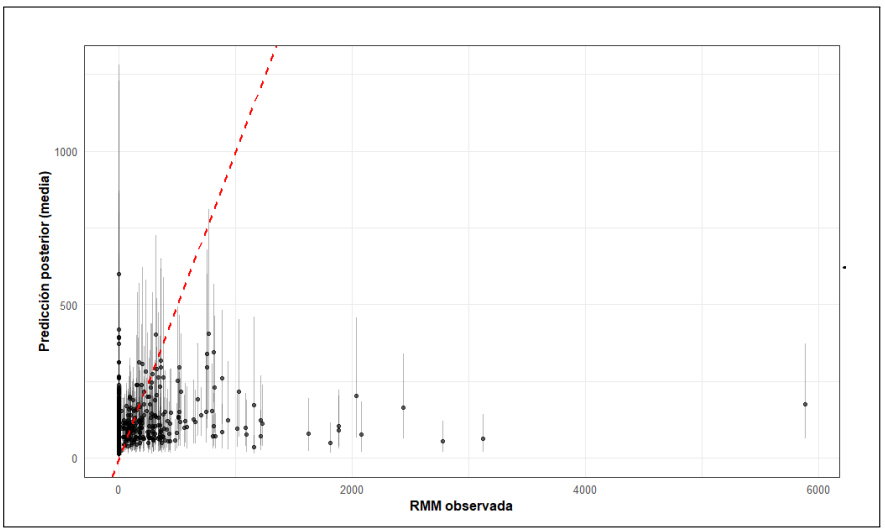

FIGURA 4.23: Predicción posterior media de la RMM frente a los valores observados por municipio.

La distribución de los residuos posteriores se presenta en la Figura 4.24. El histograma evidencia una marcada concentración de residuos alrededor de cero, lo que indica que, para la mayoría de los municipios, el modelo reproduce adecuadamente los niveles observados de RMM. La cola derecha, más extendida, corresponde a un pequeño grupo de municipios con RMM extremadamente alta, donde la predicción posterior subestima parcialmente los valores observados. Estos casos reflejan territorios con riesgos excepcionales que exceden el patrón promedio capturado por las covariables y los efectos aleatorios.

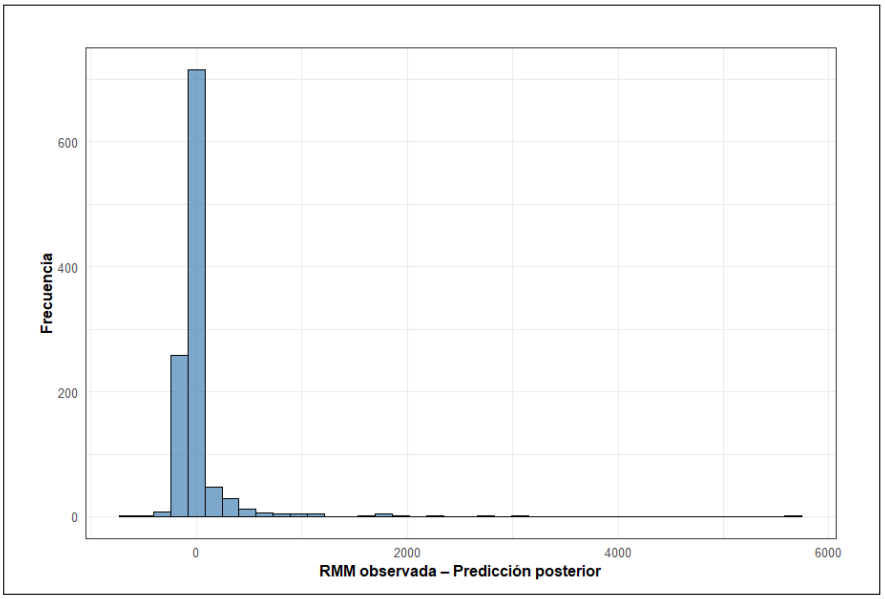

FIGURA 4.24: Distribución de los residuos posteriores (RMM observada menos predicción posterior media).

En conjunto, las predicciones posteriores permiten identificar de manera robusta los municipios con mayor riesgo estimado de mortalidad materna, principalmente en regiones con alta vulnerabilidad socioeconómica y acceso limitado a servicios de salud, así como aquellos con niveles relativamente bajos. El análisis de residuos sugiere que el modelo captura adecuadamente la estructura general del fenómeno. En particular, los residuos posteriores de la componente continua, definidos como la diferencia entre el valor observado y la media de la predicción posterior, y evaluados sobre $\log(Y_{ij})$ condicional a $Y_{ij} > 0$, se concentran alrededor de cero y no exhiben patrones sistemáticos. Este comportamiento es consistente con lo esperado en un modelo *hurdle* log-normal correctamente especificado, ya que un centrado en cero indica ausencia de sesgo global en la predicción.

No obstante, se identifica un reducido grupo de municipios con RMM excepcionalmente alta, incluso tras ajustar por covariables y estructura jerárquica. Estos casos reflejan territorios con riesgos atípicos que exceden el patrón promedio capturado por el modelo y resultan relevantes para la priorización de intervenciones. En conjunto, estos resultados complementan los hallazgos descriptivos y de clústeres, proporcionando una base cuantitativa sólida para la focalización de políticas de salud materna.

Mapas de predicción posterior y análisis espacial

Con el fin de recuperar la dimensión territorial del fenómeno, se representaron las predicciones posteriores de la Razón de Mortalidad Materna (RMM) mediante diferentes mapas temáticos. Estos mapas permiten identificar patrones espaciales, niveles de riesgo y zonas de alta incertidumbre.

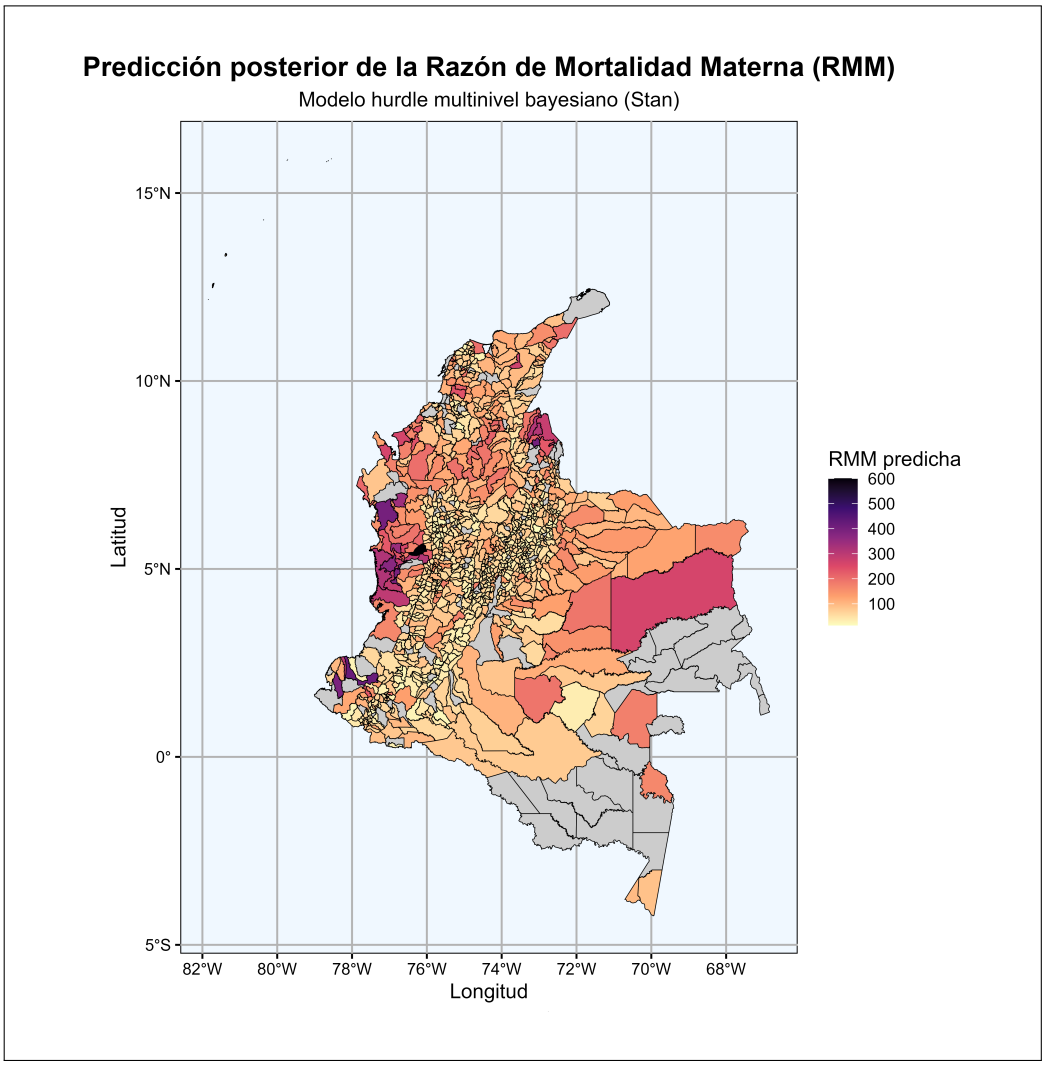

FIGURA 4.25: Predicción posterior media de la RMM a nivel municipal. Modelo hurdle multinivel bayesiano (Stan).

El mapa de la media posterior (Figura 4.25) muestra que los valores más elevados de RMM predicha se concentran en municipios del litoral Pacífico, particularmente en el Chocó (Bagadó, Bojayá, Medio San Juan, Bajo Baudó, Murindó, entre otros) y en parte del litoral nariñense (El Charco, Roberto Payán). También se identifican focos relevantes en zonas rurales de Norte de Santander. Por el contrario, los valores más bajos se localizan en el eje andino central, Bogotá, Boyacá y Antioquia.

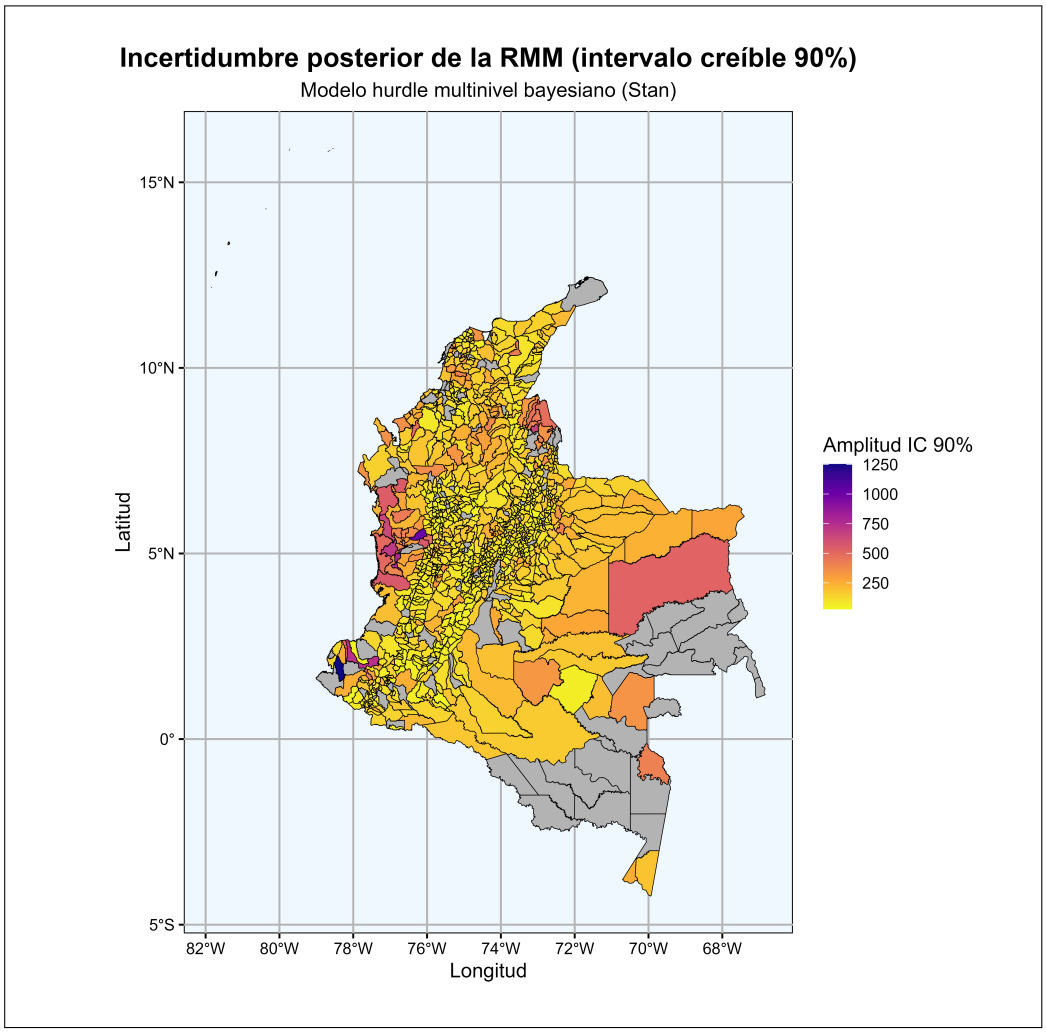

FIGURA 4.26: Incertidumbre posterior (amplitud del intervalo creíble del 90 %) de la RMM predicha a nivel municipal.

La Figura 4.26 presenta la amplitud del intervalo creíble al 90 %.Los niveles de incertidumbre más alto se concentran en municipios de baja población y difícil acceso (Amazonía, Orinoquía y parte del Pacífico). Por el contrario, los municipios urbanos e intermedios presentan estimaciones mucho más precisas, reflejando una mayor disponibilidad de información.

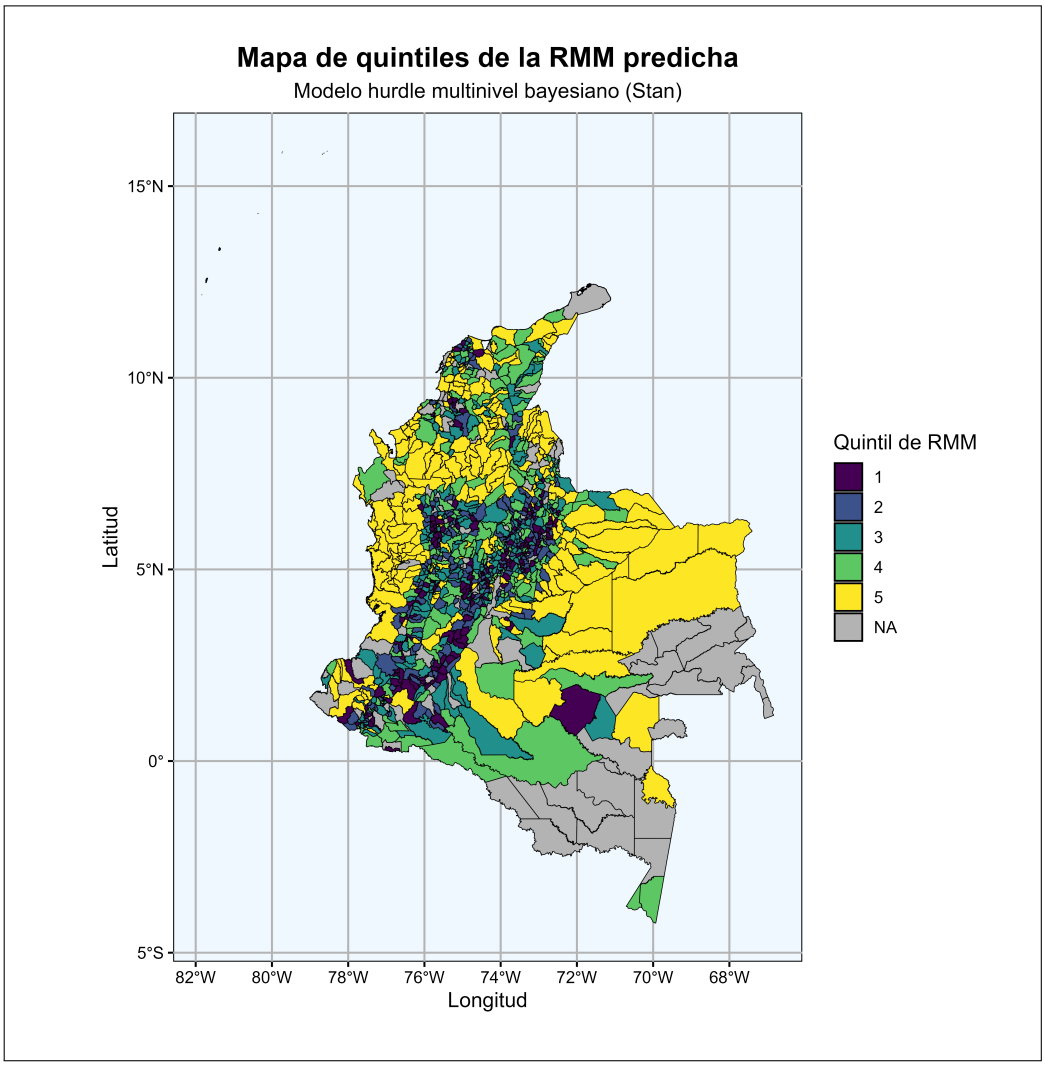

FIGURA 4.27: Clasificación municipal en quintiles según la media posterior de RMM.

La clasificación por quintiles (Figura 4.27) muestra de forma clara las desigualdades territoriales: los municipios en los quintiles más altos (4 y 5) se concentran casi exclusivamente en el Pacífico y en zonas rurales dispersas, mientras que los quintiles más bajos aparecen en ciudades principales y sus áreas aledañas.

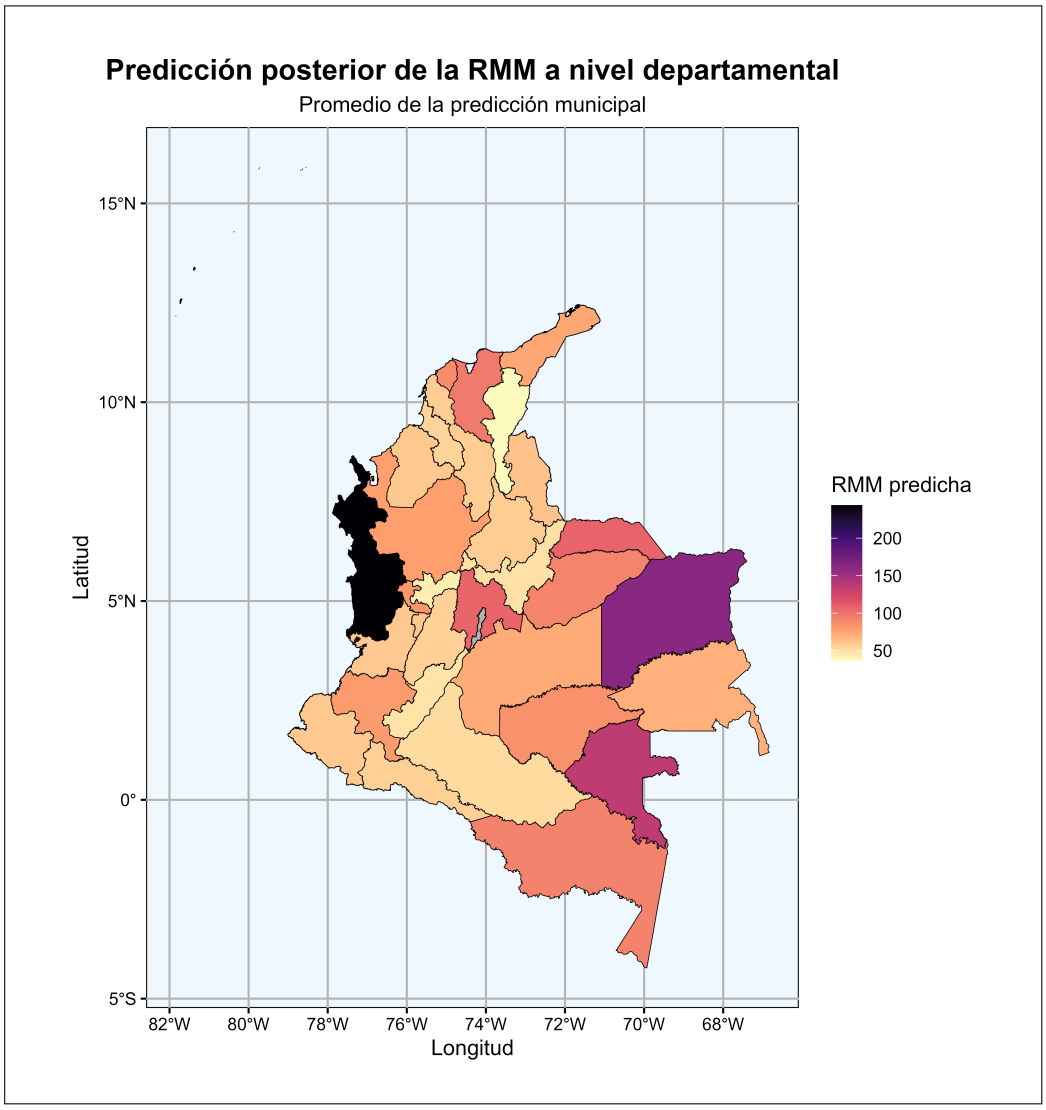

FIGURA 4.28: Predicción posterior de la RMM a nivel departamental (promedio de la predicción municipal).

La Figura 4.28 muestra la predicción posterior de la Razón de Mortalidad Materna a nivel departamental. Aunque la representación se realiza a este nivel de agregación, los valores departamentales corresponden a promedios de la predicción posterior obtenida a partir del modelo jerárquico estimado a nivel municipal, el cual incorpora efectos aleatorios departamentales. Bajo esta representación, se observan diferencias marcadas entre departamentos: Chocó, Guainía, Vaupés y La Guajira presentan los valores promedio más altos de RMM predicha, mientras que Bogotá, Boyacá, Cundinamarca y Antioquia exhiben niveles sustancialmente menores.

Síntesis del Resultado Tres

Los análisis del Resultado 3 permiten concluir lo siguiente:

- El modelo hurdle multinivel bayesiano implementado en `Stan` representa adecuadamente la estructura de los datos, separando la probabilidad de ocurrencia de muertes maternas y la magnitud de la RMM cuando estas se presentan.

- Los diagnósticos MCMC evidencian una clara convergencia y estabilidad de las cadenas, garantizando una inferencia posterior confiable.
- Se identificó variabilidad importante tanto a nivel departamental como municipal, incluso después de ajustar por covariables. Esto confirma que la estructura territorial es un componente esencial para explicar la RMM.
- Los efectos aleatorios permiten detectar territorios que mantienen riesgo residual elevado, lo que sugiere factores no observados relacionados con pobreza estructural, conflicto armado, disponibilidad de servicios y barreras geográficas.
- Los mapas de predicción posterior confirmaron fuertes desigualdades: los niveles más altos de RMM se concentran en el Pacífico, la Amazonía, la Orinoquía y algunos municipios del nororiente; mientras que los niveles más bajos se observan en el corredor andino y las principales áreas urbanas.
- Los mapas de incertidumbre indican que las estimaciones más imprecisas se concentran en municipios pequeños o de baja información, reforzando la necesidad de fortalecer los sistemas locales de registro y vigilancia epidemiológica.

En conjunto, este modelo proporciona una herramienta rigurosa para la priorización territorial de intervenciones en salud materna, permitiendo identificar municipios y departamentos críticos con alta probabilidad de mortalidad materna y, además, cuantificar la incertidumbre asociada a dichas estimaciones.

4.3. Resultado 3. Factores sociodemográficos asociados a la Razón de Mortalidad Materna en Colombia, 2023

El análisis de los factores sociodemográficos asociados a la Razón de Mortalidad Materna (RMM) en Colombia durante el año 2023 se basó en los efectos fijos estimados a partir del modelo *hurdle* multinivel bayesiano. Este enfoque permite distinguir dos procesos estocásticos complementarios: en primer lugar, la probabilidad de que un municipio registre al menos un evento de mortalidad materna, la cual se modela mediante un componente Bernoulli con función de enlace logit; y, en segundo lugar, la magnitud de la RMM condicionada a la ocurrencia del evento, que se modela mediante una distribución lognormal aplicada al logaritmo de la RMM positiva.

La identificación de factores asociados se fundamenta en la estimación posterior de los coeficientes de regresión ($\boldsymbol{\beta}$) correspondientes a las covariables sociodemográficas incluidas en el modelo. Dado que estas covariables fueron previamente estandarizadas, cada coeficiente puede interpretarse como el efecto asociado a un incremento de una desviación estándar en la covariable respectiva, lo que facilita la comparación directa de la magnitud de los efectos entre los distintos factores dentro de cada componente del modelo.

Para cada covariable, la evidencia posterior se evaluó a partir de los siguientes criterios estadísticos:

- **Media posterior del coeficiente**: utilizada como una medida puntual del efecto estimado a partir de la distribución posterior.
- **Intervalo creíble al 95 %**: rango dentro del cual se encuentra el coeficiente con una probabilidad del 95 %, y que permite evaluar la incertidumbre del efecto. Se considera que una covariable presenta evidencia estadística de asociación cuando dicho intervalo no incluye el valor cero, lo cual indica que el efecto mantiene una dirección consistente en la distribución posterior, es decir, que la mayor parte de la masa de probabilidad del coeficiente se concentra en valores con el mismo signo.
- **Probabilidad posterior de dirección (pd)**: Probabilidad de que el efecto estimado sea mayormente positivo o negativo en la distribución posterior; valores cercanos a 1 indican alta certeza sobre el signo del efecto. (Makowski et al., 2019).

En el componente Bernoulli, los coeficientes se interpretan en términos de razones de odds, obtenidas mediante la transformación exponencial de los coeficientes estimados (OR=e^{β}). Esta interpretación permite cuantificar el cambio relativo en la probabilidad de que un municipio registre mortalidad materna ante variaciones en los factores sociodemográficos. Por su parte, en el componente lognormal, la exponenciación de los coeficientes se interpreta como un factor multiplicativo sobre la RMM condicionada a que esta sea positiva (Ecuación 4.2), lo que facilita la lectura de los efectos en la escala original del indicador.

Bajo este marco analítico, los resultados que se presentan a continuación permiten identificar los factores sociodemográficos con mayor evidencia de asociación, tanto con la ocurrencia como con la magnitud de la Razón de Mortalidad Materna en Colombia durante el año 2023, en coherencia con el objetivo específico planteado.

En las secciones siguientes (4.3.1, 4.3.2) se presentan los resultados del componente Bernoulli, correspondiente a la ocurrencia de mortalidad materna, y del componente lognormal, correspondiente a la magnitud de la Razón de Mortalidad Materna condicionada a su ocurrencia. Los resultados se resumen en las Tablas 4.16 y 4.17. Para ambos componentes se reportan la media posterior del coeficiente (MP), los percentiles 2.5 y 97.5 de la distribución posterior, que delimitan el intervalo creíble al 95 %, los odds ratios o factores multiplicativos asociados a la media posterior y a dichos percentiles, la probabilidad posterior de dirección (*pd*), el indicador de significancia estadística de cada covariable (Sig), así como los estadísticos de diagnóstico de convergencia y eficiencia del muestreo, incluyendo el estadístico $\widehat{R}$ y los tamaños efectivos de muestra ESS_{bulk} y ESS_{tail}.

4.3.1. Resultados del componente Bernoulli: ocurrencia de mortalidad materna

El componente Bernoulli del modelo *hurdle* evalúa la asociación entre los factores sociodemográficos y la probabilidad de que un municipio registre al menos un evento de mortalidad materna en 2023. En este componente, los coeficientes representan efectos sobre el logaritmo de las *odds* de ocurrencia del evento (Ecuación 4.1), ajustados por la estructura multinivel y las covariables incluidas.

La evidencia de asociación se determina a partir de los intervalos creíbles al 95 % y de las probabilidades posteriores de dirección: se consideran asociadas aquellas variables cuyos intervalos no incluyen el valor cero, o se sitúan próximos a este, y que presentan probabilidades posteriores de dirección altas, de modo que la distribución posterior concentra la mayor parte de su masa en un mismo signo (Tabla 4.16). Bajo estos criterios, las variables con evidencia de asociación con la ocurrencia de mortalidad materna son las siguientes:

- Producto interno bruto per cápita estandarizado (`pib_s`)
- Índice de pobreza multidimensional (`ipm_s`)
- Proporción de población con necesidades básicas insatisfechas en el ámbito rural (`nbi_rur_s`)
- Cobertura de controles prenatales (`cont_prenat_s`)

Los resultados indican que el PIB per cápita estandarizado (`pib_s`) se asocia positivamente con la probabilidad de que un municipio registre al menos un evento de mortalidad materna, mientras que el índice de pobreza multidimensional `ipm_s`) muestra una asociación inversa con dicha probabilidad. Asimismo, la proporción de población con necesidades básicas insatisfechas en el ámbito rural (`nbi_rur_s`) presenta una asociación positiva, lo que sugiere una mayor probabilidad de ocurrencia del evento en contextos de mayor privación rural. En contraste, la cobertura de controles prenatales (`cont_prenat_s`) se asocia negativamente, indicando un efecto protector marginal frente a la ocurrencia de mortalidad materna.

Las demás covariables incluidas en el componente Bernoulli no muestran evidencia estadística suficiente de asociación, por lo que no se identifican relaciones concluyentes con la ocurrencia de mortalidad materna en 2023 bajo la especificación del modelo.

Covariable	Media posterior	$q_{2.5}$	$q_{97.5}$	OR	$OR_{2.5}$	$OR_{97.5}$	pd	Sig.	rhat	ESS_{bulk}	ESS_{tail}
pib_s	3.78	2.56	5.05	43.7	12.9	156.0	1.000	Sí	1.00	69692	31926
ipm_s	-0.65	-1.06	-0.25	0.52	0.35	0.78	0.999	Sí	1.00	30829	25130
nbi_rur_s	0.71	0.26	1.17	2.03	1.29	3.22	0.999	Sí	1.00	27784	23961
cont_prenat_s	-0.30	-0.60	-0.00	0.74	0.55	1.00	0.976	Sí	1.00	57589	31569
cob_edu_neta_muj_s	0.39	-0.06	0.88	1.47	0.94	2.40	0.952	No	1.00	34804	29219
cob_ene_urb_s	0.17	-0.07	0.45	1.18	0.94	1.56	0.907	No	1.00	66402	30811
iica_s	0.08	-0.09	0.25	1.09	0.92	1.28	0.840	No	1.00	66523	32122
partos_calif_s	-0.12	-0.36	0.12	0.89	0.70	1.13	0.834	No	1.00	47241	29954
cob_ene_rur_s	0.13	-0.15	0.41	1.13	0.87	1.50	0.816	No	1.00	54445	31323
cob_acue_rur_s	-0.12	-0.38	0.14	0.89	0.68	1.15	0.816	No	1.00	58420	31174
nbi_urb_s	0.15	-0.28	0.58	1.16	0.76	1.79	0.758	No	1.00	45399	30635
cob_edu_bruta_tot_s	-0.03	-0.52	0.41	0.97	0.60	1.51	0.541	No	1.00	36665	29508
prop_nbi_s	0.03	-0.56	0.62	1.03	0.57	1.86	0.539	No	1.00	47735	29981
cob_acue_urb_s	-0.01	-0.31	0.29	0.99	0.74	1.34	0.534	No	1.00	47632	30271

TABLA 4.16: Estimaciones posteriores de los efectos fijos de las covariables sociodemográficas sobre la probabilidad de ocurrencia de mortalidad materna (componente Bernoulli del modelo hurdle)

Dado que las covariables incluidas en el componente Bernoulli fueron estandarizadas, los *odds ratios* (OR) estimados representan el cambio relativo en las *odds* de que un municipio registre al menos un evento de mortalidad materna asociado a un incremento de una desviación estándar en cada covariable, manteniendo constantes las demás variables del modelo. En este sentido, el producto interno bruto per cápita estandarizado (`pib_s`) presenta un OR de 43.7, lo que implica un incremento aproximado del 4 270 % en las *odds* de ocurrencia del evento. No obstante, debido a la estandarización de la covariable y a que la mortalidad materna constituye un evento poco frecuente, este valor debe interpretarse como evidencia de una asociación direccional fuerte y consistente, más que como una relación causal directa de gran magnitud.

Por su parte, el índice de pobreza multidimensional (`ipm_s`) muestra un OR de 0.52, correspondiente a una reducción aproximada del 48 % en las *odds* de ocurrencia por cada desviación estándar adicional. Asimismo, la proporción de población con necesidades básicas insatisfechas en el ámbito rural (`nbi_rur_s`) presenta un OR de 2.03, lo que indica un aumento cercano al 103 % en las *odds* de ocurrencia del evento. Finalmente, la cobertura de controles prenatales (`cont_prenat_s`) exhibe un OR de 0.74, lo que sugiere un efecto protector marginal, correspondiente a una disminución aproximada del 35 % en las *odds* de que ocurra al menos un evento de mortalidad materna ante un incremento de una desviación estándar en dicha cobertura.

4.3.2. Resultados del componente lognormal: magnitud de la Razón de Mortalidad Materna condicionada a su ocurrencia

El componente lognormal del modelo hurdle analiza los factores sociodemográficos asociados a la magnitud de la Razón de Mortalidad Materna (RMM), condicionada a que un municipio haya registrado al menos un evento durante el año 2023. En este componente, los coeficientes estimados describen el efecto de las covariables sobre el logaritmo de la RMM positiva, manteniendo constantes las demás variables e incorporando la estructura jerárquica del modelo.

Los resultados del componente lognormal se presentan en la Tabla 4.17, donde se reportan las medias posteriores de los coeficientes, sus intervalos creíbles al 95 %, los factores multiplicativos obtenidos mediante la transformación exponencial y la

probabilidad posterior de dirección. La evidencia estadística de asociación se evalúa a partir de intervalos creíbles que no incluyen el valor cero y probabilidades posteriores de dirección elevadas. Con base en estos criterios, se identifica evidencia estadística de asociación para:

- Índice de pobreza multidimensional (`ipm_s`).
- Producto interno bruto per cápita estandarizado (`pib_s`).
- Proporción de población con necesidades básicas insatisfechas en el ámbito rural (`nbi_rur_s`).
- Cobertura de controles prenatales (`cont_prenat_s`).

Las restantes covariables no muestran evidencia suficiente de asociación en este componente del modelo.

La interpretación de los efectos indica que un mayor índice de pobreza multidimensional se asocia con un aumento en la magnitud de la RMM, una vez que ocurre el evento. En contraste, mayores niveles de PIB per cápita y de necesidades básicas insatisfechas en el ámbito rural se asocian con una menor magnitud de la RMM condicionada a su ocurrencia. Por su parte, la cobertura de controles prenatales presenta una asociación positiva con la magnitud de la RMM, lo que sugiere que, entre los municipios donde ocurre mortalidad materna, mayores niveles de cobertura se relacionan con valores más altos de la RMM.

En conjunto, los resultados del componente lognormal sugieren que, una vez condicionada la ocurrencia de mortalidad materna, la variación en la magnitud de la RMM a nivel municipal se encuentra asociada principalmente con condiciones socioeconómicas y con el acceso a servicios de salud, mientras que otros factores sociodemográficos no muestran una asociación estadísticamente identificable bajo los criterios de inferencia bayesiana empleados.

Covariable	Media posterior	$q_{2.5}$	$q_{97.5}$	Factor	$\text{Factor}_{2.5}$	$\text{Factor}_{97.5}$	pd	Sig.	rhat	ESS_{bulk}	ESS_{tail}
ipm_s	0.70	0.45	0.94	2.01	1.57	2.57	1.000	Sí	1.00	34256	30901
pib_s	-0.39	-0.60	-0.19	0.68	0.55	0.83	1.000	Sí	1.00	68411	30754
nbi_rur_s	-0.37	-0.59	-0.15	0.69	0.55	0.86	0.999	Sí	1.00	44601	34037
cont_prenat_s	0.23	0.02	0.43	1.25	1.02	1.54	0.983	Sí	1.00	45341	33204
cob_acue_rur_s	-0.17	-0.35	0.01	0.85	0.71	1.01	0.964	No	1.00	48686	33532
cob_edu_bruta_tot_s	-0.20	-0.58	0.17	0.82	0.56	1.19	0.859	No	1.00	39481	30886
cob_edu_neta_muj_s	0.19	-0.18	0.55	1.21	0.83	1.74	0.843	No	1.00	39999	31453
cob_ene_urb_s	0.07	-0.12	0.25	1.07	0.89	1.28	0.766	No	1.00	61422	34700
partos_calif_s	-0.05	-0.20	0.10	0.95	0.82	1.10	0.748	No	1.00	44851	32442
cob_acue_urb_s	0.06	-0.15	0.28	1.07	0.86	1.32	0.721	No	1.00	34363	31315
nbi_urb_s	-0.05	-0.33	0.24	0.96	0.72	1.27	0.627	No	1.00	38225	31373
iica_s	-0.01	-0.13	0.11	0.99	0.87	1.11	0.595	No	1.00	55624	33901
cob_ene_rur_s	-0.02	-0.18	0.14	0.98	0.84	1.15	0.579	No	1.00	53372	31714
prop_nbi_s	0.03	-0.28	0.34	1.03	0.76	1.41	0.579	No	1.00	41910	32058

TABLA 4.17: Estimaciones posteriores de los efectos fijos de las covariables sociodemográficas en la magnitud de la RMM (componente lognormal del modelo hurdle)

Dado que el componente lognormal modela el logaritmo de la Razón de Mortalidad Materna positiva, los coeficientes estimados se interpretan mediante su transformación exponencial, dando lugar a factores multiplicativos sobre la magnitud esperada de la RMM, condicionada a que el evento haya ocurrido. En este contexto, un valor del factor mayor que uno indica un incremento relativo en la RMM,

mientras que un valor menor que uno representa una disminución proporcional, asociada a un aumento de una desviación estándar en la covariable correspondiente, manteniendo constantes las demás variables del modelo. Así, el índice de pobreza multidimensional estandarizado (`ipm_s`) presenta un factor de 2.01, lo que implica que un incremento de una desviación estándar en esta variable se asocia con un aumento aproximado del 101 % en la magnitud esperada de la RMM, una vez ocurrido el evento. Por su parte, el producto interno bruto per cápita estandarizado (`pib_s`) exhibe un factor de 0.68, indicando una reducción cercana al 32 % en la RMM asociada a mayores niveles de esta variable.

De manera similar, la proporción de población con necesidades básicas insatisfechas en el ámbito rural (`nbi_rur_s`) presenta un factor de 0.69, lo que corresponde a una disminución aproximada del 31 % en la magnitud de la RMM condicionada a su ocurrencia. En contraste, la cobertura de controles prenatales (`cont_prenat_s`) muestra un factor de 1.25, sugiriendo un incremento relativo cercano al 25 % en la RMM entre los municipios donde ocurre mortalidad materna. En conjunto, estos resultados indican que, una vez condicionada la ocurrencia del evento, las condiciones socioeconómicas y el acceso a servicios de salud se asocian con cambios proporcionales en la magnitud de la RMM, reflejando heterogeneidad en la severidad del fenómeno más allá de su simple presencia.

4.3.3. Síntesis del Resultado tres

El análisis de los componentes Bernoulli y lognormal del modelo hurdle multinivel permite distinguir los factores sociodemográficos asociados a dos procesos estadísticos distintos: la ocurrencia de al menos un evento de mortalidad materna y la magnitud de la Razón de Mortalidad Materna (RMM) una vez ocurrido el evento en Colombia durante 2023. Los resultados muestran que las covariables no afectan ambos procesos de la misma forma: algunas se asocian con la probabilidad de registrar casos, mientras que otras influyen principalmente en la intensidad de la RMM. Esta diferenciación respalda la pertinencia del enfoque hurdle para el análisis de datos municipales con exceso de ceros, en concordancia con la literatura metodológica (Min, Y. and Agresti, A., 2005; Hilbe, 2011).

Los resultados empíricos evidencian que los factores asociados a la mortalidad materna no operan de manera homogénea, sino que su efecto depende del proceso analizado. En particular, se observa que algunos indicadores socioeconómicos y de acceso a servicios de salud están vinculados principalmente con la ocurrencia del evento, mientras que otros inciden sobre la magnitud de la RMM una vez que esta se presenta. Este patrón es consistente con hallazgos previos en la literatura de salud pública, que señalan que los determinantes de la mortalidad materna pueden manifestarse de forma diferenciada según se analice la presencia del evento o su intensidad, especialmente en contextos de marcada desigualdad socioeconómica y territorial (Filippi et al., 2016). En este sentido, la evidencia para 2023 sugiere que distinguir explícitamente entre ambos procesos aporta información relevante para una comprensión más precisa del comportamiento de la RMM a nivel municipal y para la identificación de factores asociados al fenómeno.

Capítulo 5

Conclusiones

El presente estudio desarrolló un modelo bayesiano multinivel de tipo hurdle para analizar la Razón de Mortalidad Materna (RMM) en los municipios de Colombia. La combinación de una parte Bernoulli para modelar la ocurrencia del evento y una parte lognormal para la magnitud de la razón, cuando esta es positiva, permitió representar adecuadamente la naturaleza altamente asimétrica del indicador y su exceso de ceros. La estructura jerárquica municipios–departamentos capturó diferencias territoriales que no pueden ser explicadas solo por las covariables disponibles, proporcionando una caracterización más precisa de las desigualdades en salud materna.

Síntesis general del estudio

El modelo hurdle se ajustó de forma adecuada al patrón empírico de la RMM, diferenciando la probabilidad de registrar mortalidad materna del nivel de la razón entre los municipios con al menos un caso. El enfoque bayesiano, implementado mediante HMC/NUTS en Stan, permitió obtener distribuciones posteriores completas para los parámetros, facilitando la evaluación de la incertidumbre y la interpretación de la variabilidad territorial. La inclusión de efectos aleatorios a nivel de departamento y municipio evidenció la importancia del contexto territorial, especialmente en un país con heterogeneidades estructurales marcadas.

Principales hallazgos

En la parte Bernoulli, se identificó que la probabilidad de registrar un caso de mortalidad materna está relacionada con condiciones sociodemográficas y territoriales, como pobreza, acceso a servicios básicos y brechas educativas, lo que sugiere que la ocurrencia del evento refleja desigualdades persistentes en bienestar y provisión de servicios esenciales. En la parte lognormal, los municipios que sí reportaron mortalidad materna mostraron niveles de RMM influenciados por factores estructurales asociados a vulnerabilidad social, limitaciones en infraestructura y desigual acceso a la atención materna. Esto indica que la intensidad del riesgo no solo depende de que el evento ocurra, sino del entorno institucional y territorial en el que tiene lugar.

De manera adicional, la estimación posterior de los coeficientes de regresión permitió identificar que no todas las covariables sociodemográficas consideradas presentan el mismo nivel de relevancia estadística en la explicación de la RMM. Los

resultados muestran que algunos factores se asocian principalmente con la probabilidad de ocurrencia de la mortalidad materna, mientras que otros influyen sobre la magnitud de la razón una vez ocurrido el evento, lo que confirma la conveniencia de analizar ambos procesos de forma diferenciada. Este hallazgo resalta la utilidad del enfoque *hurdle* multinivel para distinguir el papel específico de las covariables en cada componente del modelo y aporta evidencia empírica relevante para la comprensión de las desigualdades territoriales en salud materna.

En conjunto, el análisis confirma que la mortalidad materna en Colombia muestra profundas desigualdades, con variaciones marcadas entre departamentos y municipios que requieren especial atención

Aportes metodológicos

El uso de un modelo hurdle multinivel en un marco bayesiano constituye un aporte significativo para el estudio de eventos raros con exceso de ceros. La capacidad del modelo para integrar simultáneamente estructuras jerárquicas, múltiples fuentes de heterogeneidad y distribuciones altamente asimétricas ofrece una herramienta robusta para investigaciones en salud pública con datos territoriales. La implementación en Stan permitió manejar la complejidad del modelo y obtener estimaciones estables para un número elevado de parámetros, destacando la utilidad de enfoques bayesianos en contextos con información incompleta o irregular.

Implicaciones para la política pública

Los resultados subrayan la necesidad de intervenciones focalizadas que atiendan desigualdades territoriales en determinantes sociales, infraestructura y acceso a servicios maternos. La identificación de municipios con alta probabilidad de registrar mortalidad materna o con razones significativamente elevadas proporciona evidencia útil para la priorización territorial en políticas públicas.

Asimismo, el análisis muestra que la planificación en salud materna debe considerar no solo factores individuales, sino también las condiciones estructurales de los territorios y la capacidad institucional local para responder a emergencias obstétricas.

Limitaciones del estudio

El estudio enfrenta limitaciones derivadas principalmente de la disponibilidad y calidad de la información:

- **Integración de múltiples fuentes:** La construcción de la base de datos requirió armonizar múltiples fuentes estadísticas, con diferencias en periodos, definiciones y niveles de desagregación. La falta de información uniforme a nivel municipal obligó a un esfuerzo considerable de estandarización para generar un conjunto de datos coherente.
- **Subregistro:** La presencia de ceros y la magnitud reportada de la RMM pueden verse afectadas por subregistro en municipios dispersos, de difícil acceso o con baja capacidad institucional, lo que incide directamente en la calidad del indicador.
- **Complejidad computacional:** El modelo bayesiano multinivel utilizado presenta elevada complejidad computacional debido al número de parámetros

y a la incorporación de efectos jerárquicos, requiriendo recursos de cómputo significativos y experiencia técnica para ajustar e interpretar correctamente las cadenas MCMC

Líneas de investigación futura

Este estudio abre diversas posibilidades para investigaciones posteriores, ya que los resultados muestran que los determinantes sociodemográficos analizados no explican completamente la variabilidad de la mortalidad materna. Por ello, resulta pertinente integrar indicadores institucionales adicionales, como disponibilidad de transporte sanitario, tiempos de traslado o capacidad de la red de atención, que puedan complementar y mejorar la comprensión del riesgo.

Asimismo, futuras investigaciones podrían analizar la RMM a lo largo de varios años para evaluar la estabilidad y evolución de los patrones territoriales identificados. Este enfoque permitiría determinar si las desigualdades se mantienen, aumentan o disminuyen con el tiempo y brindar información sobre la efectividad de las políticas de salud pública. Además, ayudaría a identificar períodos de mayor vulnerabilidad, cambios en infraestructura y acceso a servicios, fortaleciendo la planificación y focalización de intervenciones subnacionales.

Capítulo 6

Discusión

La modelación jerárquica bayesiana de la Razón de Mortalidad Materna (RMM) en los municipios de Colombia permitió identificar y caracterizar de manera más detallada las desigualdades territoriales, tanto en la ocurrencia de muertes maternas como en la magnitud de la RMM cuando estas ocurren. El modelo *hurdle*, con una parte Bernoulli para la presencia de muertes maternas y una parte lognormal para las RMM positivas, resultó adecuado para datos con muchos ceros y distribución muy asimétrica. Los resultados indican que la mortalidad materna no solo está relacionada con las condiciones socioeconómicas y el acceso a servicios, sino que también varía notablemente según la ubicación geográfica.

En la parte Bernoulli del modelo, se observa que la probabilidad de registrar al menos un caso de mortalidad materna se asocia de manera diferenciada con un conjunto específico de covariables sociodemográficas. En particular, el producto interno bruto per cápita (`pib_s`), el índice de pobreza multidimensional (`ipm_s`), la proporción de población con necesidades básicas insatisfechas en el ámbito rural (`nbi_rur_s`) y la cobertura de controles prenatales (`cont_prenat_s`) concentran la mayor evidencia estadística de asociación con la ocurrencia del evento. Estos resultados indican que la presencia de mortalidad materna a nivel municipal se relaciona principalmente con condiciones económicas, niveles de privación social y acceso efectivo a servicios de salud materna, mientras que otras covariables muestran un papel secundario o no estadísticamente identificable en este componente del modelo.

De manera complementaria, el análisis basado en la estimación posterior de los coeficientes de regresión permitió establecer que la relevancia de las covariables no es homogénea entre los distintos componentes del modelo. Los resultados muestran que variables como `pib_s`, `ipm_s` y `nbi_rur_s` presentan patrones de asociación diferenciados entre la probabilidad de ocurrencia de la mortalidad materna y la magnitud de la Razón de Mortalidad Materna una vez ocurrido el evento, mientras que la cobertura de controles prenatales (`cont_prenat_s`) adquiere relevancia particular en la caracterización del riesgo asociado a la ocurrencia. Esta diferenciación, que no es capturada por enfoques uniparte tradicionales, refuerza la pertinencia del enfoque *hurdle* para analizar fenómenos de salud pública con alta heterogeneidad territorial.

Por su parte, el componente del modelo que describe la magnitud de la Razón de Mortalidad Materna entre los municipios que sí registran mortalidad materna evidencia asociaciones principalmente con factores estructurales de carácter socioeconómico. En este componente, el índice de pobreza multidimensional (`ipm_s`), el

producto interno bruto per cápita (`pib_s`) y la proporción de población con necesidades básicas insatisfechas en zonas rurales (`nbi_rur_s`) concentran la mayor evidencia de asociación con la magnitud de la RMM condicionada a su ocurrencia. Estos resultados sugieren que, una vez ocurrido el evento, la intensidad del fenómeno se encuentra asociada a desigualdades persistentes en el desarrollo relativo de los territorios, más que a factores asociados exclusivamente a la ocurrencia inicial.

En cuanto a los efectos aleatorios, la varianza residual tanto a nivel departamental como municipal evidenció heterogeneidad no explicada por los predictores incluidos, lo que sugiere la presencia de factores no observados, institucionales, geográficos o de gestión que siguen influyendo en la variabilidad del riesgo y que el modelo no alcanza a capturar por completo. Al modelar explícitamente la variación departamental y municipal, la estructura jerárquica permitió identificar que la variabilidad de la mortalidad materna en Colombia no es aleatoria, sino que responde a diferencias territoriales asociadas a procesos sociales y sanitarios complejos.

El uso de métodos bayesianos mediante *Hamiltonian Monte Carlo* (HMC/NUTS) implementados en `Stan` aportó ventajas importantes, como la obtención de distribuciones posteriores completas, una mayor estabilidad ante modelos con alta asimetría y la posibilidad de incorporar múltiples efectos aleatorios. Esto permitió capturar simultáneamente la estructura jerárquica y el comportamiento irregular de los datos de mortalidad materna. Además, el modelo *hurdle* resolvió de manera adecuada el exceso de ceros, evitando sesgos que habrían surgido al emplear modelos uniparte tradicionales.Sin embargo, este tipo de modelación implica limitaciones a considerar.

Las principales limitaciones son: (1) la estimación con HMC es computacionalmente demandante, sobre todo en modelos complejos con múltiples niveles y parámetros correlacionados; y (2) el elevado número de ceros en los datos de mortalidad materna no siempre refleja ausencia real del evento, sino también subregistro en territorios con dificultades de acceso o capacidades institucionales limitadas, lo que puede dar una falsa impresión de bajo riesgo. Aunque el modelo incluye efectos aleatorios para reflejar la heterogeneidad territorial, la falta de covariables clínicas y de información sobre la calidad de los servicios obstétricos limita la interpretación de los factores que explican las diferencias observadas. La RMM depende de aspectos como la respuesta institucional, el transporte disponible, la capacidad para manejar emergencias y los tiempos de atención, elementos que no pueden capturarse plenamente con datos agregados a nivel municipal.

En conjunto, los resultados muestran que la mortalidad materna en Colombia sigue siendo un fenómeno marcado por profundas desigualdades territoriales. El enfoque multinivel bayesiano permitió captar estas variaciones con mayor precisión, revelando patrones que los modelos tradicionales suelen pasar por alto. Los hallazgos subrayan la necesidad de políticas que enfrenten simultáneamente las inequidades socioeconómicas y las barreras de acceso al sistema de salud, con el fin de reducir las brechas territoriales en mortalidad materna.

Capítulo 7

Anexos

Anexo 1. Participación de expertos en la selección de variables

Como complemento al proceso de revisión de literatura, se llevó a cabo una consulta con expertos mediante la metodología Delphi, con el propósito de obtener consenso sobre las variables relevantes para el análisis de la Razón de Mortalidad Materna (RMM). En este proceso participaron la Doctora Erika Johanna Cantor, estadística con formación en epidemiología y experiencia en métodos multinivel y análisis de datos en salud pública, y el Doctor Jaime Mosquera Restrepo, estadístico con trayectoria académica e investigativa en modelación aplicada.

Ambos expertos contribuyeron respondiendo las rondas de consulta, revisando los listados preliminares de variables y aportando criterios técnicos para clasificar los factores como indispensables o potencialmente relevantes. Su participación se integró de manera anónima en la metodología descrita en el capítulo metodológico, y se detalla aquí únicamente con fines de transparencia y documentación del proceso.

Anexo 2. Tabla de Estadísticas por Departamento

Departamento	RMM	Cob Edu Tot	Cob Acu Rural	Cob Acu Urb	Cob EE Rural	Cob EE Urb	Cob Edu Muj	NBI Rural	NBI Urb	IPM	Prom Ctrl Prenat	Porc Partos Per	PIB	Prop Pers NBI
amazonas	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	29.84	86.94	12.25	61.17	73.16	90.43	74.66	41.06	28.07	59.31	3.13	72.60	745.25	34.78
	51.69	22.45	4.67	7.62	14.22	3.18	16.50	7.08	4.79	11.00	0.49	12.37	573.06	7.20
	173.2 %	25.8 %	38.1 %	12.5 %	19.4 %	3.5 %	22.1 %	17.3 %	17.0 %	18.5 %	15.8 %	17.0 %	76.9 %	20.7 %
antioquia	126	126	126	126	126	126	126	126	126	126	126	126	126	126
	64.92	97.05	52.64	95.73	95.52	99.16	86.75	24.24	12.14	37.46	6.23	98.10	3497.14	17.98
	187.77	13.94	25.56	12.36	5.84	0.88	12.42	18.31	12.63	15.48	0.85	5.37	21829.30	14.43
	289.2 %	14.4 %	48.6 %	12.9 %	6.1 %	0.9 %	14.3 %	75.5 %	104.0 %	41.3 %	13.7 %	5.5 %	624.2 %	80.3 %
arauca	8	8	8	8	8	8	8	8	8	8	8	8	8	8
	89.14	102.80	24.12	92.79	70.11	95.27	92.58	43.12	25.36	44.19	4.62	94.67	2159.00	31.27
	78.39	18.20	18.86	4.99	26.19	4.24	16.02	6.71	5.24	4.26	0.49	3.70	2852.46	5.31
	87.9 %	17.7 %	78.2 %	5.4 %	37.4 %	4.4 %	17.3 %	15.6 %	20.6 %	9.6 %	10.6 %	3.9 %	132.1 %	17.0 %
atlántico	24	24	24	24	24	24	24	24	24	24	24	24	24	24
	53.53	92.28	56.34	93.74	85.37	98.10	84.29	26.96	18.26	36.73	5.63	99.65	5601.67	18.67
	115.33	18.66	27.04	3.90	13.21	2.55	18.45	12.18	8.70	13.14	0.36	0.45	16225.42	7.78
	215.4 %	20.2 %	48.0 %	4.2 %	15.5 %	2.6 %	21.9 %	45.2 %	47.6 %	35.8 %	6.5 %	0.4 %	289.7 %	41.7 %
bolívar	47	47	47	47	47	47	47	47	47	47	47	47	47	47
	124.73	105.66	40.27	77.30	77.06	97.77	91.85	46.82	38.16	56.77	5.23	99.06	2288.10	43.13
	247.13	12.43	27.44	28.02	19.87	1.93	10.86	17.66	19.14	11.02	0.64	0.96	9444.10	16.84
	198.1 %	11.8 %	68.2 %	36.2 %	25.8 %	2.0 %	11.8 %	37.7 %	50.2 %	19.4 %	12.2 %	1.0 %	412.7 %	39.0 %
boyacá	124	124	124	124	124	124	124	124	124	124	124	124	124	124
	30.72	93.58	60.88	98.63	92.32	98.82	87.69	18.76	8.02	38.03	5.79	98.24	655.86	15.26
	195.93	17.05	24.48	1.63	12.42	1.13	13.72	11.57	4.81	13.21	0.84	6.12	3839.32	9.72
	637.7 %	18.2 %	40.2 %	1.7 %	13.5 %	1.1 %	15.6 %	61.6 %	60.0 %	34.7 %	14.5 %	6.2 %	585.4 %	63.7 %
caldas	28	28	28	28	28	28	28	28	28	28	28	28	28	28
	123.51	97.20	56.91	97.69	97.70	99.10	86.56	15.86	8.01	32.21	6.74	98.88	1797.62	11.58
	346.61	9.36	14.88	2.91	1.51	1.65	8.04	4.59	3.01	7.18	0.66	2.21	5194.11	3.01
	280.6 %	9.6 %	26.1 %	3.0 %	1.5 %	1.7 %	9.3 %	29.0 %	37.5 %	22.3 %	9.8 %	2.2 %	288.9 %	26.0 %
caquetá	17	17	17	17	17	17	17	17	17	17	17	17	17	17
	20.21	83.61	14.71	95.60	60.33	96.46	76.68	37.28	18.66	51.36	4.35	91.95	713.10	28.11
	52.27	10.81	10.91	3.24	24.51	2.67	9.38	8.36	5.05	10.37	0.66	7.57	1564.94	8.73
	258.6 %	12.9 %	74.2 %	3.4 %	40.6 %	2.8 %	12.2 %	22.4 %	27.0 %	20.2 %	15.2 %	8.2 %	219.5 %	31.1 %
casanare	20	20	20	20	20	20	20	20	20	20	20	20	20	20
	78.94	91.57	27.82	95.37	76.43	97.53	83.90	30.58	13.14	37.58	4.65	94.70	2263.17	21.38
	150.93	14.32	17.15	6.57	15.90	2.07	13.87	11.79	5.58	10.51	0.74	7.87	5169.16	8.69
	191.2 %	15.6 %	61.6 %	6.9 %	20.8 %	2.1 %	16.5 %	38.6 %	42.5 %	28.0 %	15.9 %	8.3 %	228.4 %	40.6 %
cauca	43	43	43	43	43	43	43	43	43	43	43	43	43	43
	39.64	90.83	42.70	89.26	84.60	97.78	80.47	22.07	20.27	45.08	5.44	90.18	1262.07	22.22
	150.95	15.78	27.55	18.96	20.40	2.33	14.40	10.78	18.20	15.64	0.64	12.26	4402.29	12.83
	380.8 %	17.4 %	64.5 %	21.2 %	24.1 %	2.4 %	17.9 %	48.8 %	89.8 %	34.7 %	11.8 %	13.6 %	348.8 %	57.7 %
cesar	26	26	26	26	26	26	26	26	26	26	26	26	26	26
	62.18	97.44	48.97	94.54	80.01	98.13	88.91	35.41	22.98	45.05	4.96	96.26	2635.82	26.89
	104.03	10.89	20.37	4.21	20.07	2.25	9.29	14.71	6.61	8.87	0.74	11.17	6954.60	9.27
	167.3 %	11.2 %	41.6 %	4.5 %	25.1 %	2.3 %	10.4 %	41.5 %	28.8 %	19.7 %	14.9 %	11.6 %	263.9 %	34.5 %
chocó	31	31	31	31	31	31	31	31	31	31	31	31	31	31
	224.22	101.73	25.42	52.90	56.66	92.32	83.30	56.21	55.05	66.54	2.74	61.83	437.12	57.19
	321.40	23.66	18.00	30.63	26.46	15.83	20.00	15.65	24.66	11.60	1.40	27.07	1256.84	14.16
	143.3 %	23.3 %	70.8 %	57.9 %	46.7 %	17.1 %	24.0 %	27.8 %	44.8 %	17.4 %	51.2 %	43.8 %	287.5 %	24.8 %
cundinamarca	117	117	117	117	117	117	117	117	117	117	117	117	117	117
	111.10	103.25	59.63	98.55	94.20	98.94	92.44	12.72	6.62	27.46	5.86	99.38	1615.96	10.38
	604.86	18.67	27.34	3.05	5.26	1.20	18.34	7.89	4.08	12.71	0.56	1.29	9267.15	6.57
	544.4 %	18.1 %	45.8 %	3.1 %	5.6 %	1.2 %	19.8 %	62.0 %	61.6 %	46.3 %	9.5 %	1.3 %	573.5 %	63.3 %
córdoba	31	31	31	31	31	31	31	31	31	31	31	31	31	31
	87.72	97.51	37.61	86.50	92.44	98.44	86.14	51.66	31.66	53.16	5.41	98.81	1765.34	43.08
	123.88	11.75	22.20	16.44	7.06	1.32	10.68	20.94	10.40	11.99	0.50	2.11	5124.91	17.57
	141.2 %	12.1 %	59.0 %	19.0 %	7.6 %	1.3 %	12.4 %	40.5 %	32.9 %	22.6 %	9.2 %	2.1 %	290.3 %	40.8 %
guainía	3	3	3	3	2	2	3	2	2	2	2	3	3	3
	24.31	79.32	5.65	36.05	35.13	84.69	68.55	77.83	39.77	70.44	2.04	76.39	342.97	61.85
	42.11	2.91	2.37	2.88	0.43	0.00	2.50	2.00	0.00	4.59	0.20	17.33	241.74	12.38
	173.2 %	3.7 %	42.0 %	8.0 %	1.2 %	0.0 %	3.6 %	2.6 %	0.0 %	6.5 %	10.0 %	22.7 %	70.5 %	20.0 %
guaviare	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	37.24	82.40	8.35	70.28	36.08	93.43	69.36	47.65	18.91	53.03	3.77	91.12	496.89	32.76
	83.28	25.61	4.16	18.49	13.48	6.56	23.31	4.18	8.01	9.64	0.58	3.70	526.66	7.30
	223.6 %	31.1 %	49.9 %	26.3 %	37.4 %	7.0 %	33.6 %	8.8 %	42.4 %	18.2 %	15.4 %	4.1 %	106.0 %	22.3 %
huila	38	38	38	38	38	38	38	38	38	38	38	38	38	38
	19.35	95.77	58.64	98.30	89.17	98.41	89.54	18.45	11.94	40.98	5.96	97.88	1353.82	15.80
	52.53	11.15	17.77	1.24	8.33	1.17	9.68	6.41	3.54	9.57	0.45	5.67	4423.43	4.51
	271.4 %	11.6 %	30.3 %	1.3 %	9.3 %	1.2 %	10.8 %	34.8 %	29.7 %	23.4 %	7.6 %	5.8 %	326.7 %	28.5 %
la guajira	16	16	16	16	16	16	16	16	16	16	16	16	16	16
	139.17	96.82	22.06	80.21	48.15	92.88	87.32	57.14	30.63	50.33	3.98	94.43	2502.75	40.60
	133.22	18.01	18.19	22.60	25.49	6.23	15.57	19.55	20.18	18.30	0.87	6.42	5122.33	21.40
	95.7 %	18.6 %	82.5 %	28.2 %	52.9 %	6.7 %	17.8 %	34.2 %	65.9 %	36.4 %	21.9 %	6.8 %	204.7 %	52.7 %
magdalena	31	31	31	31	31	31	31	31	31	31	31	31	31	31
	113.67	120.53	44.28	85.90	77.82	98.70	89.78	40.97	34.35	55.28	5.39	99.03	1375.98	34.29
	209.38	15.46	21.88	21.35	19.34	1.14	10.80	13.14	17.11	9.84	0.67	1.83	4116.34	10.67
	184.2 %	12.8 %	49.4 %	24.8 %	24.9 %	1.2 %	12.0 %	32.1 %	49.8 %	17.8 %	12.4 %	1.8 %	299.2 %	31.1 %
meta	30	30	30	30	30	30	30	30	30	30	30	30	30	30
	116.26	99.48	36.41	90.85	70.35	97.33	90.62	29.94	13.26	36.32	4.63	95.53	3512.01	21.85
	279.91	15.04	25.70	12.29	27.95	2.37	14.41	19.20	7.49	14.05	1.00	7.97	10023.71	14.76
	240.8 %	15.1 %	70.6 %	13.5 %	39.7 %	2.4 %	15.9 %	64.1 %	56.4 %	38.7 %	21.7 %	8.3 %	285.4 %	67.5 %
nariño	65	65	65	65	65	65	65	65	65	65	65	65	65	65
	60.57	86.72	59.08	86.37	84.74	98.43	79.55	25.45	26.33	45.73	6.15	96.71	700.99	26.59
	209.07	18.11	31.61	28.40	25.98	3.45	18.45	17.38	24.37	16.13	1.16	6.04	3056.66	19.52
	345.2 %	20.9 %	53.5 %	32.9 %	30.7 %	3.5 %	23.2 %	68.3 %	92.6 %	35.3 %	18.8 %	6.2 %	436.0 %	73.4 %
norte de santander	41	41	41	41	41	41	41	41	41	41	41	41	41	41
	231.17	91.73	31.96	96.57	88.06	98.33	85.05	31.46	14.42	49.24	4.95	96.36	1197.58	25.33
	579.97	14.85	17.36	3.97	10.43	1.47	12.81	12.17	6.50	14.74	0.78	4.45	4354.17	10.92
	250.9 %	16.2 %	54.3 %	4.1 %	11.8 %	1.5 %	15.1 %	38.7 %	45.1 %	29.9 %	15.8 %	4.6 %	363.6 %	43.1 %
putumayo	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	75.70	89.49	41.76	77.73	62.78	97.17	81.81	24.37	11.82	38.24	5.40	94.48	816.83	17.89
	114.28	17.48	35.79	32.23	27.56	2.18	16.97	11.78	4.74	11.89	1.05	4.26	1507.81	7.61
	151.0 %	19.5 %	85.7 %	41.5 %	43.9 %	2.2 %	20.7 %	48.4 %	40.1 %	31.1 %	19.4 %	4.5 %	184.6 %	42.5 %
quindío	13	13	13	13	13	13	13	13	13	13	13	13	13	13
	248.58	110.93	75.98	99.17	97.59	99.30	101.74	10.79	6.90	23.42	6.07	98.92	1987.05	8.32
	764.73	23.09	25.71	0.54	2.07	0.37	23.51	3.56	1.82	5.99	0.44	1.68	3785.16	2.29
	307.6 %	20.8 %	33.8 %	0.5 %	2.1 %	0.4 %	23.1 %	33.0 %	26.4 %	25.6 %	7.3 %	1.7 %	190.5 %	27.5 %
risaralda	15	15	15	15	15	15	15	15	15	15	15	15	15	15
	35.24	103.70	58.95	98.99	97.56	99.70	90.51	20.66	6.79	34.29	5.71	90.15	3357.01	15.27
	82.38	12.15	19.07	1.06	2.48	0.19	9.71	16.34	1.59	16.43	1.45	19.45	7290.97	14.14
	233.7 %	11.7 %	32.3 %	1.1 %	2.5 %	0.2 %	10.7 %	79.1 %	23.4 %	47.9 %	25.5 %	21.6 %	217.2 %	92.6 %
san andrés y providencia	2	2	2	2	2	2	2	2	2	2	2	2	1	1
	0.00	91.27	67.86	60.37	98.49	99.57	85.11	5.75	15.62	15.09	6.43	100.00	199.45	8.54
	0.00	16.31	28.60	41.18	0.95	0.04	15.27	0.18	5.19	1.97	0.28	0.00	NA	NA
	NaN %	17.9 %	42.1 %	68.2 %	1.0 %	0.0 %	17.9 %	3.2 %	33.2 %	13.1 %	4.4 %	0.0 %	NA %	NA %
santander	88	88	88	88	88	88	88	88	88	88	88	88	88	88
	63.13	96.74	49.91	97.32	97.19	99.07	87.64	19.04	8.74	37.55	5.62	98.89	2115.63	15.46
	359.55	17.99	21.14	4.67	1.87	1.03	15.50	8.12	4.63	11.40	0.55	2.07	11204.46	7.36
	569.6 %	18.6 %	42.4 %	4.8 %	1.9 %	1.0 %	17.7 %	42.6 %	53.0 %	30.4 %	9.7 %	2.1 %	529.6 %	47.7 %
sucre	27	27	27	27	27	27	27	27	27	27	27	27	27	27
	38.10	104.26	66.99	92.07	91.20	97.89	90.33	37.51	32.91	53.25	5.34	99.46	912.36	34.84
	88.25	9.94	17.86	9.23	4.33	2.19	9.73	11.54	12.19	10.99	0.58	0.67	2543.55	11.17
	231.6 %	9.5 %	26.7 %	10.0 %	4.7 %	2.2 %	10.8 %	30.8 %	37.1 %	20.6 %	10.9 %	0.7 %	278.8 %	32.1 %
tolima	48	48	48	48	48	48	48	48	48	48	48	48	48	48
	39.87	102.47	53.54	97.91	94.10	98.79	93.17	20.69	11.99	38.31	5.51	97.14	1353.41	16.92
	104.97	13.25	20.57	3.24	3.17	0.90	11.69	6.82	5.01	11.93	0.64	4.76	5096.68	6.92
	263.3 %	12.9 %	38.4 %	3.3 %	3.4 %	0.9 %	12.5 %	33.0 %	41.8 %	31.1 %	11.6 %	4.9 %	376.6 %	40.9 %

TABLA 7.1: Tabla de Estadísticas por Departamento
n, Media, Desviación estándar y Coeficiente de variación

Departamento	RMM	Cob Edu Tot	Cob Acu Rural	Cob Acu Urb	Cob EE Rural	Cob EE Urb	Cob Edu Muj	NBI Rural	NBI Urb	IPM	Prom Ctrl Prenat	Porc Partos Per	PIB	Prop Pers NBI
valle del cauca	43	43	43	43	43	43	43	43	43	43	43	43	43	43
	69.80	85.73	73.90	98.21	95.45	99.11	75.93	12.57	7.03	23.51	6.36	97.97	6771.57	9.23
	152.65	9.33	15.19	2.88	6.52	1.34	8.40	5.39	2.55	8.44	0.59	4.09	24773.14	3.42
	218.7 %	10.9 %	20.6 %	2.9 %	6.8 %	1.4 %	11.1 %	42.9 %	36.3 %	35.9 %	9.2 %	4.2 %	365.8 %	37.1 %
vaupés	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	59.88	47.28	3.87	82.70	15.35	92.27	41.95	82.59	30.92	80.08	2.13	54.80	202.89	70.89
	119.76	27.46	1.20	9.24	8.35	2.70	20.47	14.00	11.25	4.51	0.58	22.32	204.23	5.46
	200.0 %	58.1 %	30.9 %	11.2 %	54.4 %	2.9 %	48.8 %	17.0 %	36.4 %	5.6 %	27.1 %	40.7 %	100.7 %	7.7 %
vichada	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	135.63	94.57	10.75	87.35	19.92	92.50	86.70	71.74	31.83	68.75	2.70	72.15	416.85	52.21
	188.06	34.88	7.57	8.36	13.35	2.75	33.81	15.84	4.40	15.16	1.55	19.48	371.32	14.60
	138.7 %	36.9 %	70.4 %	9.6 %	67.0 %	3.0 %	39.0 %	22.1 %	13.8 %	22.1 %	57.2 %	27.0 %	89.1 %	28.0 %

TABLA 7.2: Tabla de Estadísticas por Departamento
n, Media, Desviación estándar y Coeficiente de variación

Anexo 3. Graficos densidades de las covaribles

Las siguientes figuras presentan las densidades posteriores de los coeficientes de regresión estimados en el modelo hurdle multinivel. Estas se incluyen como material complementario para ilustrar la forma y dispersión de las distribuciones posteriores, mientras que la inferencia principal se basa en los resúmenes presentados en el cuerpo del documento.

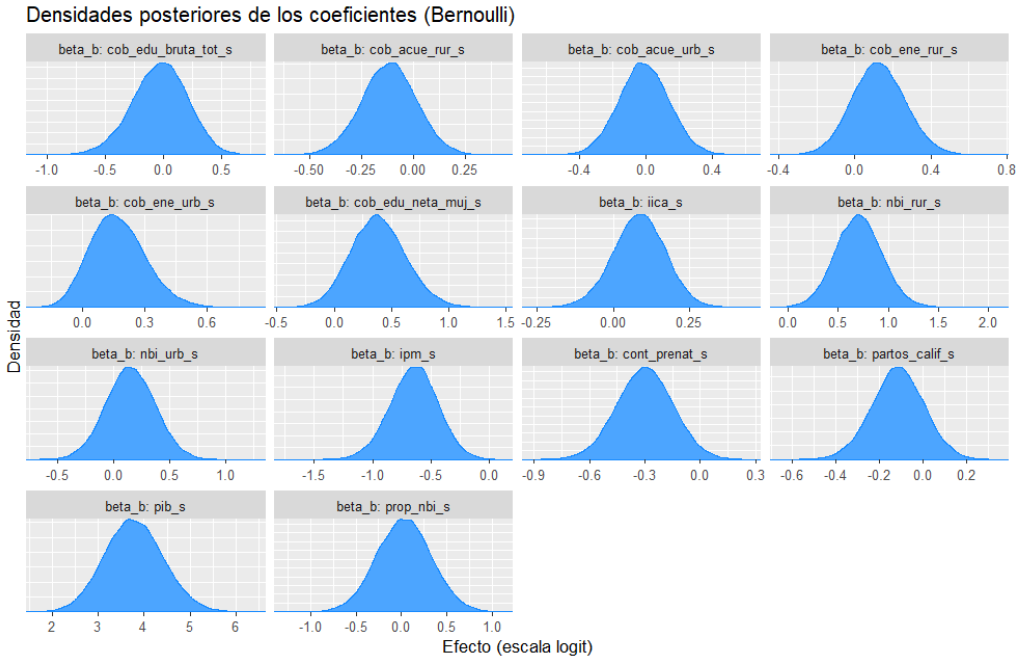

FIGURA 7.1: Densidades posteriores de los coeficientes de regresión correspondientes al componente Bernoulli del modelo hurdle multinivel.

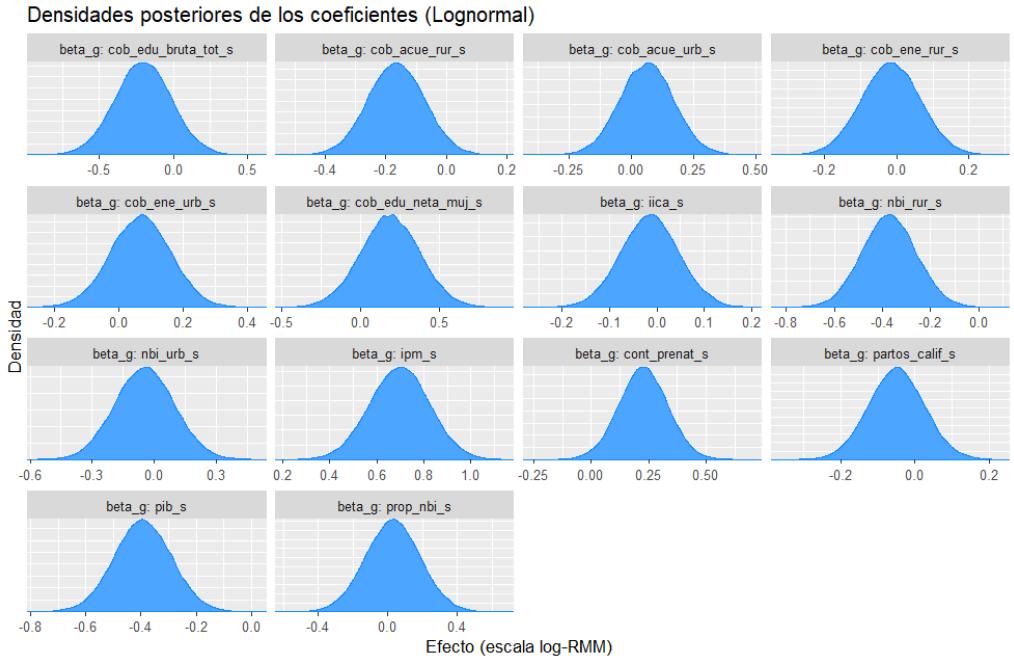

FIGURA 7.2: Densidades posteriores de los coeficientes de regresión correspondientes al componente lognormal del modelo hurdle multinivel.

Anexo 4. Creación base de datos - Stan

```
# ================================================================ #
# 0. Cargar base (como en tu script INLA)
# ================================================================ #

library(readxl); library(dplyr); library(rstan)

# file.choose()
Archivo    <- # Direccion donde se encuentra la base de datos
Base.Unica <- read_excel(Archivo, sheet = "Base", .name_repair = "minimal")

Base.Unica <- Base.Unica %>% rename(
    Departamento      = "Departamento",
    Municipio         = "Municipio",
    RMM               = "RMM",
    cod_mpio          = "Codigo Municipio",
    cob_edu_bruta_tot = "Cobertura bruta en educación - Total (2022,12)",
    cob_acue_rur      = "Cobertura de acueducto rural (Censo) (2018,12)",
    cob_acue_urb      = "Cobertura de acueducto urbana (Censo) (2018,12)",
    cob_ene_rur       = "Cobertura de Energía Eléctrica Rural (Censo) (2018,12)",
    cob_ene_urb       = "Cobertura de Energía Eléctrica Urbana (Censo) (2018,12)",
    cob_edu_neta_muj  = "Cobertura neta en educación total - Mujeres (2022,12)",
    iica              = "Índice de Incidencia del Conflicto Armado - IICA (2021,12)",
    nbi_rur           = "Índice de Necesidades Básicas Insatisfechas - NBI - en el área rural
    ↪  (2018,12)",
    nbi_urb           = "Índice de Necesidades Básicas Insatisfechas - NBI - en el área urbana
    ↪  (2018,12)",
    ipm               = "Índice de pobreza multidimensional - IPM\r\r\n (2018,12)",
    cont_prenat       = "Promedio de controles prenatales (2020,12)",
    partos_calif      = "Porcentaje de partos atendidos por personal calificado (2021,12)",
    pib               = "PIB.2023p",
    prop_nbi          = "Prop de Personas en NBI (%).1",
    irca              = "IRCA.2023",
    long              = "Longitud",
    lat               = "Latitud"
  )

```

```
# ================================================================ #
# 1. Crear Z, IDs y estandarizar covariables
# ================================================================ #

covs <- c("cob_edu_bruta_tot","cob_acue_rur","cob_acue_urb","cob_ene_rur",
          "cob_ene_urb","cob_edu_neta_muj","iica","nbi_rur","nbi_urb",
          "ipm","cont_prenat","partos_calif","pib","prop_nbi","irca")

Base.Unica <- Base.Unica %>%
  mutate(Departamento = as.factor(Departamento),
         Municipio    = as.factor(Municipio),
         dep_id       = as.integer(Departamento),
         mpio_id      = as.integer(Municipio),
         Z            = as.integer(RMM > 0))

for (v in covs) {
  if (v %in% names(Base.Unica)) {
    Base.Unica[[v]]  <- as.numeric(as.character(Base.Unica[[v]]))
    Base.Unica[[paste0(v,"_s")]] <- as.vector(scale(Base.Unica[[v]]))
  }}

# ================================================================ #
# 2. Seleccionar las mismas covariables que en INLA (sin irca)
# ================================================================ #

covs_names_s <- c("cob_edu_bruta_tot_s", "cob_acue_rur_s", "cob_acue_urb_s",
                  "cob_ene_rur_s", "cob_ene_urb_s", "cob_edu_neta_muj_s",
                  "iica_s", "nbi_rur_s", "nbi_urb_s", "ipm_s",
                  "cont_prenat_s", "partos_calif_s", "pib_s", "prop_nbi_s")

# ================================================================ #
# 3. Eliminar NAs y construir X
# ================================================================ #

vars_needed <- c("RMM", "Z", "dep_id", "mpio_id", covs_names_s)
Base.Unica.stan <- Base.Unica %>%
  dplyr::select(all_of(vars_needed)) %>% na.omit()
X_matrix <- as.matrix(Base.Unica.stan[, covs_names_s])

# ================================================================ #
# 4. Índices y datos positivos
# ================================================================ #

ii_pos <- which(Base.Unica.stan$RMM > 0)
y_pos  <- log(Base.Unica.stan$RMM[ii_pos])
pos_idx_lookup         <- rep(0, nrow(Base.Unica.stan))
pos_idx_lookup[ii_pos] <- seq_along(ii_pos)

# ================================================================ #
# 5. Reindexar IDs (contiguos)
# ================================================================ #

Base.Unica.stan <- Base.Unica.stan %>%
  mutate(dep_id_stan  = as.integer(as.factor(dep_id)),
         mpio_id_stan = as.integer(as.factor(mpio_id)))

# ================================================================ #
# 6. Lista final de datos para Stan
# ================================================================ #

stan_data <- list(
  N_total = nrow(Base.Unica.stan),
  N_pos   = length(ii_pos),
  K       = ncol(X_matrix),
  N_dep   = max(Base.Unica.stan$dep_id_stan),
  N_mpio  = max(Base.Unica.stan$mpio_id_stan),
  Z       = Base.Unica.stan$Z,
  y_pos   = y_pos,
  X       = X_matrix,
  dep_id  = Base.Unica.stan$dep_id_stan,
  mpio_id = Base.Unica.stan$mpio_id_stan,
  ii_pos  = ii_pos,
```

```
  pos_idx_lookup = pos_idx_lookup)
```

Anexo 5. Modelo en formato C++

```
data {
  int<lower=1> N_total;
  int<lower=1> N_pos;
  int<lower=1> K;
  int<lower=1> N_dep;
  int<lower=1> N_mpio;

  array[N_total] int<lower=0, upper=1> Z;
  vector[N_pos] y_pos;
  matrix[N_total, K] X;

  array[N_total] int<lower=1, upper=N_dep>  dep_id;
  array[N_total] int<lower=1, upper=N_mpio> mpio_id;

  array[N_pos] int<lower=1, upper=N_total> ii_pos;
  array[N_total] int<lower=0, upper=N_pos> pos_idx_lookup;
}

parameters {
  real alpha_0;
  vector[K] beta_b;
  vector[N_dep]  alpha_dep_raw;
  vector[N_mpio] alpha_mpio_raw;
  real<lower=0> sigma_dep_b;
  real<lower=0> sigma_mpio_b;

  real mu_0;
  vector[K] beta_g;
  vector[N_dep]  mu_dep_raw;
  vector[N_mpio] mu_mpio_raw;
  real<lower=0> sigma_dep_g;
  real<lower=0> sigma_mpio_g;

  real<lower=0> sigma_y;
}

transformed parameters {
  vector[N_total] eta_b;
  vector[N_total] eta_g;

  vector[N_dep]  alpha_dep  = alpha_dep_raw * sigma_dep_b;
  vector[N_mpio] alpha_mpio = alpha_mpio_raw * sigma_mpio_b;
  vector[N_dep]  mu_dep     = mu_dep_raw * sigma_dep_g;
  vector[N_mpio] mu_mpio    = mu_mpio_raw * sigma_mpio_g;

  for (n in 1:N_total) {
    eta_b[n] = alpha_0 + X[n] * beta_b + alpha_dep[dep_id[n]] + alpha_mpio[mpio_id[n]];
    eta_g[n] = mu_0    + X[n] * beta_g + mu_dep[dep_id[n]]    + mu_mpio[mpio_id[n]];
  }
}

model {
  // Priors para Desviaciones Estándar de la parte Binomial (Z)
  sigma_dep_b  ~ cauchy(0, 1);    // Cauchy(0, 1) - Flexible para el efecto superior
  sigma_mpio_b ~ cauchy(0, 0.5);  // Cauchy(0, 0.5) - Más estricta para el efecto anidado

  // Priors para Desviaciones Estándar de la parte Log-Normal (y_pos)
  sigma_dep_g  ~ cauchy(0, 0.25); // Cauchy(0, 0.25) - Basado en su prior original
 sigma_mpio_g ~ normal(0, 0.05);// sigma_mpio_g ~ cauchy(0, 0.2);  // Cauchy(0, 0.2) - Más
  ↪  estricta, corrige la no-convergencia de sigma_mpio_g

  // Prior para la Desviación Estándar Residual (sigma_y)
  sigma_y ~ normal(0, 1.5);   // Reemplaza el Prior de Jeffreys por una Half-Cauchy
   ↪  regularizadora

```

```
  // Nota: Las priors para los "raw" effects permanecen std_normal()
  alpha_dep_raw  ~ std_normal();
  alpha_mpio_raw ~ std_normal();
  mu_dep_raw     ~ std_normal();
  mu_mpio_raw    ~ std_normal();

  // Priors para Interceptos y Coeficientes Fijos (si son débilmente informativas)
  alpha_0 ~ normal(0, 1.0);
  mu_0    ~ normal(0, 1.0);
  beta_b  ~ normal(0, 1.0);  // Asumiendo que las covariables (X) están estandarizadas (muy
  ↪  importante)
  beta_g  ~ normal(0, 1.0);

  Z ~ bernoulli_logit(eta_b);

  for (i in 1:N_pos)
    y_pos[i] ~ normal(eta_g[ii_pos[i]], sigma_y);
}

generated quantities {
  vector[N_total] log_lik;
  vector[N_total] E_RMM_pred;

  for (n in 1:N_total) {
    real p_Z1 = inv_logit(eta_b[n]);
    real E_pos = exp(eta_g[n] + 0.5 * square(sigma_y));
    E_RMM_pred[n] = p_Z1 * E_pos;

    if (Z[n] == 0) {
      log_lik[n] = bernoulli_logit_lpmf(0  eta_b[n]);
  else
 int pos_idx = pos_idx_lookup[n];
 log_lik[n] = bernoulli_logit_lpmf(1 eta_b[n])
                   + normal_lpdf(y_pos[pos_idx] | eta_g[n], sigma_y);
    }
  }
}
```

Anexo 6. Algoritmo principal

```
# ================================================================ #
# ---------------------------------------------------------------- #
# Pasos inciales
# ---------------------------------------------------------------- #
# ================================================================ #

library(ggplot2); library(bayesplot); library(dplyr)
library(tidyr); library(posterior);library(bayesplot)
library(ggplot2);library(patchwork);library(bayesplot)
library(ggtext);library(ggplot2); library(patchwork)

# Seleccionar la Version de Rstools (Una unica ocasion)
# ---------------------------------------------------------------- #

Sys.which("make")  # file.edit("~/.Rprofile")
# Sys.setenv(PATH = "C:/rtools44/usr/bin;C:/rtools44/mingw64/bin")

# Configurar directorio de trabajo
# ---------------------------------------------------------------- #

# choose.files()
Directorio <- "C:/Users/Jose Alejandro E.M/OneDrive/Proyecto De Grado/RStudio/Modelo Stan"
getwd();setwd(Directorio);getwd()

list.files()
source("Data_STAN_RMM.R" )   # Anexo 4

# Instalar el paquete cmdstanr
# ---------------------------------------------------------------- #
```

```
30
31  # Instala el paquete cmdstanr desde el repositorio oficial de Stan
32  # install.packages("cmdstanr", repos =
    ↪  c("https://mc-stan.org/r-packages/",getOption("repos")))
33
34  library(cmdstanr)
35  # El proceso (+) solo se lleva a cabo en una sola ocasion
36  # cmdstanr::install_cmdstan(cores = 4,overwrite = TRUE, quiet = FALSE)
37  # cmdstanr::cmdstan_version()
38
39  cmdstan_path()      # 1. Donde esta instalado el "cmdstan"
40  # rebuild_cmdstan() # 2. Reconstruir cmdstan
41
42  # ================================================================ #
43  # ---------------------------------------------------------------- #
44  # 1. Modelo Hurled ----
45  # ---------------------------------------------------------------- #
46  # ================================================================ #
47
48  mod_hurdle <- cmdstan_model("hurdle_rmm.stan") # Modelo y especificaciones del mismo
49
50  fit_hurdle_final <- mod_hurdle$sample(
51    data            = stan_data,  # Datos que el modelo Stan necesita (lista declarada en
      ↪  'data{}')
52    chains          = 4,       # Número de cadenas MCMC independientes
53    parallel_chains = 4,       # Cuántas cadenas correr al mismo tiempo (acelera el proceso)
54    iter_warmup     = 5000,    # Iteraciones de calentamiento (ajuste interno del sampler)
55    iter_sampling   = 10000,   # Iteraciones que sí se usan para la posterior (muestras finales)
56    seed            = 123,     # Semilla para obtener resultados reproducibles
57    adapt_delta     = 0.995,   # Controla la precisión del sampler; alto = menos divergencias
58    max_treedepth   = 15)      # Profundidad máxima de NUTS; evita saturación en modelos
      ↪  complejos
59
60  class(fit_hurdle_final)
61
62  # ---------------------------------------------------------------- #
63  # Guardar el modelo
64  # ---------------------------------------------------------------- #
65
66  # saveRDS(fit_hurdle_final, file = "modelo_hurdle_guardado.rds") # Guardar
67  fit_hurdle_final <- readRDS("modelo_hurdle_guardado.rds")       # Cargar
68  class(fit_hurdle_final2)
69
70  # ================================================================ #
71  ## 1.1.Parametros principales del modelo multinivel hurled  ----
72  # ================================================================ #
73
74  pars_main <- c("alpha_0","mu_0",              # Interceptos: parte binaria (0) y parte
    ↪  continua (0)
75                 "sigma_dep_b","sigma_mpio_b",  # Desv. estándar efectos aleatorios
                   ↪  (departamento/municipio) parte binaria
76                 "sigma_dep_g","sigma_mpio_g",  # Desv. estándar efectos aleatorios
                   ↪  (departamento/municipio) parte gaussiana
77                 "sigma_y")                     # Desv. estándar residual de la parte gaussiana
                   ↪  (ruido)
78
79  # Resumen. Parámetros principales (pars_main)
80  fit_hurdle_final$summary(pars_main)
81
82  # Resumen. Indicadores de diagnóstico MCMC
83  #   (1). Rhat (Gelman-Rubin diagnostic)          - Convergencia (Ecxelente: 1.00 - 1.01)
84  #   (2). ESS_bulk (Effective Sample Size - bulk) - muestreo en la zona central (ess_bulk alt)
85  #   (3). ESS_tail (Effective Sample Size - tail) - muestreo en las colas (ess_tail alto)
86
87  fit_hurdle_final$summary(pars_main)[, c("variable","rhat","ess_bulk","ess_tail")]
88
89  #Resumen. Diagnósticos del muestreo NUTS:
90  #   (1) "num_divergent":     Número de divergencias durante el muestreo
    ↪  - 0 = perfecto, no hubo problemas de exploración.
91  #   (2) "num_max_treedepth": Número de veces que NUTS alcanzó el máximo de profundidad del
    ↪  árbol - 0 = nunca se saturó, trayectorias adecuadas.
92  #   (3) "ebfmi":             E-BFMI (Energy Bayesian Fraction of Missing Information)
    ↪  - >0.3 aceptable, >0.8 excelente (muestreo es muy estable y bien mezclado)
```

```
93
94  fit_hurdle_final$diagnostic_summary()
95
96  # Cálculo y gráfico de la RMM predicha (media posterior)
97  draws           <- fit_hurdle_final$draws("E_RMM_pred")         # muestras que Stan genera de
    ↪  la distribución posterior de cada parámetro. Cada draw = un posible valor que el parámetro
    ↪  puede tomar según el modelo y los datos.
98  E_RMM_pred_mean <- posterior::summarise_draws(draws)$mean       # valor esperado de RMM
    ↪  predicha por cada unidad
99  E_RMM_pred_mean <- fit_hurdle_final$summary("E_RMM_pred")$mean  # Extraer directamente
100
101   # Valor esperado de RMM por unidad (Municipios)
102
103   plot(E_RMM_pred_mean,type = "l")
104
105   plot(E_RMM_pred_mean, type = "n", xlab = "Unidades", # Municipios
106        ylab = "RMM Predicha", las = 1);grid()
107   lines(E_RMM_pred_mean, col = "steelblue", lwd = 2)
108
109   # Valor esperado de RMM por unidad (Departamentos)
110
111   E_RMM_pred_dep <- data.frame(dep_id   = stan_data$dep_id,          # Vector para identificar
      ↪  al departamento
112                                RMM_pred = E_RMM_pred_mean) %>%      # Media posterior por
                                   ↪  unidad (Departamento)
113     group_by(dep_id) %>% summarise(RMM_pred_mean = mean(RMM_pred))  # media por departamento
114
115   plot(E_RMM_pred_dep, type = "n", xlab = "Unidades", # Municipios
116        ylab = "RMM Predicha", las = 1);grid()
117   lines(E_RMM_pred_dep, col = "steelblue", lwd = 2)
118
119 # ------------------------------------------------------------------ #
120 # ------------------------------------------------------------------ #
121
122 library(cmdstanr) ; library(posterior)  # para as_draws_*
123 library(bayesplot)                      # para mcmc_*()
124 color_scheme_set("brightblue")          # opcional
125
126 # extraer draws
127 draws <- fit_hurdle_final$draws()
128
129 # ================================================================== #
130 ## 1.2. Histogramas y densidades marginales ----
131 # ================================================================== #
132
133 mcmc_hist(draws, pars = pars_main)   # Histogramas
134 mcmc_dens(draws, pars = pars_main)   # Densidades
135 mcmc_trace(draws, pars = pars_main)
136
137 # Histogramas marginales de los parametros principales
138 # ------------------------------------------------------------------ #
139
140 {windows(width = 30, height = 20)
141 mcmc_hist(draws, pars = pars_main,
142           facet_args = list(ncol = 3, nrow = 3),
143           binwidth   = NULL,          # deja que bayesplot decida
144           alpha      = 1,           # transparencia
145           fill       = "skyblue") +   # color del histograma
146   theme_minimal() +
147   theme(
148     panel.spacing    = unit(0.5, "cm"),
149    strip.background = element_rect(fill = "lightgray", color = "black"),
150     strip.text       = element_text(face = "bold", size = 10),
151     panel.border     = element_rect(color = "gray20", fill = NA, size = 0.1)) +
152   scale_y_continuous(expand = expansion(mult = c(0, 0.1)))}
153
154 # Densidades marginales de los parametros principales
155 # ------------------------------------------------------------------ #
156
157 {windows(width = 30, height = 20)
158 mcmc_dens(draws, pars = pars_main,
159           facet_args  = list(ncol = 3, nrow = 3),  # Pantalla
```

```
160           alpha       = 1) +                        # Transparencia
161   theme_minimal() +
162   geom_density(aes(y = ..density..), color = "black", size = 0.2,fill = "steelblue") +
163   theme(
164     panel.spacing    = unit(0.5, "cm"),
165    strip.background = element_rect(fill = "lightgray", color = "black"),
166     strip.text       = element_text(face = "bold", size = 10),
167     panel.border     = element_rect(color = "gray20", fill = NA, size = 0.1)) +
168   scale_y_continuous(expand = expansion(mult = c(0, 0.1)))  }
169
170 #  Traceplots de los  parametros principales
171 # ------------------------------------------------------------------ #
172
173 {windows(width = 30, height = 20);
174 mcmc_trace(draws, pars = pars_main,
175            facet_args = list(ncol = 3, nrow = 3)) +
176   theme_minimal() +
177   theme(
178     panel.spacing    = unit(0.5, "cm"),
179    strip.background = element_rect(fill = "lightgray", color = "black"),
180     strip.text       = element_text(face = "bold", size = 10),
181     panel.border     = element_rect(color = "gray20", fill = NA, size = 0.1),
182     legend.position  = "none")}
183
184 # Graficar áreas
185 # ------------------------------------------------------------------ #
186
187 mcmc_areas(draws, pars = pars_main) +
188   theme_minimal(base_size = 12) +
189   theme(
190     panel.border  = element_rect(color = "gray20", fill = NA, size = 0.2),  # recuadro
191     panel.spacing = unit(0.5, "cm"))
192
193 # Opcion 2
194
195 library(bayesplot); library(ggtext); library(patchwork)
196
197 plots_list <- lapply(pars_main, function(p){
198   mcmc_areas(draws, pars       = p,
199              # prob       = 0.9,     # 90% intervalo de credibilidad
200              # prob_outer = 1,       # interno
201              # point_est  = "median"
202   ) + labs(title = p) + theme_minimal(base_size = 8) +
203     theme(
204       plot.title = element_textbox_simple(fill = "lightgray", color = "black", halign = 0.5,
205         face = "bold", padding = margin(t = 2, r = 5, b = 2, l = 5),linewidth = 0.2, linetype
          ↪  = 1),
206       panel.border = element_rect(color = "gray20", fill = NA, size = 0.2),
207       axis.title.y = element_blank(),
208       axis.text.y  = element_blank())})
209
210 {windows(width = 30, height = 20); wrap_plots(plots_list, ncol = 3)}
211
212 # ------------------------------------------------------------------ #
213 # final_plot <- wrap_plots(plots_list, ncol = 3)
214 # ggsave("mis_graficos.png", plot = final_plot, width = 30, height = 20, units = "cm", dpi =
    ↪  600)
215 # ------------------------------------------------------------------ #
216
217 # :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::: #
218 ### (Opcion 2). Histogramas y densidades marginales ----
219 # :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::: #
220
221 library(posterior)
222
223 df2_long <- df2 %>% as.data.frame() %>%      # convierte a data.frame normal
224   mutate(iter = row_number()) %>%
225   pivot_longer(cols = all_of(pars_main),names_to = "variable", values_to = "value")
226
227 # Etiquetas matemáticas
228 pars_labels <- c(
229   "alpha_0"      = "alpha[0]",
```

```
    "mu_0"         = "mu[0]",
    "sigma_dep_b"  = "sigma[dep]^{(b)}",
    "sigma_mpio_b" = "sigma[mpio]^{(b)}",
    "sigma_dep_g"  = "sigma[dep]^{(g)}",
    "sigma_mpio_g" = "sigma[mpio]^{(g)}",
    "sigma_y"      = "sigma[y]")

# Histogramas marginales de los parametros principales
# ------------------------------------------------------------------ #

{windows(width = 30, height = 20);
ggplot(df2_long, aes(x = value)) +
  geom_histogram(bins = 30, fill = "steelblue", color = "black", alpha = 0.6) +
  facet_wrap(~variable, scales = "free", labeller = as_labeller(pars_labels, label_parsed)) +
  theme_minimal(base_size = 12) +
  theme(
    panel.border     = element_rect(color = "gray20", fill = NA, size = 0.2),
    panel.spacing    = unit(0.2, "cm"),
    strip.text       = element_text(face = "bold", size = 12),
    axis.title.x     = element_text(size = 16, face = "bold"),  # tamaño y estilo eje X
    axis.title.y     = element_text(size = 16, face = "bold") )}

# Densidades marginales de los parametros principales
# ------------------------------------------------------------------ #

{windows(width = 30, height = 20);
ggplot(df2_long, aes(x = value)) +
  geom_density(fill = "steelblue", color = "black", alpha = 0.6) +
  facet_wrap(~variable, scales = "free", labeller = as_labeller(pars_labels, label_parsed)) +
  theme_minimal(base_size = 12) +
  theme(
    panel.border     = element_rect(color = "gray20", fill = NA, size = 0.2),
    panel.spacing    = unit(0.2, "cm"),
    strip.text       = element_text(face = "bold", size = 12))}

#  Traceplots de los  parametros principales
# ------------------------------------------------------------------ #

{windows(width = 30, height = 20);
ggplot(df2_long, aes(x = iter, y = value)) +
  geom_line(color = "steelblue", alpha = 0.8) +
  facet_wrap(~variable, scales = "free_y", labeller = as_labeller(pars_labels, label_parsed))
  ↪  +
  theme_minimal(base_size = 12) +
  theme(
    panel.border   = element_rect(color = "gray20", fill = NA, size = 0.2),
    panel.spacing  = unit(0.2, "cm"),
    strip.text     = element_text(face = "bold", size = 12))
  #+ xlab("Iteración") + ylab("Valor del parámetro")
  }

# :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::: #


# ================================================================== #
## 1.3. Intervalos de credibilidad ----
# ================================================================== #

?mcmc_intervals

# ------------------------------------------------------------------ #
# Departamentos - Parte Bernoulli
# ------------------------------------------------------------------ #
# Intervalos posteriores de los interceptos departamentales (_dep)

draws_alpha_dep <- fit_hurdle_final$draws("alpha_dep")  # draws de los interceptos
↪  departamentales de la parte binaria

# Grafico: Intervalos para los desviaciones estándar de los efectos aleatorios
#          a nivel de departamento

mcmc_intervals(draws_alpha_dep)
```

```
mcmc_intervals(draws_alpha_dep, pars = paste0("alpha_dep[", 1:10, "]"))

{windows(width = 30, height = 20);
mcmc_intervals(draws_alpha_dep #,  pars = paste0("alpha_dep[", 1:10, "]"),
               #prob       = 0.5,       # intervalo más oscuro          (Por defecto 0.5)
               #prob_outer = 0.9,       # intervalo más claro           (Por defecto 0.9)
               #point_est  = "median"   # Medidas de tendencia central (Por defecto median)
) + theme_minimal(base_size = 12) +
    geom_vline(xintercept = 0, color = "black", linetype = "solid", size = 0.7) +
  theme(panel.border  = element_rect(color = "gray20", fill = NA, size = 0.2),
    panel.spacing = unit(0.5, "cm"),
    axis.text.y   = element_text(face = "bold"),
    axis.title    = element_text(face = "bold"),
    plot.title    = element_text(face = "bold", size = 14)) +
  scale_color_manual(values = c("steelblue", "skyblue")) }

# ------------------------------------------------------------------ #
# Municipios - Parte Bernoulli
# ------------------------------------------------------------------ #
# Intervalos posteriores de los interceptos municipales (_mun)"

draws_alpha_mpio <- fit_hurdle_final$draws("alpha_mpio")

# Grafico: Intervalos para los desviaciones estándar de los efectos aleatorios
#          a nivel de municipios

mcmc_intervals(draws_alpha_mpio)  # Demasiados municipios
mcmc_intervals(draws_alpha_mpio, pars = paste0("alpha_mpio[", 1:20, "]"))

mcmc_intervals(draws_alpha_mpio,
               pars       = paste0("alpha_mpio[", 1:20, "]"),  # selecciona solo los primeros
               ↪  20
               prob       = 0.8,
               prob_outer = 0.95,
               point_est  = "median") +
  theme_minimal(base_size = 12) +
  theme(panel.border   = element_rect(color = "gray20", fill = NA, size = 0.2),
    panel.spacing  = unit(0.5, "cm"),
    axis.text.y    = element_text(face = "bold"),
    axis.title     = element_text(face = "bold"),
    plot.title     = element_text(face = "bold", size = 14)) +
  scale_color_manual(values = c("steelblue", "skyblue"))

# ------------------------------------------------------------------ #
# Departamentos - Parte continua (gaussiana)
# ------------------------------------------------------------------ #
# Intervalos posteriores de los interceptos departamentales

draws_mu_dep <- fit_hurdle_final$draws("mu_dep")  # Draws de interceptos departamentales de la
↪  parte continua

mcmc_intervals(draws_mu_dep)

{windows(width = 30, height = 20);
mcmc_intervals(draws_mu_dep,           # pars = paste0("alpha_dep[", 1:10, "]"),
               prob       = 0.5,       # intervalo más oscuro          (Por defecto 0.5)
               prob_outer = 0.9,       # intervalo más claro           (Por defecto 0.9)
               point_est  = "median"   # Medidas de tendencia central (Por defecto median)
) + theme_minimal(base_size = 12) +
  theme(panel.border  = element_rect(color = "gray20", fill = NA, size = 0.2),
        panel.spacing = unit(0.5, "cm"),
        axis.text.y   = element_text(face = "bold"),
        axis.title    = element_text(face = "bold"),
        plot.title    = element_text(face = "bold", size = 14)) +
  scale_color_manual(values = c("steelblue", "skyblue"))}

# ------------------------------------------------------------------ #
# Municipios - Parte continua (gaussiana)
# ------------------------------------------------------------------ #
# Intervalos posteriores de los interceptos municipales

draws_mu_mpio <- fit_hurdle_final$draws("mu_mpio")
```

```
mcmc_intervals(draws_mu_mpio, pars = paste0("mu_mpio[", 1:20, "]"))

# ================================================================ #
# Resumen de los parámetros principales definidos en pars_main
# ================================================================ #

fit_hurdle_final$summary(pars_main)[,
  c("variable","mean","sd","rhat","ess_bulk","ess_tail")]

# Extraer draws de todos los parámetros con todas las cadenas

draws_sigmas <- fit_hurdle_final$draws(c("sigma_dep_g","sigma_mpio_g","sigma_y"))
                mcmc_pairs(draws_sigmas, diag_fun = "dens", off_diag_fun = "hex")

# Desviaciones estándar (sigmas)
sigmas       <- c("sigma_dep_b", "sigma_mpio_b", "sigma_dep_g", "sigma_mpio_g", "sigma_y")
draws_sigmas <- fit_hurdle_final$draws(sigmas)
mcmc_pairs(draws_sigmas, diag_fun = "dens", off_diag_fun = "hex")
rm(sigmas, draws_sigmas)

# Interceptos globales

interceptos_globales <- c("alpha_0", "mu_0")
draws_global         <- fit_hurdle_final$draws(interceptos_globales)
mcmc_pairs(draws_global, diag_fun = "dens", off_diag_fun = "hex")
rm(interceptos_globales, draws_global)

# ================================================================ #
# ---------------------------------------------------------------- #
# BLOQUE 1 - Diagnostico Del Modelo Bayesiano (MCMC) ----
# ---------------------------------------------------------------- #
# ================================================================ #

library(cmdstanr) ; library(posterior)  ; library(bayesplot)

# 1. Resumen de las "MCMC": Rhat, ESS, cuantiles (1.2 del modelo hurled)

tabla_diagnostico <- fit_hurdle_final$summary(pars_main)[,

                      c("variable","mean","sd","q5","median","q95","rhat","ess_bulk","ess_tail")]
tabla_diagnostico <- `colnames<-`(cbind(tabla_diagnostico[,1], round(tabla_diagnostico[,-1],
  3)), colnames(tabla_diagnostico))
tabla_diagnostico

# 2. Extraer los "DRAWS"
# draws <- fit_hurdle_final$draws(pars_main) # Se hizo antes

# 3. Graficos diagnosticos (Ya estan anterioretne en el codigo)

# Traceplots
mcmc_trace(draws, pars = pars_main)
# Densidades posteriores
mcmc_dens(draws, pars = pars_main)
# Áreas posteriores (intervalos creíbles)
mcmc_areas(draws, pars = pars_main, prob = 0.90)

# ================================================================ #
# ---------------------------------------------------------------- #
# BLOQUE 2 - Estimación De Parametros Principales ----
# ---------------------------------------------------------------- #
# ================================================================ #

library(cmdstanr) ; library(posterior)  ; library(bayesplot)
library(dplyr)    ; library(ggplot2)

# ---------------------------------------------------------------- #
# 1. Tabla completa de los "Parametros Principales"
# ---------------------------------------------------------------- #

tabla_diagnostico

# ---------------------------------------------------------------- #
```

```
# 2. Forest Plot (intervalos creíbles)
# ------------------------------------------------------------- #

draws_main <- as_draws_df(fit_hurdle_final$draws(pars_main))

# Opcion 1  ---> CONFÍRMA QUE SE GENERA SIN ERROR
p_forest  <- mcmc_intervals(draws_main, pars = pars_main)

# Opcion 2
p_forest <- mcmc_intervals(draws_main, pars = pars_main,
                           prob       = 0.5,       # intervalo más oscuro           (Por defecto 0.5)
                           prob_outer = 0.9,       # intervalo más claro            (Por defecto 0.9)
                           point_est  = "median"   # Medidas de tendencia central (Por defecto median)
) +
  ggplot2::labs(title = "Intervalos creíbles de los parámetros principales") +
  theme_minimal(base_size = 12) +
  theme(
    panel.border   = element_rect(color = "gray20", fill = NA, size = 0.2),
    panel.spacing  = unit(0.5, "cm"),
    axis.text.y    = element_text(face = "bold"),
    axis.title     = element_text(face = "bold"),
    plot.title     = element_text(face = "bold", size = 14),
    legend.position = "none") +
  scale_color_manual(values = c("steelblue", "skyblue"))

{windows(width = 30, height = 20);p_forest}

# Opcion 3

plots_list_intervals <- lapply(pars_main, function(p){
  mcmc_intervals(draws, pars = p,
                       # prob = 0.5,
                       # prob_outer = 0.9,
                       # point_est = "median"
  ) + labs(title = p) + theme_minimal(base_size = 8) +
    theme(
      plot.title = element_textbox_simple(fill = "lightgray", color = "black", halign = 0.5, face = "bold", padding = margin(t = 2, r = 5, b = 2, l = 5), linewidth = 0.2, linetype = 1),
      panel.border = element_rect(color = "gray20", fill = NA, size = 0.2),
      axis.title.y = element_blank(),
      axis.text.y = element_blank(),
      legend.position = "none") +
    scale_color_manual(values = c("steelblue", "skyblue"))})

windows(width = 30, height = 20); wrap_plots(plots_list_intervals, ncol = 3)

# ------------------------------------------------------------- #
# 3. Densidades posteriores superpuestas
# ------------------------------------------------------------- #

p_dens <- mcmc_dens(draws_main, pars = pars_main) +
  ggplot2::labs(title = "Densidades posteriores") +
  theme_minimal(base_size = 12)

p_dens # ---> CONFÍRMA QUE SE GENERA SIN ERROR

# ------------------------------------------------------------- #
# 4. Areas creibles Ordenadas (opcional pero recomendable)
# ------------------------------------------------------------- #

p_areas <- mcmc_areas(draws_main, pars = pars_main, prob = 0.90) +
  ggplot2::labs(title = "Áreas creíbles al 90% de los parámetros") +
  theme_minimal(base_size = 12)

p_areas

# ============================================================= #
# ------------------------------------------------------------- #
```

```
# BLOQUE 3 - Efectos Aleatorios (Departamentales Y Municipales) ----
# ------------------------------------------------------------- #
# ============================================================= #

library(cmdstanr) ; library(posterior)  ; library(bayesplot)
library(dplyr)    ; library(ggplot2)

# ============================================================= #
# 1. Extraer efectos aleatorios POSTERIORES
# ------------------------------------------------------------- #
# Importante: Usa los nombres EXACTOS del modelo STAN
#             (mu_dep, mu_mpio) o (alpha_dep, alpha_mpio)
# ============================================================= #

# Parte Bernoulli (probabilidad de >0)
draws_dep_b  <- fit_hurdle_final$summary("alpha_dep")   # efectos departamentales
draws_mpio_b <- fit_hurdle_final$summary("alpha_mpio")  # efectos municipales

# Parte Lognormal (RMM >0)
draws_dep_g  <- fit_hurdle_final$summary("mu_dep")      # efectos departamentales
draws_mpio_g <- fit_hurdle_final$summary("mu_mpio")     # efectos municipales

# ============================================================= #
# 2. Construir tablas con medias e intervalos creíbles (departamentos)
# ============================================================= #

tabla_dep_b   <- as.data.frame(draws_dep_b) %>% mutate(across(where(is.numeric), ~ round(.,
↪  3)))
tabla_dep_g   <- as.data.frame(draws_mpio_b) %>% mutate(across(where(is.numeric), ~ round(.,
↪  3)))

tabla_dep_b[, c("variable","mean","sd","q5","median","q95")]
tabla_dep_g[, c("variable","mean","sd","q5","median","q95")]

# ============================================================= #
# 3. Construir tablas con medias e intervalos creíbles (municipios)
# ============================================================= #

tabla_mpio_b  <- as.data.frame(draws_dep_g) %>% mutate(across(where(is.numeric), ~ round(.,
↪  3)))
tabla_mpio_g  <- as.data.frame(draws_mpio_g) %>% mutate(across(where(is.numeric), ~ round(.,
↪  3)))

tabla_mpio_b[, c("variable","mean","sd","q5","median","q95")]
tabla_mpio_g[, c("variable","mean","sd","q5","median","q95")]

# ============================================================= #
# 4. Forest Plots Ordenados para "Efectos Aleatorios"
# ============================================================= #

# Convertir draws a formato para bayesplot:
# representa las muestras posteriores de los parámetros generadas por Stan.

  # Departamentos
  draws_dep_b_draws <- as_draws_df(fit_hurdle_final$draws("alpha_dep"))
  draws_dep_g_draws <- as_draws_df(fit_hurdle_final$draws("mu_dep"))

  # Municipios
  draws_mpio_b_draws <- as_draws_df(fit_hurdle_final$draws("alpha_mpio"))
  draws_mpio_g_draws <- as_draws_df(fit_hurdle_final$draws("mu_mpio"))

# Forest plot Departamentos (Bernoulli)
p_dep_b <- mcmc_intervals(draws_dep_b_draws, prob = 0.90) +
  labs(title = "Efectos aleatorios departamentales (Bernoulli)")

#Forest plot Departamentos (Lognormal)
p_dep_g <- mcmc_intervals(draws_dep_g_draws, prob = 0.90) +
  labs(title = "Efectos aleatorios departamentales (Lognormal)")

# Forest plot Municipios (Bernoulli)
p_mpio_b <- mcmc_intervals(draws_mpio_b_draws, prob = 0.90) +
  labs(title = "Efectos aleatorios municipales (Bernoulli)")
```

```
574
575 # Forest plot Municipios (Lognormal)
576 p_mpio_g <- mcmc_intervals(draws_mpio_g_draws, prob = 0.90) +
577   labs(title = "Efectos aleatorios municipales (Lognormal)")
578
579 # ================================================================ #
580 # 5. Ranking de Efectos aleatorios
581 # ================================================================ #
582
583 # ordena los departamentos/municipios de mayor a menor efecto Bernoulli, es decir,
584 # cuáles departamentos tienen mayor tendencia a que ocurra un valor positivo
585 # en la variable RMM según la parte binaria del modelo.
586
587 ranking_dep_b <- tabla_dep_b %>% arrange(desc(mean))
588 ranking_dep_g <- tabla_dep_g %>% arrange(desc(mean))
589
590 ranking_mpio_b <- tabla_mpio_b %>% arrange(desc(mean))
591 ranking_mpio_g <- tabla_mpio_g %>% arrange(desc(mean))
592
593 head(ranking_dep_b)
594 head(ranking_dep_g)
595 head(ranking_mpio_b)
596 head(ranking_mpio_g)
597
598 # ================================================================ #
599 # OBJETIVO 3 (2023): Identificar factores sociodemográficos asociados a la RMM
600 # Basado en efectos fijos (betas) del modelo hurdle multinivel en Stan
601 # ================================================================ #
602
603 library(posterior)
604 library(dplyr)
605 library(stringr)
606 library(tidyr)
607
608 # 1) Nombres de covariables en el mismo orden de X
609 X_names <- covs_names_s  # <-- este es el match correcto con beta_[k]
610
611 # 2) Extraer draws SOLO de betas (más rápido que extraer todo)
612 draws_betas <- as_draws_df(fit_hurdle_final$draws(c("beta_b", "beta_g")))
613
614 # 3) Pasar a formato largo y resumir posterior
615 regex_index <- "(?<=\\[)\\d+(?=\\])"  # escapamos los backslashes
616 betas_long <- draws_betas %>%
617   pivot_longer(
618     cols = everything(),
619     names_to = "param",
620     values_to = "value"
621   ) %>%
622   mutate(
623     comp = case_when(
624       str_detect(param, "^beta_b\\[") ~ "Bernoulli (logit): Pr(RMM>0)",
625       str_detect(param, "^beta_g\\[") ~ "Lognormal: log(RMM) | RMM>0",
626       TRUE ~ NA_character_
627     ),
628     k = as.integer(str_extract(param, regex_index)),
629     covariable = X_names[k]
630   )
631
632 tabla_betas <- betas_long %>%
633   group_by(comp, covariable, k) %>%
634   summarise(
635     mean   = mean(value),
636     sd     = sd(value),
637     q025   = quantile(value, 0.025),
638     median = median(value),
639     q975   = quantile(value, 0.975),
640     p_gt0  = mean(value > 0),
641     p_lt0  = mean(value < 0),
642     pd     = pmax(p_gt0, p_lt0),
643     sig95  = (q025 > 0) | (q975 < 0),
644     .groups = "drop"
645   ) %>%
```

```
  arrange(comp, desc(sig95), desc(pd), desc(abs(mean)))

# 4) Interpretación en escalas útiles
tabla_betas_interpret <- tabla_betas %>%
  mutate(
    OR_mean    = ifelse(str_detect(comp, "Bernoulli"), exp(mean), NA_real_),
    OR_q025    = ifelse(str_detect(comp, "Bernoulli"), exp(q025), NA_real_),
    OR_q975    = ifelse(str_detect(comp, "Bernoulli"), exp(q975), NA_real_),
    Factor_mean = ifelse(str_detect(comp, "Lognormal"), exp(mean), NA_real_),
    Factor_q025 = ifelse(str_detect(comp, "Lognormal"), exp(q025), NA_real_),
    Factor_q975 = ifelse(str_detect(comp, "Lognormal"), exp(q975), NA_real_))

print(tabla_betas_interpret, n = Inf, width = Inf)

# 5) Tablas separadas para el Objetivo 3

tabla_obj3_bernoulli <- tabla_betas_interpret %>%
  filter(str_detect(comp, "Bernoulli")) %>%
  select(covariable, mean, q025, q975, OR_mean, OR_q025, OR_q975, pd, sig95)
tabla_obj3_lognormal <- tabla_betas_interpret %>%
  filter(str_detect(comp, "Lognormal")) %>%
  select(covariable, mean, q025, q975, Factor_mean, Factor_q025, Factor_q975, pd, sig95)

print(tabla_obj3_bernoulli, n = Inf, width = Inf)
print(tabla_obj3_lognormal, n = Inf, width = Inf)
```

Bibliografía

Instituto Nacional de Salud (INS). (2022a). Informe de evento Mortalidad Materna (Código 551) [Consultado el 29 de mayo de 2025]. https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20MATERNA%20INFORME%20PRIMER%20SEMESTRE%202022.pdf

Departamento Administrativo Nacional de Estadística (DANE). (2021). Informe estadístico sociodemográfica aplicada (Número 9) [Consultado el 29 de mayo de 2025]. https://www.dane.gov.co/files/investigaciones/poblacion/informes-estadisticas-sociodemograficas/2021-12-20-mortalidad-materna-en-colombia-en-la-ultima-decada.pdf

Instituto Nacional de Salud (INS). (2022b). Informe de mortalidad materna, Colombia, 2022 [Consultado el 29 de mayo de 2025]. https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20MATERNA%20INFORME%202022.pdf#:~:text=El%20prop%C3%B3sito%20de%20este%20documento%20es

Min, Y. and Agresti, A. (2005). Modelos de efectos aleatorios para medidas repetidas de datos de conteo inflados a cero [Consultado el 29 de mayo de 2025]. https://journals.sagepub.com/doi/abs/10.1191/1471082x05st084oa

Jabessa, S., & Jabessa, D. (2021). Modelo bayesiano multinivel sobre mortalidad materna en Etiopía [Consultado el 29 de mayo de 2025].

Oliveira, I. V. G, et all. (2023). Mortalidad materna en el noreste de Brasil 2009-2019: distribución espacial, tendencia y factores asociados [Consultado el 29 de mayo de 2025]. https://www.scopus.com/pages/publications/85175530046

Instituto Nacional de Salud (INS). (2020). Informe del evento - Mortalidad Materna Colombia 2022 [Consultado el 29 de mayo de 2025]. https://www.ins.gov.co/buscador-eventos/Informesdeevento/MORTALIDAD%20MATERNA_2020.pdf

Marca País. (2025). ¿Cómo es la organización político-administrativa de Colombia? [Consultado el 3 de marzo de 2025]. https://colombia.co/pais-colombia/como-es-la-organizacion-politico-administrativa-de-colombia

Sandoval-Vargas, Y. G. and Eslava-Schmalbach, J. H. (2013). Inequidades en mortalidad materna por departamentos en Colombia para los años (2000-2001),(2005-2006) y (2008-2009) [Consultado el 29 de mayo de 2025]. https://www.scielosp.org/pdf/rsap/2013.v15n4/579-591/es

U.S. Census Bureau. (2022). Measuring maternal mortality: Select topics in international censuses [Consultado el 29 de mayo de 2025]. https://www.census.gov/content/dam/Census/programs-surveys/international-programs/stic/maternal-mortality-in-a-census-6.pdf

Varela-Ruiz, M., Díaz-Bravo, L., & García-Durán, R. (2012). Descripción y usos del método Delphi en investigaciones del área de la salud. *Investigación en Educación Médica*, *1*(2), 90-95. https://www.elsevier.es/es-revista-investigacion-educacion-medica-343-articulo-descripcion-usos-del-metodo-delphi-X2007505712427047

Hritcu, R. O. S. (2015). Multilevel models: Conceptual framework and applicability [Consultado el 29 de mayo de 2025]. https://journals.univ-danubius.ro/index.php/oeconomica/article/view/2959

Ghosal, S, et all. (2020). Un modelo jerárquico de obstáculos de efectos mixtos para datos de recuento espaciotemporal y su aplicación para identificar factores que afectan la escasez de profesionales sanitarios [Consultado el 29 de mayo de 2025]. https://academic.oup.com/jrsssc/article-abstract/69/5/1121/7058668

Berger, J. (2006). El caso del análisis bayesiano objetivo [Consultado el 29 de mayo de 2025]. https://projecteuclid.org/journals/bayesian-analysis/volume-1/issue-3/The-case-for-objective-Bayesian-analysis/10.1214/06-BA115.short

Gelman, A. (2013). Bayesian Data Analysis [Consultado el 29 de mayo de 2025]. https://sites.stat.columbia.edu/gelman/book/BDA3.pdf

Ros, D. P. (2022). Métodos Monte Carlo basados en cadenas de Markov [Consultado el 29 de mayo de 2025]. https://zaguan.unizar.es/record/125196/files/TAZ-TFG-2022-2905.pdf?version=1

Chen, Y.-C. (2019). Lecture 9: Hamiltonian Monte Carlo. https://faculty.washington.edu/yenchic/19A_stat535/Lec9_HMC.pdf

Carpenter, B, et all. (2017). Stan: A probabilistic programming language [Consultado el 29 de mayo de 2025]. https://www.jstatsoft.org/article/view/v076i01/0

Wang, X. and Yue, Y. R. and Faraway, J. J. (2018). Modelado de regresión bayesiana con INLA [Consultado el 29 de mayo de 2025]. https://www.taylorfrancis.com/books/mono/10.1201/9781351165761/bayesian-regression-modeling-inla-xiaofeng-wang-yu-ryan-yue-julian-faraway

Wierzchoslawska, J. (2023). Modelado espacial con INLA para el análisis de la atención desigual en Skåne [Consultado el 29 de mayo de 2025]. https://lup.lub.lu.se/student-papers/search/publication/9111119

Gómez-Rubio, Virgilio. (2020). Inferencia bayesiana con INLA [Consultado el 29 de mayo de 2025]. https://becarioprecario.bitbucket.io/inla-gitbook/#welcome

Vehtari, A, et all. (2021). https://projecteuclid.org/journals/bayesian-analysis/volume-16/issue-2/Rank-Normalization-Folding-and-Localization--An-Improved-R%cb%86-for/10.1214/20-BA1221.full

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Índices de efecto, existencia y significación en el marco bayesiano. *Fronteras en la psicología*, *10*, 2767. https://doi.org/10.3389/fpsyg.2019.02767

Lovric, M. M. (2025). Credible Intervals [Disponible en la página institucional del autor: https://sites.radford.edu/ mlovric/]. https://sites.radford.edu/~mlovric/Credible_Intervals.html

VanderPlas, J. (2014, junio). Frequentism and Bayesianism III: Confidence, Credibility, and Why Frequentism and Science Don't Mix [Parte III de una serie de cinco entradas sobre frecuentismo y bayesianismo]. https://jakevdp.github.io/blog/2014/06/12/frequentism-and-bayesianism-3-confidence-credibility/

Ramírez, M. L. (2016). Autocorrelación espacial: analogías y diferencias entre el Índice de Moran y el Índice Getis y Ord [Consultado el 29 de mayo de 2025]. https://repositorio.unne.edu.ar/handle/123456789/1694

Zeileis, A. and Kleiber, C. and Jackman, S. (2008). Modelos de regresión para datos de conteo en R [Consultado el 29 de mayo de 2025]. https://www.jstatsoft.org/article/view/v027i08/0

Montes Mora, C. L. (2024). Metodologías estadísticas para el análisis de la asociación entre dos estructuras de datos de tres vías [Tesis de maestría, Universidad del Valle] [Consultado el 29 de mayo de 2025]. https://shre.ink/od1y

Karimi, J., Cherono, A., Alegana, V., Mutua, M., Kiarie, H., Muthee, R., Temmerman, M., & Gichangi, P. (2025). Geographic inequalities and socio-demographic determinants of reproductive, maternal and child health at sub-national levels in Kenya. *BMC Public Health*, *25*, 1656. https://doi.org/10.1186/s12889-025-22583-w

Monnahan, C. C. (2024). Hacia buenas prácticas para evaluaciones bayesianas de poblaciones pesqueras ricas en datos utilizando un flujo de trabajo estadístico moderno [Consultado el 29 de mayo de 2025]. https://acortar.link/Fr41pw

Oliveira, I. V. G, et all. (2024). Mortalidad materna en Brasil: un análisis de las tendencias temporales y la agrupación espacial [Consultado el 29 de mayo de 2025]. https://www.scopus.com/pages/publications/85204512305

Lee, D. (2013). CARBayes: un paquete R para modelado espacial bayesiano con priors autoregresivos condicionales. *Revista de Software Estadístico*, *55*, 1-24. https://www.jstatsoft.org/article/view/v055i13/0

Gelman, A. (2006). Prior Distributions for Variance Parameters in Hierarchical Models [Comment on article by Browne and Draper. Source file: `Gelman 2006.pdf`]. *Bayesian Analysis*, *1*(3), 515-534. https://doi.org/10.1214/06-BA117A

Cortés, D. and Vargas, J. F. (2012). Inequidad regional en Colombia [Consultado el 29 de mayo de 2025]. https://repositorio.uniandes.edu.co/items/1ec1e038-0574-49e4-b842-1013eb311b99

Siabato, W. and Guzmán-Manrique, J. (2019). La autocorrelación espacial y el desarrollo de la geografía cuantitativa [Consultado el 29 de mayo de 2025]. http://www.scielo.org.co/scielo.php?pid=S0121-215X2019000100001&script=sci_arttext

Sánchez-Cantalejo, E. and Ocaña-Riola, R. (1999). Los modelos multinivel o la importancia de la jerarquía [Consultado el 29 de mayo de 2025]. https://www.sciencedirect.com/science/article/pii/S0213911199713907

Masconi, K. L., Matsha, T. E., Echouffo-Tcheugui, J. B., Erasmus, R. T., & Kengne, A. P. (2015). Reporte y manejo de datos faltantes en investigación predictiva para la diabetes mellitus tipo 2 prevalente no diagnosticada: una revisión sistemática. *EPMA Journal*, *6*(1), 7. https://link.springer.com/article/10.1186/s13167-015-0028-0

Centraal Bureau voor de Statistiek. (s.f.). Indexen en trends (TRIM). https://www.cbs.nl/nl-nl/maatschappij/natuur-en-milieu/indexen-en-trends--trim--/trim-frequently-asked-questions/which-proportion-of-missing-values-in-the-data-is-allowed-

Hilbe, J. M. (2011). *Negative Binomial Regression*. Cambridge University Press.

Filippi, V., Chou, D., Ronsmans, C., Graham, W., & Say, L. (2016). Levels and causes of maternal mortality and morbidity. *Reproductive Health*, *13*(2), S1-S9. https://europepmc.org/article/nbk/nbk361917