

UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

**Universidad Nacional de Colombia**

Modelos Multinivel Bayesianos: Análisis Profundo de los  
Resultados de la Prueba Saber 11 para la Configuración Nacional,  
Departamental y Municipal en Colombia

Autor:

Juan Daniel Diaz Alvarez  
juadiazal@unal.edu.co

Juan Sosa  
Ph.D Estadística

Noviembre del 2023

# Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Modelos Bayesianos y su estructura jerarquica</b>	<b>3</b>
<b>3. Relaciones entre la incidencia de la pobreza monetaria, la cobertura educativa y los puntajes de las pruebas Saber 11</b>	<b>6</b>
<b>4. Ajuste de los modelos</b>	<b>8</b>
4.1. Diagnostico y comparación de modelos . . . . .	9
4.2. Bondad de ajuste del modelo . . . . .	11
4.3. Ranking Bayesianos y frecuentistas . . . . .	11
4.4. Clasificación de los departamentos y municipios . . . . .	12
<b>5. Ajuste modelos predictivos</b>	<b>14</b>
<b>6. Apendices</b>	<b>17</b>
6.1. Distribuciones Condicionales Completas - DDC . . . . .	17
6.2. Diagnósticos de convergencia . . . . .	19
<b>7. Referencias</b>	<b>22</b>

# 1. Introducción

De acuerdo con la Guía de Usuario del Examen Saber 11, este examen, realizado semestralmente por el Icfes, tiene múltiples objetivos. En primer lugar, sirve como un criterio de admisión para estudiantes que desean ingresar a las Instituciones de Educación Superior. Además, su propósito incluye el monitoreo de la calidad de la educación ofrecida en instituciones de educación media y la generación de información relevante para la estimación del valor agregado de la educación superior (“Guía de Usuario examen Saber 11”, 2014).

Siguiendo los lineamientos de la Prueba Saber 11, este examen produce resultados a nivel individual de estudiantes que se encuentran en la fase final de su educación media. Estos resultados se expresan en puntajes obtenidos en cinco pruebas genéricas: Matemáticas, Lectura, Ciencias, Sociales e Inglés. Los puntajes están en una escala establecida en la segunda aplicación del año 2014, con un promedio de 50 y una desviación estándar de 10. Esta fijación de la media y la desviación estándar permite establecer una línea de base y proporciona un punto de referencia para las estimaciones. Además, se calcula un puntaje global, que se obtiene mediante un promedio ponderado de los puntajes en las cinco pruebas genéricas (“Documentación del examen Saber 11”, 2014).

Así, el puntaje global (PG) de la prueba Saber 11 está dado por:

$$PG = 5 \cdot \left( \frac{5 \cdot M + 3 \cdot L + 3 \cdot C + 3 \cdot S + 1 \cdot I}{13} \right)$$

donde M, L, C, S e I son los puntajes en las pruebas de Matemáticas, Lectura, Ciencias, Sociales, e Inglés, respectivamente. Por lo tanto, el puntaje global está diseñado de forma que asuma valores entre 0 puntos y 500 puntos, con una media de 250 puntos y una desviación estándar de 50 puntos.

El propósito de este estudio consiste en la configuración de modelos multinivel Bayesianos, empleando como conjunto de datos de entrenamiento el puntaje global de los estudiantes, para así modelar los resultados de las pruebas a nivel nacional, desglosados por Municipio y Departamento, con los siguientes objetivos:

- Establecer un ranking y una segmentación probabilística de los departamentos basados en su puntaje global promedio.
- Crear un ranking y una segmentación probabilística de los municipios en función de su puntaje global promedio.
- Desarrollar un modelo predictivo de la incidencia de la pobreza monetaria basado en el puntaje global promedio por departamento.

- Establecer un modelo predictivo de la cobertura neta de educación secundaria a partir del puntaje global promedio por municipio.

Para cumplir con los objetivos del estudio emplearemos los resultados de las pruebas Saber 11 desglosados municipio y departamento, acompañado de la incidencia de la pobreza monetaria y por último la cobertura neta secundaria. Por lo que excluirémos del estudio datos faltantes en la ubicación del colegio por Municipio, Departamento y Puntaje Global, igualmente se excluirán las estudiantes que no sean de nacionalidad colombiana y que no residen en Colombia, además se excluirá San Andrés y por último se usarán los registros que el Icfes haya catalogado como *Publicar*.

En la sección 2, se analizarán en detalle los modelos Bayesianos propuestos, así como su estructura jerárquica. En la sección 3, se investigará la relación entre la incidencia monetaria por departamento y los puntajes de las pruebas Saber 11, además de examinar la cobertura neta de la educación por municipio y los puntajes de las pruebas Saber 11 por municipio. En la sección 4, se procederá a ajustar los modelos Bayesianos y a establecer rankings Bayesianos basados en los puntajes obtenidos por departamento y municipio. En la sección 5, se llevará a cabo el ajuste de modelos predictivos de la incidencia de la pobreza monetaria y la cobertura neta en educación secundaria basados en los puntajes de las pruebas Saber 11. Y Finalmente, en la Sección 6 presentaremos resúmenes de cada uno de los modelos.

## 2. Modelos Bayesianos y su estructura jerárquica

Consideraremos 5 modelos Bayesianos basados en la distribución normal, dadas las características de la prueba Saber 11, con esto veremos el comportamiento desde el modelo más simple hasta el más complejo, ganando flexibilidad y extrapolación de datos, pero con un costo computacional alto. A continuación se especifican los modelos considerados:

### 1. Modelo Normal.

#### Distribución muestral:

$$y_{ij} \mid \theta, \sigma^2 \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2)$$

Para  $i = 1, \dots, n_j$  y  $j = 1, \dots, m$  donde  $y_{ij}$  es el puntaje global del estudiante  $i$  en el departamento  $j$ .

**Distribución previa:**

$$\theta \sim N(\mu_0, \tau_0^2) \quad \sigma^2 \sim \text{Gl} \left( \frac{v_0}{2}, \frac{v_0 \sigma_0^2}{2} \right)$$

Donde  $\mu_0, \tau_0^2, v_0, \sigma_0^2$  son los hiperparámetros del modelo.

## 2. Modelo Normal con medias específicas por departamento.

**Distribución muestral:**

$$y_{ij} \mid \theta_j, \sigma^2 \stackrel{\text{ind}}{\sim} N(\theta_j, \sigma^2)$$

**Distribución previa:**

$$\begin{aligned} \theta_j \mid \mu, \tau^2 &\stackrel{\text{iid}}{\sim} N(\mu, \tau^2), \quad \mu \sim N(\mu_0, \gamma_0^2), \quad \tau^2 \sim \text{Gl} \left( \frac{\eta_0}{2}, \frac{\eta_0 \tau_0^2}{2} \right) \\ \sigma^2 &\sim \text{Gl} \left( \frac{v_0}{2}, \frac{v_0 \sigma_0^2}{2} \right) \end{aligned}$$

Donde  $\mu_0, \gamma_0^2, \tau_0^2, v_0, \sigma_0^2$  son los hiperparámetros del modelo.

## 3. Modelo Normal con medias y varianzas específicas por departamento.

**Distribución muestral:**

$$y_{ij} \mid \theta_j, \sigma_j^2 \stackrel{\text{ind}}{\sim} N(\theta_j, \sigma_j^2)$$

**Distribución previa:**

$$\begin{aligned} \theta_j \mid \mu, \tau^2 &\stackrel{\text{iid}}{\sim} N(\mu, \tau^2), \quad \mu \sim N(\mu_0, \gamma_0^2), \quad \tau^2 \sim \text{Gl} \left( \frac{\eta_0}{2}, \frac{\eta_0 \tau_0^2}{2} \right), \\ \sigma_j^2 \mid v, \sigma^2 &\stackrel{\text{iid}}{\sim} \text{Gl} \left( \frac{v}{2}, \frac{v \sigma^2}{2} \right), \quad v = \text{Constante}, \quad \sigma^2 \sim \text{G} \left( \frac{\alpha_0}{2}, \frac{\beta_0}{2} \right) \end{aligned}$$

Donde  $\mu_0, \gamma_0^2, \eta_0, \tau_0^2, v, \alpha_0, \beta_0$  son los hiperparámetros del modelo.

## 4. Modelo Normal con medias específicas por municipio y departamento.

**Distribución muestral:**

$$y_{ijk} \mid \zeta_{jk}, \kappa^2 \stackrel{\text{ind}}{\sim} N(\zeta_{jk}, \kappa^2)$$

Para  $i = 1, \dots, n_{jk}$ ,  $j = 1, \dots, n_k$ ,  $k = 1, \dots, m$  donde  $y_{ijk}$  es el puntaje global del estudiante  $i$  en el municipio  $j$  del departamento  $k$ .

**Distribución previa:**

$$\begin{aligned}\zeta_{jk} | \theta_k, \sigma^2 &\stackrel{\text{ind}}{\sim} N(\theta_k, \sigma^2), & \kappa^2 &\sim \text{Gl}\left(\frac{\xi_0}{2}, \frac{\xi_0 \kappa_0^2}{2}\right) \\ \theta_k | \mu, \tau^2 &\stackrel{\text{iid}}{\sim} N(\mu, \tau^2), & \mu &\sim N(\mu_0, \gamma_0^2), & \tau^2 &\sim \text{Gl}\left(\frac{\eta_0}{2}, \frac{\eta_0 \tau_0^2}{2}\right) \\ \sigma^2 &\sim \text{Gl}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)\end{aligned}$$

Donde  $\xi_0, \kappa_0^2, \mu_0, \gamma_0^2, \eta_0, \tau_0^2, \nu_0, \sigma_0^2$  son los hiperparámetros del modelo.

**5. Modelo Normal con medias específicas por municipio y departamento.**

**Distribución muestral:**

$$y_{ijk} | \zeta_{jk}, \kappa^2 \stackrel{\text{ind}}{\sim} N(\zeta_{jk}, \kappa^2)$$

**Distribución previa:**

$$\begin{aligned}\zeta_{jk} | \theta_k, \sigma_k^2 &\stackrel{\text{ind}}{\sim} N(\theta_k, \sigma_k^2), & \kappa^2 &\sim \text{Gl}\left(\frac{\xi_0}{2}, \frac{\xi_0 \kappa_0^2}{2}\right) \\ \theta_k | \mu, \tau^2 &\stackrel{\text{iid}}{\sim} N(\mu, \tau^2), & \mu &\sim N(\mu_0, \gamma_0^2), & \tau^2 &\sim \text{Gl}\left(\frac{\eta_0}{2}, \frac{\eta_0 \tau_0^2}{2}\right) \\ \sigma_k^2 | \nu, \sigma^2 &\sim \text{Gl}\left(\frac{\nu}{2}, \frac{\nu \sigma^2}{2}\right), & \nu &= \text{Constante}, & \sigma^2 &\sim \text{G}\left(\frac{\alpha_0}{2}, \frac{\beta_0}{2}\right)\end{aligned}$$

Donde  $\xi_0, \kappa_0^2, \mu_0, \gamma_0^2, \eta_0, \nu, \alpha_0, \beta_0$  son los hiperparámetros del modelo.

En la Figura 1, se presentan visualmente los Grafos Acíclicos Dirigidos (DAG, por sus siglas en inglés) correspondientes a los modelos previamente discutidos, con la excepción del Modelo 1, conocido como el 'Modelo Normal', el cual, debido a su simplicidad y características claramente establecidas, no se incluye en esta representación gráfica. Este conjunto de representaciones gráficas de los modelos forma la jerarquía de modelos, en la que en la última jerarquía se encuentran los hiperparámetros, en los que se logra una representación visual de cómo influyen en la estructura de los modelos los cuales nos permitirán darle vida a los modelos, y estos serán ajustados de acuerdo a la información que se obtiene de las pruebas Saber 11.

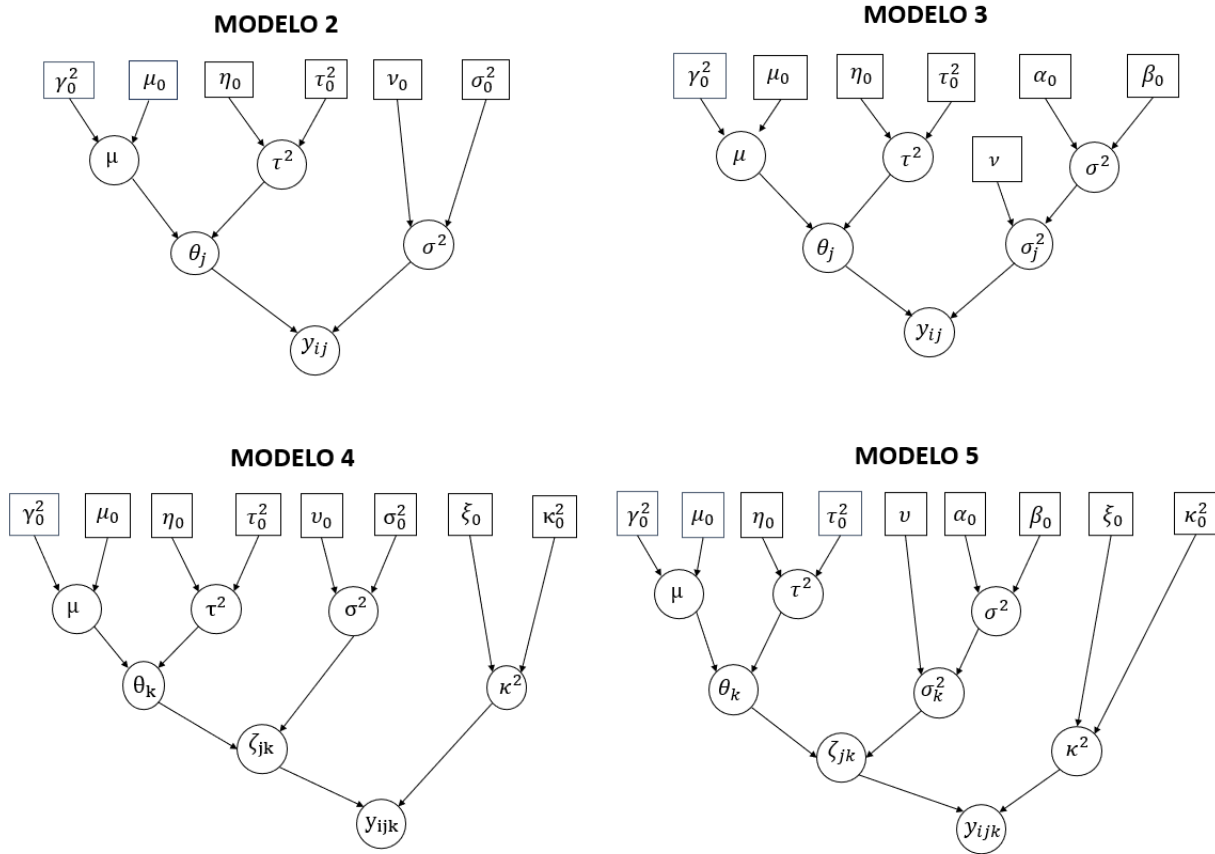


Figura 1: DAGs

Cada modelo será ajustado usando el muestreador de Gibbs, con un total de 101000 iteraciones, donde las primeras 1000 iteraciones del muestreador serán el período de calentamiento de la cadena. Una vez culminado este período, se realizará un muestreo sistemático con una amplitud de 10 y así completar un total de 10000 iteraciones para las inferencias posteriores.

### 3. Relaciones entre la incidencia de la pobreza monetaria, la cobertura educativa y los puntajes de las pruebas Saber 11

Las estadísticas de la incidencia de la pobreza monetaria se define como el porcentaje de las personas que son catalogadas como pobres. Teniendo como umbral el límite de la pobreza (LP) para determinar la caracterización. (Consejo Nacional de Política Económica y Social, 2012). Por otro lado el umbral, o bien llamado línea de pobreza, la CEPAL (como se citó en “Pobreza monetaria en Colombia: Resultados 2020”) establece que: “La línea de pobreza representa un valor monetario en el cual se consideran dos componentes: el costo de adquirir una canasta básica de alimentos y

el costo de los demás bienes y servicios, expresado sobre la base de la relación entre el gasto total y el gasto en alimentos". Sobre esta base se construyen las estadísticas de la pobreza en Colombia. Por lo tanto, en la Figura 2 se muestra los puntajes medios de las pruebas Saber 11 y la incidencia de la pobreza monetaria para el año 2018 por departamento. Hay que aclarar que las estadísticas de la incidencia de la pobreza monetaria solo se obtuvieron para 23 departamentos.

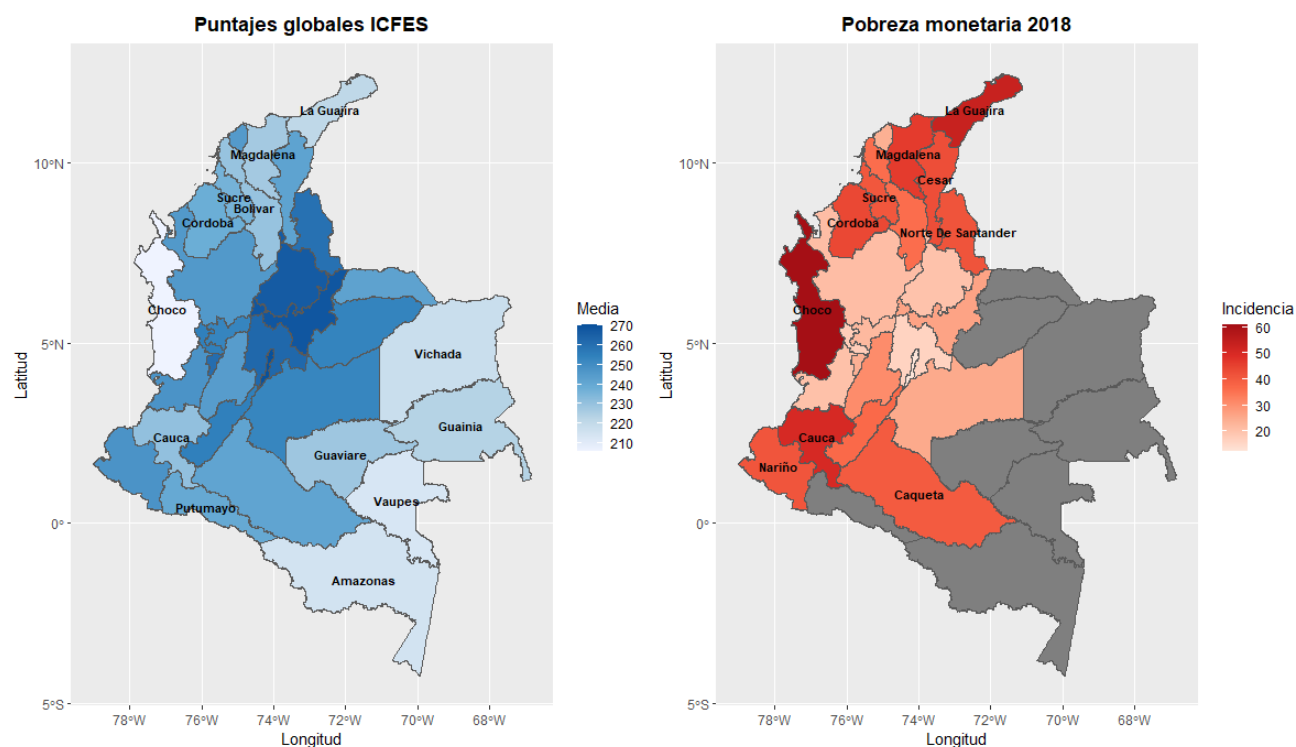


Figura 2: Pobreza monetaria 2018 y puntajes Saber 11 2022-2 por departamento

En la Figura 2 se destacan los departamentos donde el puntaje medio de la prueba Saber 11 fue menor a 240 puntos, mientras que para la pobreza monetaria se destaca los departamentos que tiene una incidencia mayor al 40%. De manera particular se puede destacar el Departamento del Chocó, quien tuvo el puntaje medio en la prueba más bajo de todo del todo país y con la tasa de pobreza monetaria más alta. Lo mismo parece suceder con el departamento de La Guajira. Esto podría sugerir que las dificultades económicas pueden estar afectando la calidad de la educación o el rendimiento académico de los estudiantes en estos departamentos.

Para la Figura 3 se muestran los puntajes medios obtenidos en la prueba Saber 11 y la Cobertura Neta Secundaria (CNS) por municipio. Este último, es un indicador que mide el acceso y la permanencia de la población en edad escolar al sistema educativo y se calcula dividiendo el número



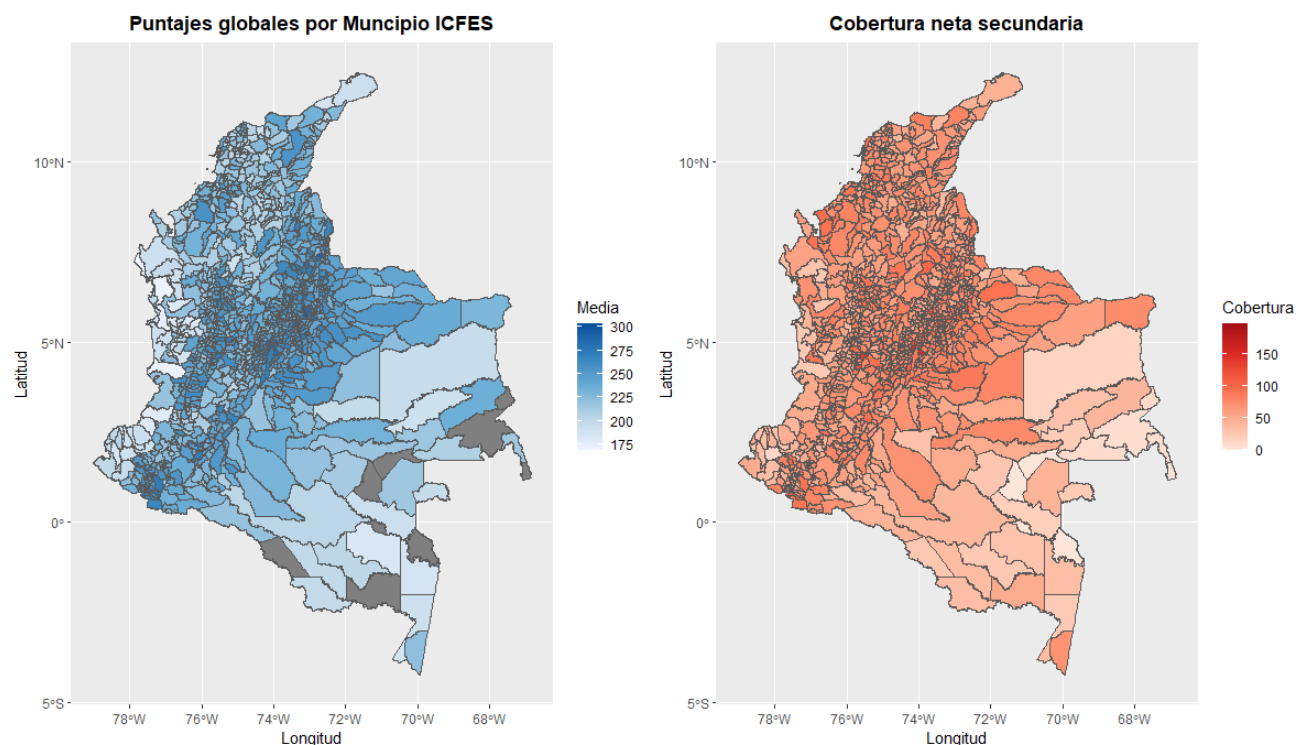


Figura 3: Cobertura neta secundaria y puntajes Saber 11 2022-2 por municipio

de estudiantes matriculados en secundaria que tienen la edad teórica para cursarlo por el total de la población correspondiente a esa misma edad (Ministerio de Educación Nacional, [2023](#)).

El mapa sugiere una variación en las puntuaciones medias de los municipios como en la tasa de cobertura. Algunos municipios tienen puntuaciones altas y una alta tasa cobertura, lo que podría indicar un alto nivel de acceso y permanencia en el sistema educativo. Sin embargo, otros municipios tienen puntuaciones medias más bajas y una menor tasa de cobertura, lo que podría indicar desafíos en términos de acceso y permanencia en el sistema educativo. Lo que podría indicar que la tasa de cobertura neta secundaria esté afectando los resultados de los estudiantes en estos municipios.

## 4. Ajuste de los modelos

Como se mencionó en la Sección 2 se implementará el muestreador de Gibbs. Para tal fin, usaremos previas difusas, atendiendo además, a la información que proporciona la documentación

de la prueba Saber 11. Así los hiperparámetros de cada modelo, estarán dispuestos a continuación:

- $\mathbf{M}_1 : \mu_0 = 250, \gamma_0^2 = 50^2, v_0 = 1, \sigma_0^2 = 50^2.$
- $\mathbf{M}_2 : \mu_0 = 250, \gamma_0^2 = 50^2, \eta_0 = 1, \tau_0^2 = 50^2, v_0^2 = 1, \sigma_0^2 = 50^2.$
- $\mathbf{M}_3 : \mu_0 = 250, \gamma_0^2 = 50^2, \eta_0 = 1, \tau_0^2 = 50^2, v = 1, \alpha_0 = 1, \beta_0^2 = 1/50^2.$
- $\mathbf{M}_4 : \xi_0 = 1, \kappa_0^2 = 50^2, \mu_0 = 250, \gamma_0^2 = 50^2, \eta_0 = 1, \tau_0^2 = 50^2, v_0^2 = 1, \sigma_0^2 = 50^2.$
- $\mathbf{M}_5 : \xi_0 = 1, \kappa_0^2 = 50^2, \mu_0 = 250, \gamma_0^2 = 50^2, \eta_0 = 1, \tau_0^2 = 50^2, v = 1, \alpha_0 = 1, \beta_0^2 = 1/50^2.$

#### 4.1. Diagnostico y comparación de modelos

Dado que estamos usando el muestreo de Gibbs, es necesaria la validación de la convergencia. Por lo tanto, a continuación se presenta la log verosimilitud en los Modelos 2, 3, 4 y 5 en función del número de iteraciones finales.

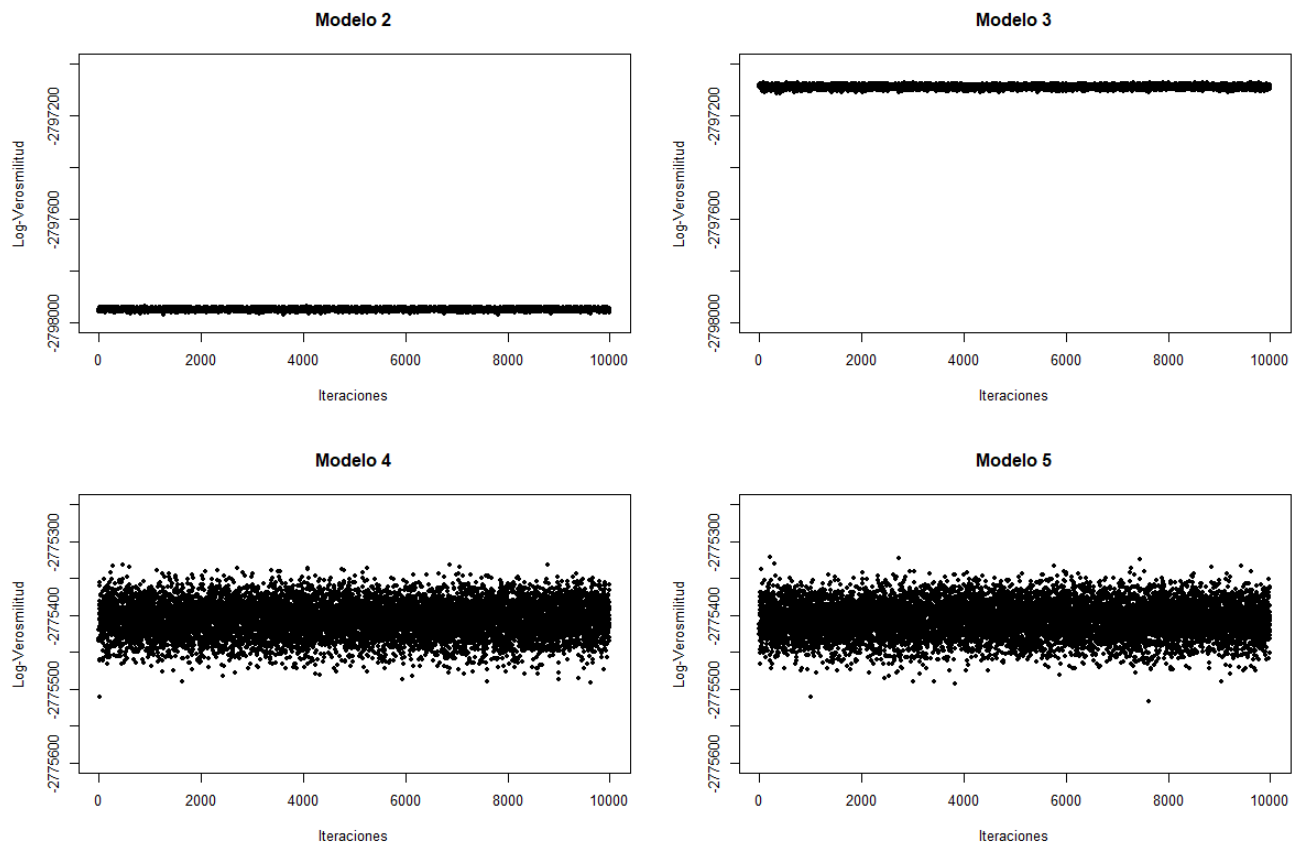


Figura 4: Convergencias del muestreador

La Figura 4 muestra claramente que los modelos 3, 4 y 5 superan al modelo 2 en log-verosimilitud y los modelos 4 y 5 son mayores que el 3 en log-verosimilitud. Aunque los modelos 4 y 5 son similares, ya que no varían sustancialmente, ambos son los más informativos. Además, la convergencia de los muestreadores se evidencia por la ausencia de patrones en las gráficas.

Por otra parte, haciendo uso del Criterio de Información de la Devianza (DIC, por sus siglas en inglés) y el Criterio de Información de Watanabe - Akaike (WAIC, por sus siglas en inglés), se comparan los 5 modelos obteniendo los siguientes resultados:

Modelo	DIC	WAIC
1	5636785	5636785
2	5595932	5595930
3	5594240	5594580
4	5551814	5587380
5	5551802	5586841

Tabla 1: Criterios de información

Con la información de la Tabla 1 se puede apreciar que el Modelo 5 sobresale en comparación con los otros modelos atendiendo al menor DIC y WAIC. No obstante, es importante notar que las diferencias entre el Modelo 5 y el Modelo 4 en relación a los criterios de información, no son lo suficientemente sustanciales para elegir el modelo 5 sobre el modelo 4, dada la complejidad del modelo 5.

A continuación se comparan las estadísticas posteriores basadas en  $\mu$  para cada modelo, de manera tabular, en la que se presentan para los 4 últimos modelos.

Modelo	Media	CV %	IC Inf	IC Sup
1	250.27	0.028	250.13	250.41
2	242.29	1.408	235.56	248.92
3	242.26	1.404	235.65	249.01
4	232.51	1.472	225.79	239.28
5	232.47	1.484	225.63	239.22

Tabla 2: Estadísticas posterior para  $\mu$

Con la información de la Tabla 2 es evidente que los modelos 1, 2 y 3 tienen medias similares, mientras que los modelos 4 y 5 tienen medias más bajas en comparación a los tres primeros. Al igual que con los modelos 2 y 3, el CV % es más alto y los IC son más amplios que para el modelo 1, indicando una mayor variabilidad e incertidumbre en las predicciones.

## 4.2. Bondad de ajuste del modelo

A continuación valideramos la bondad de ajuste del modelo por municipios usando el modelo 5, para esto emplearemos como estadísticos de prueba los siguientes: media, mediana, desviación estandar - DE, rango intercuatílico - RIC, mínimo - Min y máximo - Max. Se calcularon los valores p predictivo posterior - PPP para cada estadística de prueba en cada municipio y se obtuvieron los siguientes resultados:

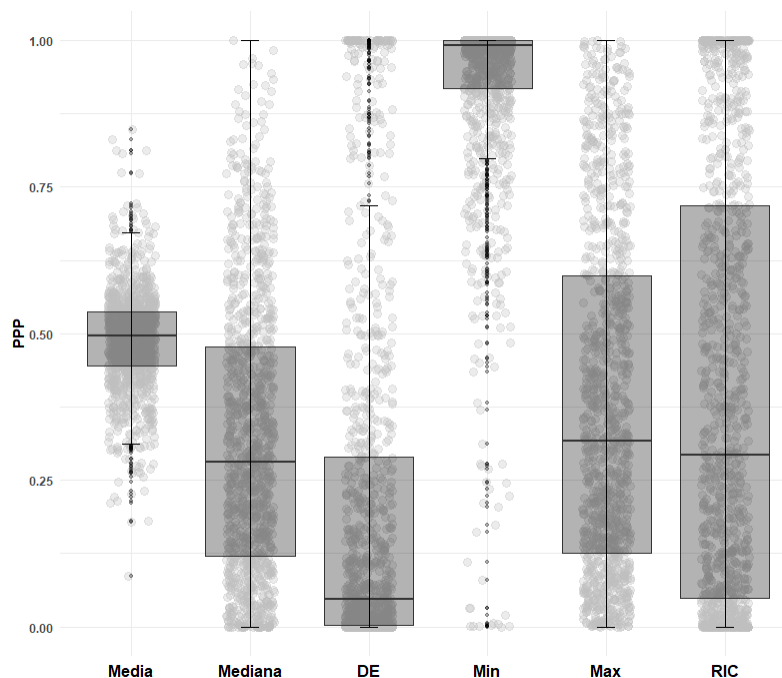


Figura 5: PPP por estadística en cada municipio

Las medianas de los PPP de la media, la mediana, el rango intercuatílico y el máximo, están cercanos al valor a 0.5, lo que induce a que el modelo los está estimando correctamente. Esto no parece suceder con el desvío estandar y el mínimo, ya que parece que el modelo está subestimando la desvío estandar y sobreestimando el mínimo, dado el comportamiento de los PPP para estas estadísticas.

## 4.3. Ranking Bayesianos y frecuentistas

Teniendo en cuenta el modelo 5, se estableció un ranking con los puntajes medios de cada departamento del país, con sus respectivos intervalos de credibilidad y se comparó con un ranking frecuentista basados en la media muestral con sus intervalos de confianza. Por lo que se obtuvieron las siguiente clasificación:

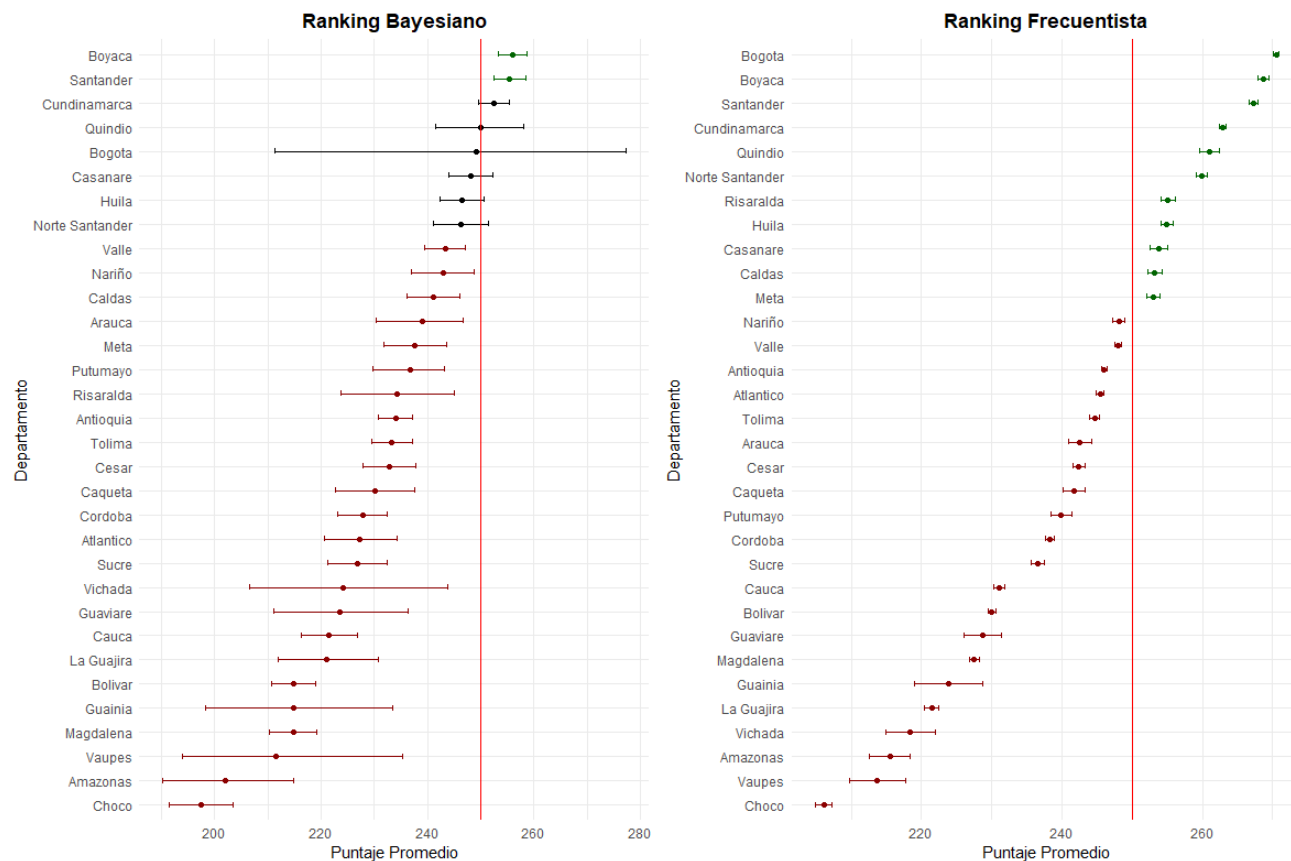


Figura 6: Comparación entre el ranking Bayesiano y frecuentista

En la Figura 6 es notorio que la incertidumbre en el modelo Bayesiano es mayor que para el modelo frecuentista, debido a la amplitud de los intervalos de credibilidad. Por otro lado, el ranking ajustado con el modelo Bayesiano difiere con el frecuentista, como es el caso de Bogotá, el cual pasa de primer lugar a estar en quinto lugar, de igual manera sucede con otros departamentos.

#### 4.4. Clasificación de los departamentos y municipios

En este apartado realizaremos clasificación de los departamentos y municipios en 5 y 8 grupos, respectivamente. Para llevar a cabo esta clasificación usaremos la técnica de *kmeans*, la cual se realizará en cada iteración del muestreador para la construcción de la matriz de incidencia y con la media posterior para la clasificación final. Por lo tanto se presentan las clasificación de los municipios y los departamento en Figura 7 y la matriz de incidencia en Figura 8 respectivamente.

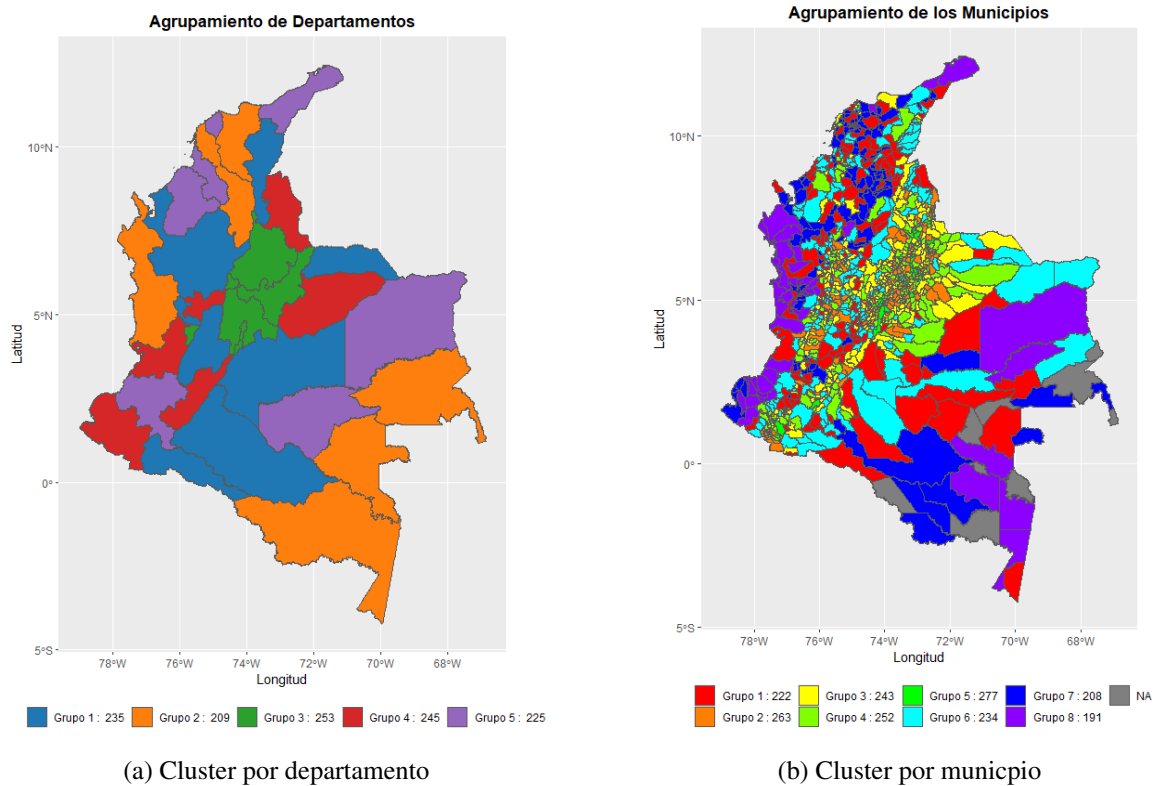
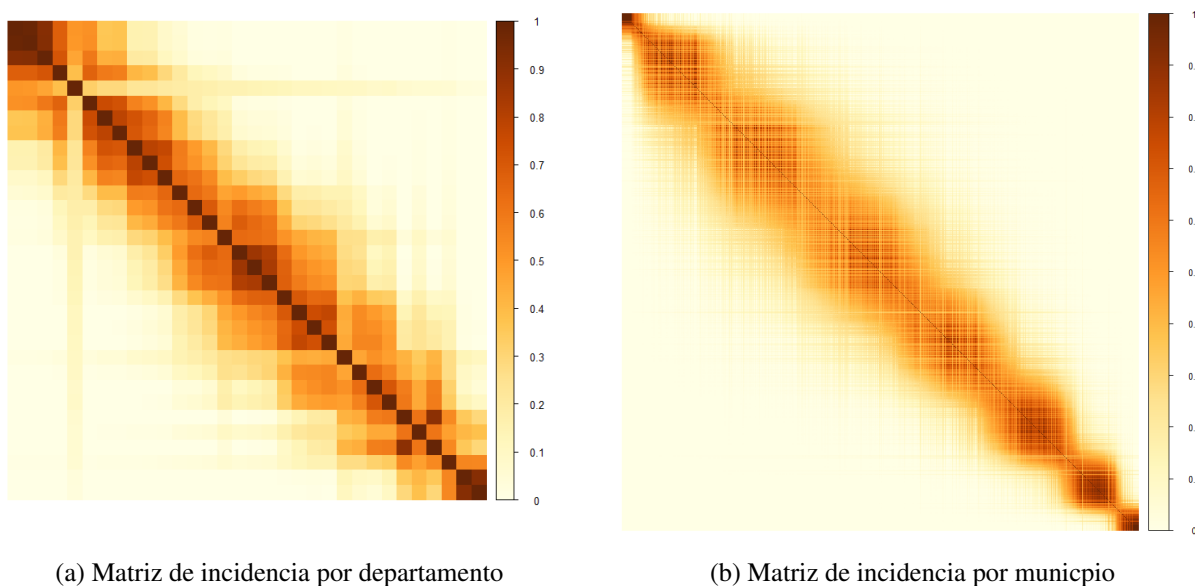
Figura 7: Agrupamiento usando *kmeans*

Figura 8: Matriz de incidencia

Los mapas de la Figura 7 sugieren que en el centro del país se concentran los puntajes de la Prueba Saber 11 más altos, tanto para los departamentos como para los municipios. Por otro lado,

teniendo en cuenta la Figura 8, es evidente que la matriz de departamentos en comparación con la matriz por municipios muestra una menor probabilidad (colores más claros) de que los municipios pertenezcan al mismo grupo. Esto podría indicar que hay más variación en los puntajes de las pruebas Saber 11 a nivel de municipio que a nivel de departamento.

## 5. Ajuste modelos predictivos

En la sección Sección 3, hemos presentado de manera visual la relación existente entre los puntajes de las pruebas Saber 11 por departamento y municipios con la incidencia de la pobreza monetaria y la CNS, respectivamente. Sin embargo, al observar detenidamente la Figura 9, esta relación se aclara aún más atendiendo a que los puntajes de las pruebas Saber 11 tienen una relación inversamente proporcional con la incidencia de la pobreza monetaria, mientras que la relación entre los puntajes y la CNS es directamente proporcional.

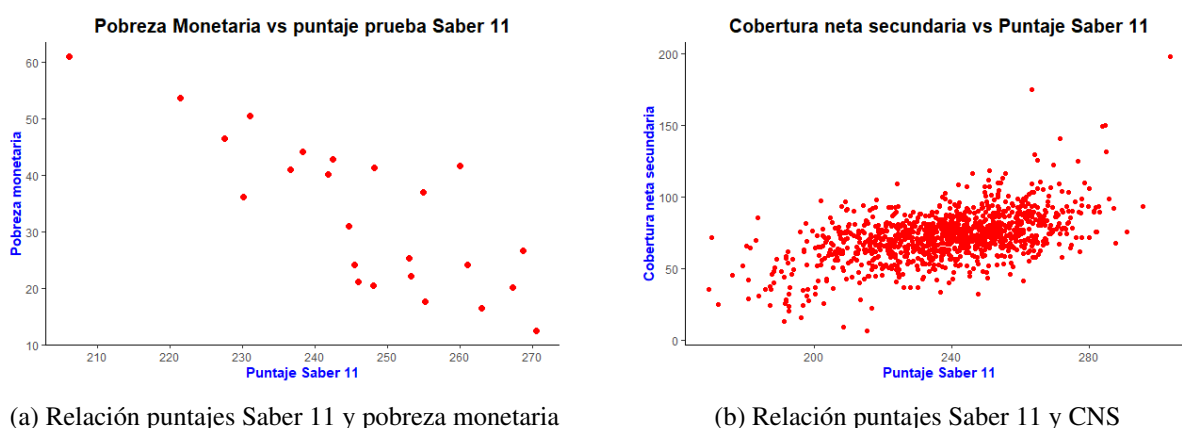


Figura 9: Puntajes Saber 11 vs Estadísticas DANE

Teniendo en cuenta que el DANE para el año 2018 no registró información para la Incidencia de la Pobreza Monetaria (IPM) para 8 de los 32 departamentos y para el año 2022 no se registró la tasa de CNS para 2 municipios. Se propone estimar la incidencia de la pobreza monetaria para los departamentos que no fueron medidos, al igual que la tasa de CNS para los municipios. Esto ajustando un modelo de regresión lineal simple usando como variables predictoras los puntajes de las pruebas Saber 11 de los departamentos y los municipios. Ajustado los modelos para cada estadística del DANE se obtuvieron los siguientes resultados:

Departamento	Media	IC Inf	IC Sup
Amazonas	52.730	43.821	60.789
Vaupes	47.088	33.087	58.200
Guainia	45.098	34.023	55.590
Guaviare	40.007	32.286	47.581
Vichada	39.674	28.003	50.367
Putumayo	32.218	28.226	36.305
Arauca	30.830	26.150	36.018
Casanare	25.427	22.524	28.339

Municipio	Media	IC Inf	IC Sup
Belén de Bajirá	57.25	54.065	60.414
Mapiripana	55.25	46.249	63.666

Tabla 4: Estimación de CNS

Tabla 3: Estimación de la incidencia de la pobreza monetaria

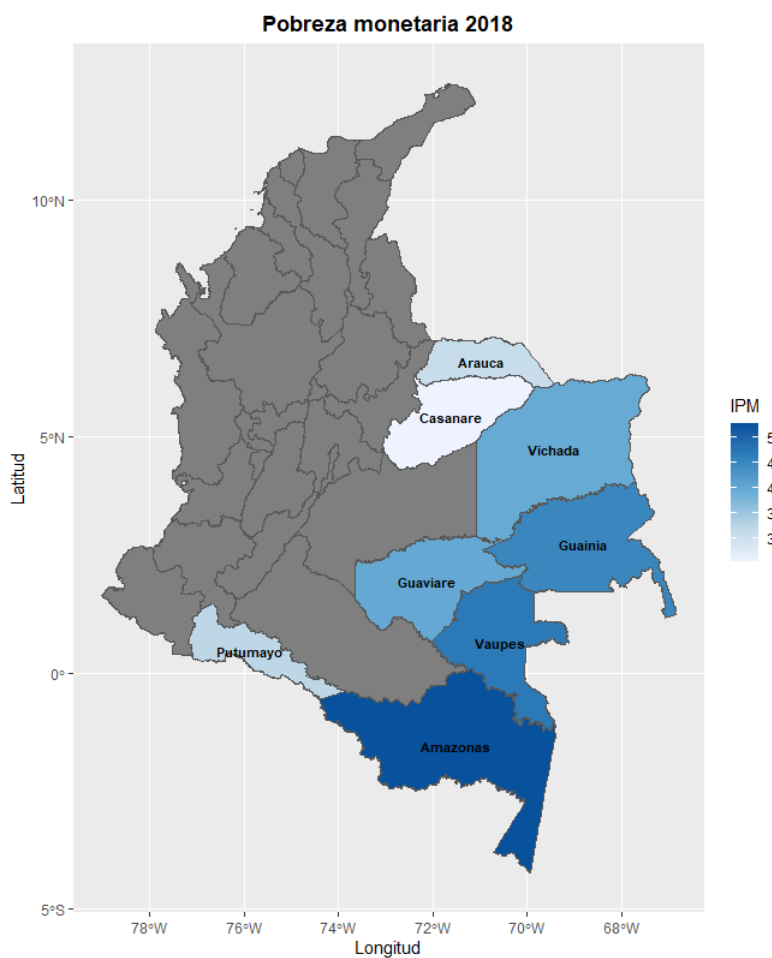


Figura 10: Estimación de IPM



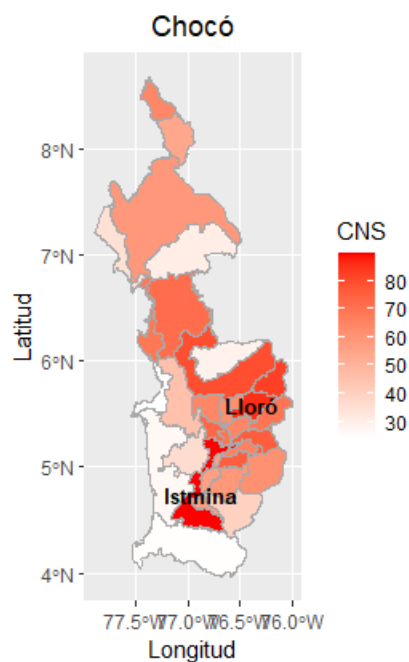


Figura 11: Estimación de Mapiripana

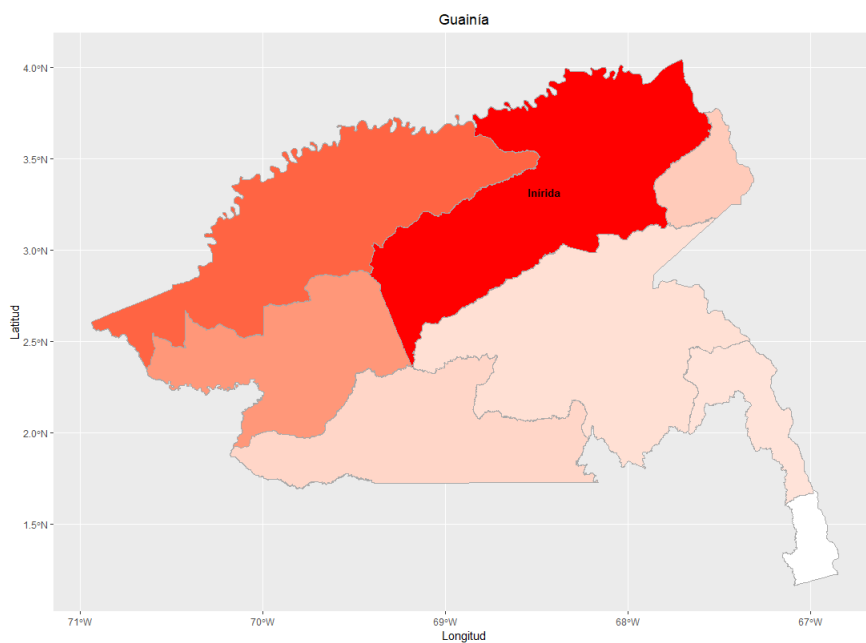


Figura 12: Estimación de Belén de Bajirá

Teniendo en cuenta la información de la Tabla 4 encontramos a *Belén de Bajirá*, un corregimiento del municipio de Barrancominas situado en el departamento de Guainía y *Mapiripana* un municipio del departamento del Chocó. Para estos no se obtuvo información georeferenciada del DANE lo que dificultó la generación del mapa de estos con las estimaciones,

Sin embargo, se destaca que la CNS de estos está por encima del 50%. Por lo contrario, en la Tabla 3 se observa las estimaciones de la incidencia de la pobreza monetaria, donde se destacan que la mitad de estos, tienen una incidencia monetaria mayor a 40%, los cuales también en la Figura 2 tuvieron puntajes bajos en la prueba Saber 11.

## 6. Apendices

### 6.1. Distribuciones Condicionales Completas - DDC

- **Modelo 1.** Para este modelo, se cuenta con el siguiente espacio de parámetros,  $\Phi = (\theta, \sigma^2)$ . Por lo tanto, las DCC están dadas por:

$$\theta | - \sim N\left(\frac{\mu_0/\tau_0^2 + n\bar{y}/\sigma^2}{1/\tau^2 + n/\sigma^2}, \frac{1}{1/\tau^2 + n/\sigma^2}\right)$$

$$\sigma^2 | - \sim \text{Gl}\left(\frac{v_0 + n}{2}, \frac{v_0\sigma_0^2 + \sum_{i=1}^n (y_i - \theta)^2}{1/\tau^2 + n/\sigma^2}\right)$$

- **Modelo 2.** El espacio de parámetros está definido por  $\Phi = (\Theta, \sigma^2, \mu, \tau^2)$ . Donde  $\Theta = (\theta_1, \dots, \theta_m)$ .

$$\theta_j | - \sim N\left(\frac{\mu/\tau^2 + n_j\bar{y}_j/\sigma^2}{1/\tau^2 + n_j/\sigma^2}, \frac{1}{1/\tau^2 + n_j/\sigma^2}\right)$$

$$\sigma^2 | - \sim \text{Gl}\left(\frac{v_0 + \sum_{j=1}^m n_j}{2}, \frac{v_0\sigma_0^2 + \sum_{j=1}^m \sum_{i=1}^{n_j} (y_{i,j} - \theta_j)^2}{2}\right)$$

$$\mu | - \sim N\left(\frac{\mu_0/\gamma_0^2 + m\bar{\theta}/\tau^2}{1/\gamma_0^2 + m/\tau^2}, \frac{1}{1/\gamma_0^2 + m/\tau^2}\right)$$

$$\tau^2 | - \sim \text{Gl}\left(\frac{\eta_0 + m}{2}, \frac{\eta_0\tau_0^2 + \sum_{j=1}^m (\theta_j - \mu)^2}{2}\right)$$

- **Modelo 3.** Tenemos el siguiente espacio de parámetros:  $\Phi = (\Theta, \Sigma, \mu, \tau^2, \sigma^2)$ . Teniendo en cuenta que  $\Sigma = (\sigma_1^2, \dots, \sigma_m^2)$ ,  $\Theta = (\theta_1, \dots, \theta_m)$ . Las DCC de  $\mu$  y  $\tau^2$  se calcularon

previamente, por lo tanto:

$$\theta_j|-\sim \text{N}\left(\frac{n_j\bar{y}_j+\mu/\tau^2}{n_j/\sigma_j^2+1/\tau^2}, \frac{1}{n_j/\sigma_j^2+1/\tau^2}\right)$$

$$\sigma_j^2|-\sim \text{GI}\left(\frac{\nu+n_j}{2}, \frac{\nu\sigma^2+\sum_{i=1}^{n_j}(y_{i,j}-\theta_j)^2}{2}\right)$$

$$\sigma^2|-\sim \text{G}\left(\frac{\alpha_0+\nu m}{2}, \frac{\beta_0}{2}+\frac{\nu}{2}\sum_{j=1}^m\frac{1}{\sigma_j^2}\right)$$

- **Modelo 4.** Como espacios de parámetros tenemos que:  $\Phi = (\zeta, \Theta, \kappa^2, \mu, \tau^2, \sigma^2)$ . Donde  $\zeta = (\zeta_{1,1}, \dots, \zeta_{n_k,m})$ ,  $\Theta = (\theta_1, \dots, \theta_m)$ . Las DCC de  $\mu, \tau^2$  se calcularon previamente, por lo tanto:

$$\theta_k|-\sim \text{N}\left(\frac{n_k\bar{\zeta}_k/\sigma_k^2+\mu/\tau^2}{n_k/\sigma_k^2+1/\tau^2}, \frac{1}{n_k/\sigma_k^2+1/\tau^2}\right)$$

$$\zeta_{j,k}|-\sim \text{N}\left(\frac{n_{jk}\bar{y}_{jk}/\kappa^2+\theta_k/\sigma_k^2}{n_{jk}/\kappa^2+1/\sigma_k^2}, \frac{1}{n_{jk}/\kappa^2+1/\sigma_k^2}\right)$$

$$\sigma^2|-\sim \text{GI}\left(\frac{\nu_0+\sum_{k=1}^m n_k}{2}, \frac{\nu_0\sigma_0^2+\sum_{k=1}^m \sum_{j=1}^{n_{j,k}}(\zeta_{j,k}-\theta_k)^2}{2}\right)$$

$$\kappa^2|-\sim \text{GI}\left(\frac{\xi_0+n}{2}, \frac{\xi_0\kappa_0^2+\sum_{k=1}^m \sum_{j=1}^{n_k} \sum_{i=1}^{n_{jk}}(y_{ijk}-\zeta_{jk})^2}{2}\right)$$

- **Modelo 5:** las DCC de  $\theta_k, \mu, \tau^2, \zeta_{jk}$  y  $\kappa^2$  se calcularon en modelos anteriores. Tenemos el siguiente espacios de parámetros  $\Phi = (\zeta, \Theta, \Sigma, \kappa^2, \mu, \tau^2, \sigma^2)$ . Donde  $\zeta = (\zeta_{1,1}, \dots, \zeta_{n_k,m})$ ,  $\Theta = (\theta_1, \dots, \theta_m)$ ,  $\Sigma = (\sigma_1^2, \dots, \sigma_m^2)$ .

$$\sigma_k^2|-\sim \text{GI}\left(\frac{\nu+n_k}{2}, \frac{\nu\sigma_0^2+\sum_{j=1}^{n_k}(\zeta_{jk}-\theta_k)^2}{2}\right)$$

$$\sigma^2|-\sim \text{G}\left(\frac{\alpha_0+\nu\cdot n}{2}, \frac{\beta_0+\nu\sum_{k=1}^m 1/\sigma_k^2}{2}\right)$$

## 6.2. Diagnósticos de convergencia

Se presentan resúmenes de la Cadena de Markov de Monte Carlo - CMCM, tales como: Error de Monte Carlo - EMC, Coeficiente de Variación de Monte Carlo - CVMC, y tamaños efectivos de muestras - neff. Para cada uno de los parámetros de cada modelo. Por lo tanto se obtuvieron los siguientes resultados:

### ■ MODELO 1

Estadísticos	$\theta$	$\sigma^2$
EMC	0.016	0.213
neff	10000	10000
CVMC	0.007	0.265

Tabla 5: Diagnosticos Modelo 1

### ■ MODELO 2

Parámetros	EMC	neff	CVMC
$\sigma^2$	0.050	9627	0.002
$\mu$	0.034	10000	0.014
$\tau^2$	1.026	10000	0.271

Tabla 6: Diagnóstico Modelo 2

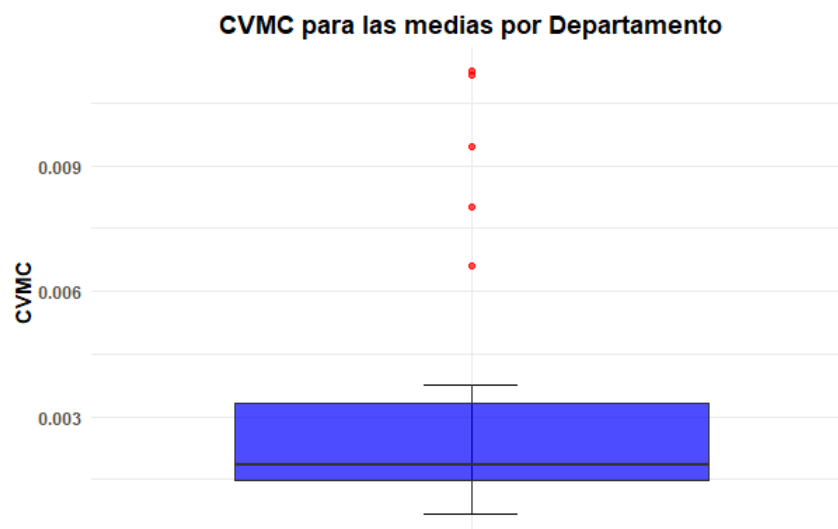


Figura 13: CVMC para las medias de los Departamentos

### ■ MODELO 3

Parámetros	EMC	neff	CVMC
$\mu$	0.034	10000	0.014
$\tau^2$	1.017	10000	0.269
$\sigma^2$	5.898	9517	0.254

Tabla 7: Diagnóstico Modelo 3

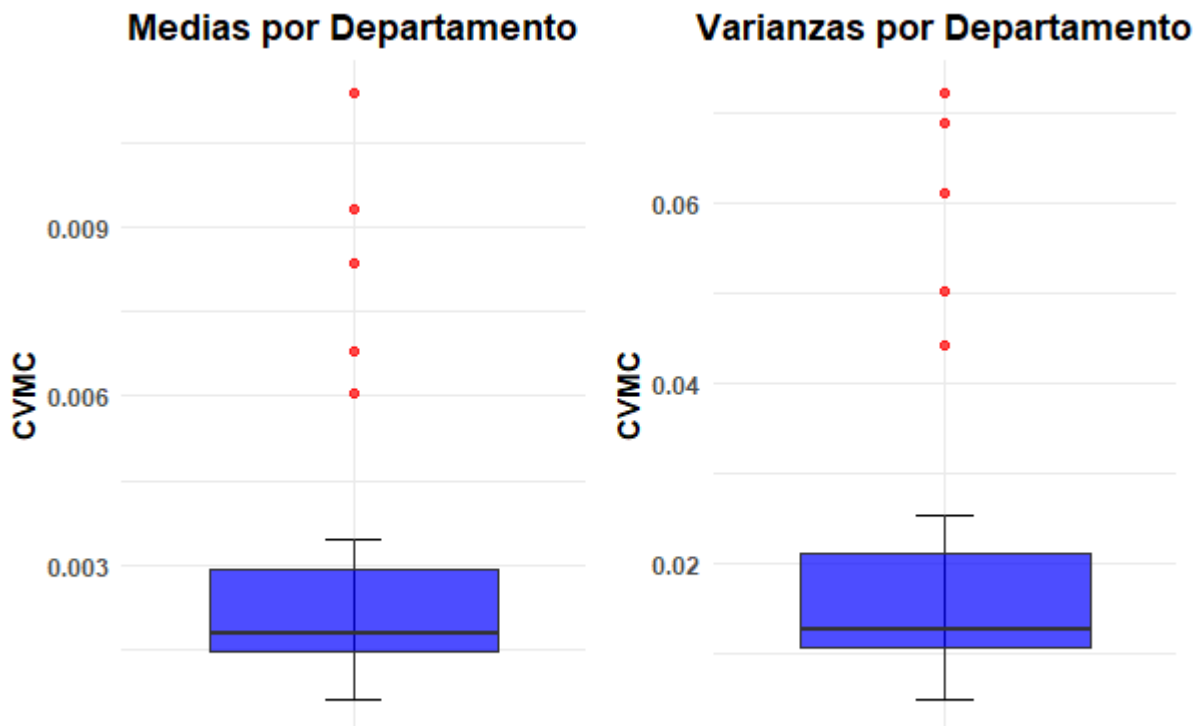


Figura 14: CVMC para las medias y varianzas de los departamentos

### ■ MODELO 4

Parámetros	EMC	neff	CVMC
$\kappa^2$	0.045	10000	0.002
$\mu$	0.034	10000	0.015
$\tau^2$	1.011	10000	0.287
$\sigma^2$	0.105	10000	0.047

Tabla 8: Diagnóstico Modelo 4

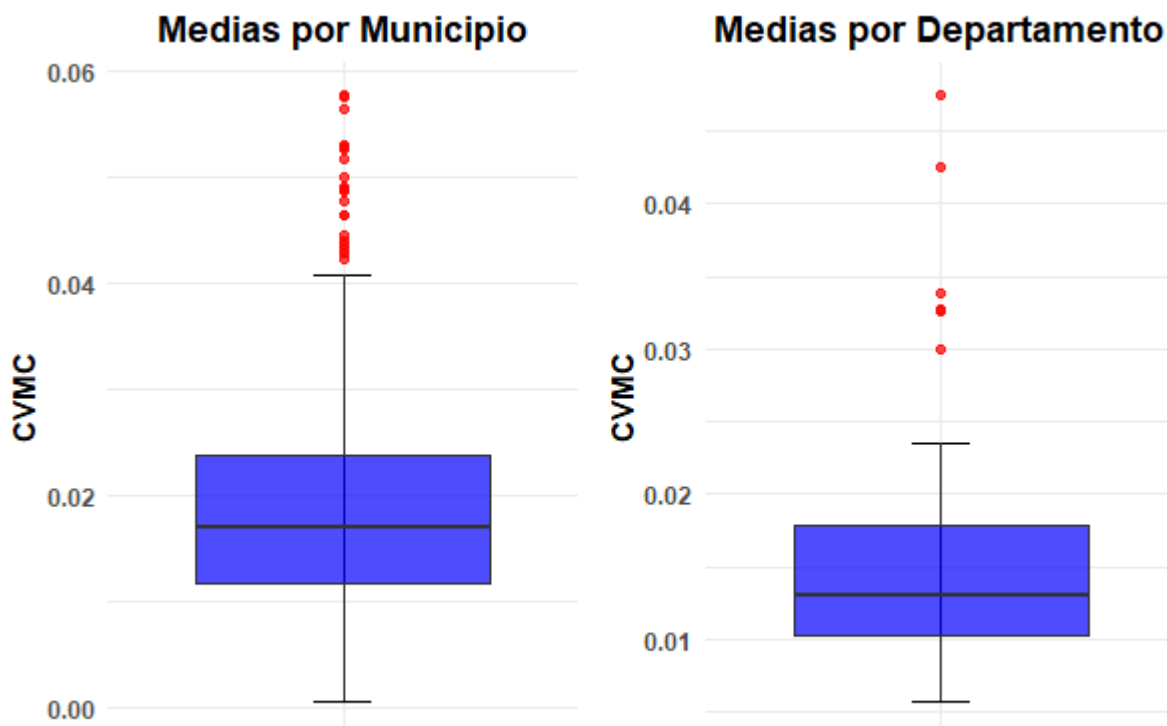


Figura 15: CVMC para las medias por Departamento y Municipio

#### ■ MODELO 5

Parámetros	EMC	neff	CVMC
$\mu$	0.034	10000	0.015
$\tau^2$	1.011	10000	0.290
$\sigma^2$	0.448	10255	0.273
$\kappa^2$	0.045	10000	0.002

Tabla 9: Diagnóstico Modelo 5

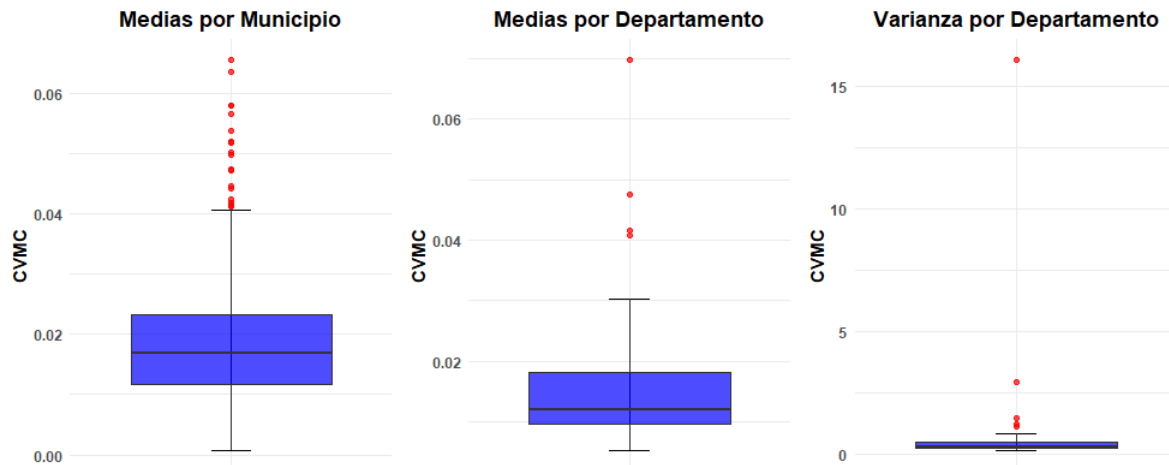


Figura 16: CVMC para las medias y varianzas por departamento y las medias por municipios

## 7. Referencias

### Referencias

- Consejo Nacional de Política Económica y Social. (2012). *Metodologías oficiales y arreglos institucionales para la medición de la pobreza en Colombia* (inf. téc.). Departamento Nacional de Planeación.
- Departamento Administrativo Nacional de Estadística (DANE). (2021). Pobreza monetaria en Colombia: Resultados 2020. [https://www.dane.gov.co/files/investigaciones/condiciones\\_vida/pobreza/2020/Presentacion-pobreza-monetaria\\_2020.pdf](https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2020/Presentacion-pobreza-monetaria_2020.pdf)
- Documentación del examen Saber 11. (2014).
- Guía de Usuario examen Saber 11. (2014).
- Ministerio de Educación Nacional. (2023). Tasa de Cobertura Neta Nacional [Accedido el 16 de octubre de 2023]. <http://bi.mineducacion.gov.co:8380/eportal/web/planeacion-basica/tasa-de-cobertura-neta1>