

Universidad Nacional de Colombia

Facultad de Ciencias
Departamento de Estadística

Primer caso de estudio Estadística Bayesiana

Juan Andrés Camacho Zárate
jucamachoz@unal.edu.co
Camilo Alejandro Raba Gomez
craba@unal.edu.co

Docente

Juan Camilo Sosa Martínez

Diciembre 2024

Introducción

La Prueba Saber 11 es un examen estandarizado diseñado para evaluar las competencias de los estudiantes que están a punto de finalizar la educación media. Esta prueba se aplica semestralmente a lo largo de todo el territorio nacional con el propósito de supervisar la calidad de la formación en la educación media y servir como requisito para acceder a la educación superior. El examen está construido de tal forma que los resultados de cada una de las cinco áreas evaluadas (Matemáticas, Lectura, Ciencias, Sociales e Inglés) se presenten en una escala discreta con límite inferior cero y límite superior 100. Además, se espera que los resultados tengan un promedio de 50 y una desviación estándar de 10 unidades.

Para este trabajo, el enfoque se centrará en la competencia de matemáticas, un área clave que evalúa el dominio de conocimientos básicos como la aritmética o la geometría, así mismo como habilidades esenciales como la interpretación de gráficos y razonamiento cuantitativo. Se analizarán los resultados obtenidos en Matemáticas recogidos de muestras aleatorias de los 32 departamentos del país y Bogotá durante los períodos comprendidos entre 2015 y 2023. Se ajustará un modelo Bayesiano Normal-Gamma Inversa-Normal utilizando los datos recolectados de la Prueba Saber 11 con los objetivos de:

1. Estimar parámetros para los puntajes de cada departamento.
2. Realizar un análisis comparativo de los resultados.
3. Evaluar la evolución temporal de los resultados y analizar las diferencias entre períodos.
4. Implementar un procedimiento de agrupamiento Bayesiano.

Modelo

Se considera el siguiente modelo Bayesiano:

1. **Distribución muestral:** Se supone que las observaciones y_i , para $i = 1, \dots, n$, son independientes e idénticamente distribuidas (iid) según una distribución Normal con media θ y varianza σ^2 :

$$y_i | \theta, \sigma^2 \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2).$$

2. **Distribución previa de θ condicional en σ^2 :** El parámetro θ sigue una distribución Normal con media μ_0 y varianza σ^2/κ_0 , donde $\kappa_0 > 0$ controla la precisión de la distribución previa:

$$\theta | \sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right).$$

3. **Distribución previa de σ^2 :** El parámetro σ^2 sigue una distribución Gamma Inversa con parámetro de forma $\nu_0/2$ y parámetro de escala $(\nu_0\sigma_0^2)/2$. Esta distribución es útil porque es conjugada para la varianza en el caso normal:

$$\sigma^2 \sim GI\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right).$$

4. **Distribución posterior:** Bajo el **modelo Normal-Gamma Inversa-Normal** se tiene que la **distribución posterior** de θ es

$$p(\theta, \sigma^2 | y) = p(\theta | \sigma^2, y) p(\sigma^2 | y)$$

donde:

- $p(\theta | \sigma^2, y) = N\left(\theta | \mu_n, \frac{\sigma^2}{\kappa_n}\right)$, donde

$$\mu_n = \frac{\kappa_0}{\kappa_n} \mu_0 + \frac{n}{\kappa_n} \bar{y} \quad \text{con} \quad \kappa_n = \kappa_0 + n,$$

- $p(\sigma^2 | y) = \text{GI} \left(\sigma^2 | \frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2} \right)$, donde

$$\nu_n = \nu_0 + n \quad \text{y} \quad \nu_n \sigma_n^2 = \nu_0 \sigma_0^2 + (n-1)s^2 + n \frac{\kappa_0}{\kappa_n} (\bar{y} - \mu_0)^2,$$

Este modelo se ajustará de forma independiente para cada departamento en cada período de tiempo, utilizando los hiperparámetros $\mu_0 = 50$, $\kappa_0 = 1$, $\nu_0 = 1$ y $\sigma_0^2 = 10^2$, mediante simulaciones de Monte Carlo con $B = 10000$ iteraciones.

Preguntas

1. Para Bogotá en 2023-2 ($t = 9$), calcule la media posterior, el coeficiente de variación posterior y el intervalo de credibilidad al 95% de confianza para la media θ , la desviación estándar σ y el coeficiente de variación $\zeta = 100 \frac{\sigma}{|\theta|}\%$. Reporte los resultados tabularmente.

Tabla 1: Inferencia sobre el puntaje promedio de Matemáticas de Bogotá para el periodo de 2023-2

Estimación	CV	L. Inf	L. Sup
55.082	0.008	54.239	55.925

Tabla 2: Inferencia sobre la desviación estandar del puntaje promedio de Matemáticas de Bogotá para el periodod de 2023-2

Estimación	CV	L. Inf	L. Sup
11.994	0.025	11.424	12.607

Tabla 3: Inferencia sobre el coeficiente de variación del puntaje promedio de Matemáticas de Bogotá para el periodo de 2023-2

Estimación	CV	L. Inf	L. Sup
21.776	0.026	20.673	22.963

2. En 2023-2 ($t = 9$), elabore el ranking de los cinco departamentos con mejores calificaciones promedio. Para cada uno de estos departamentos calcule la media posterior, el coeficiente de variación posterior y el intervalo de credibilidad al 95% de confianza para la media θ . Reporte los resultados tabularmente. Además, genere una visualización de estos cinco departamentos mediante un mapa de Colombia que utilice una escala de colores adecuada para representar la media posterior de la media θ .

Tabla 4: Top 5 departamentos con mejores puntajes promedio de matemáticas

	Estimación	CV	L. Inf	L. Sup
Santander	55.403	0.014	53.885	56.942
Bogotá, D.C.	55.082	0.008	54.239	55.925
Huila	54.678	0.021	52.404	56.978
Boyacá	54.025	0.017	52.201	55.815
Cundinamarca	53.848	0.010	52.751	54.955

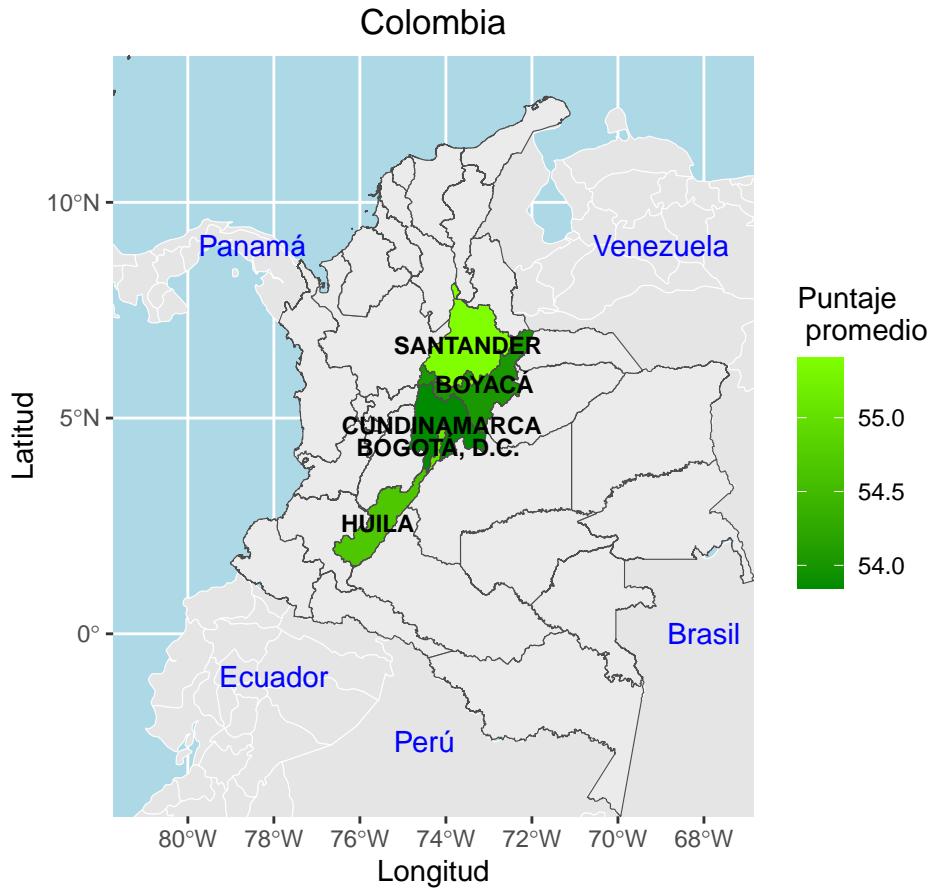


Figura 1: Top 5 departamentos con mejores puntajes promedio de matematicas en 2023-2.

3. En 2023-2 ($t = 9$), elabore el ranking de los cinco departamentos con peores calificaciones promedio. Para cada uno de estos departamentos calcule la media posterior, el coeficiente de variación posterior y el intervalo de credibilidad al 95% de confianza para la media θ . Reporte los resultados tabularmente. Además, genere una visualización de estos cinco departamentos mediante un mapa de Colombia que utilice una escala de colores adecuada para representar la media posterior de la media θ .

Tabla 5: Top 5 departamentos con peores puntajes promedio de matemáticas

	Estimación	CV	L. Inf	L. Sup
Vichada	38.024	0.147	27.099	49.058
Chocó	38.376	0.037	35.642	41.153
Amazonas	45.319	0.108	35.540	55.078
San Andrés	45.454	0.103	36.192	54.751
La Guajira	45.851	0.025	43.589	48.101

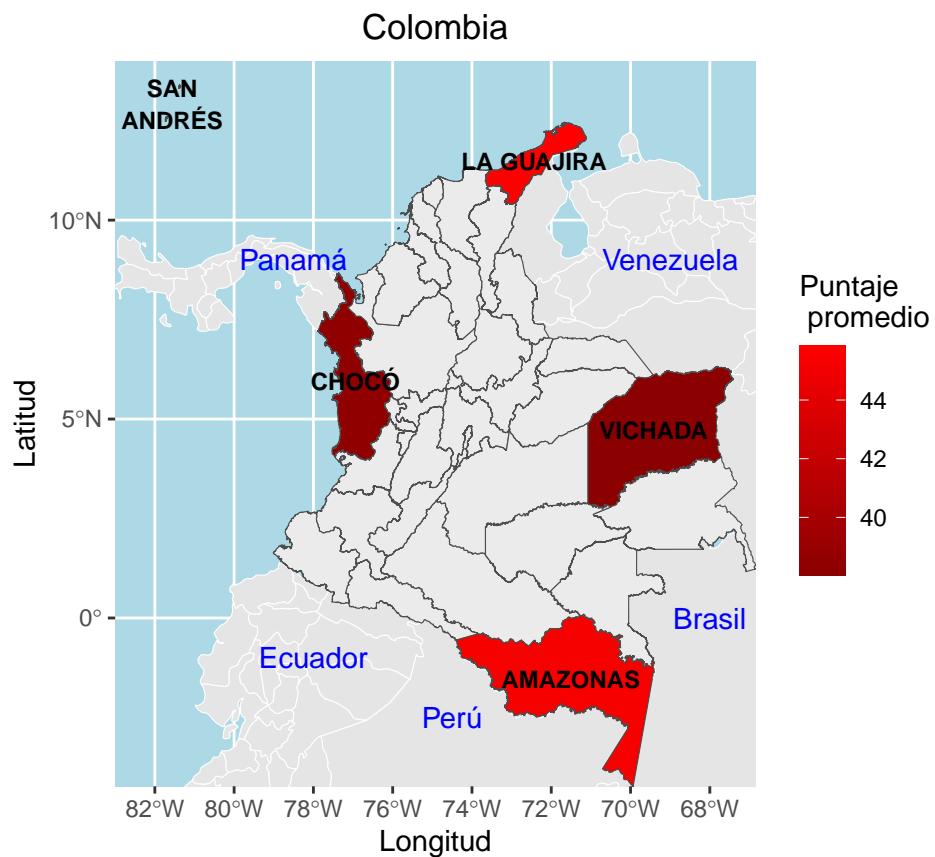


Figura 2: Top 5 departamentos con peores puntajes promedio de matematicas en 2023-2.

4. Para Bogotá en todos los períodos ($t = 1, \dots, 9$), calcule la media posterior y el intervalo de credibilidad al 95% de confianza para la media θ . Presente los resultados mediante una visualización que muestre simultáneamente las medias posteriores y los intervalos de credibilidad.

Tabla 6: Inferencia sobre el puntaje promedio de Bogotá a través del tiempo

	Estimación	CV	L. Inf	L. Sup
2015-2	54.162	0.007	53.369	54.961
2016-2	55.625	0.007	54.891	56.352
2017-2	54.096	0.007	53.350	54.831
2018-2	55.405	0.009	54.440	56.386
2019-2	54.819	0.009	53.850	55.803
2020-2	55.614	0.010	54.513	56.734
2021-2	54.345	0.008	53.544	55.152
2022-2	54.650	0.008	53.828	55.485
2023-2	55.082	0.008	54.239	55.925

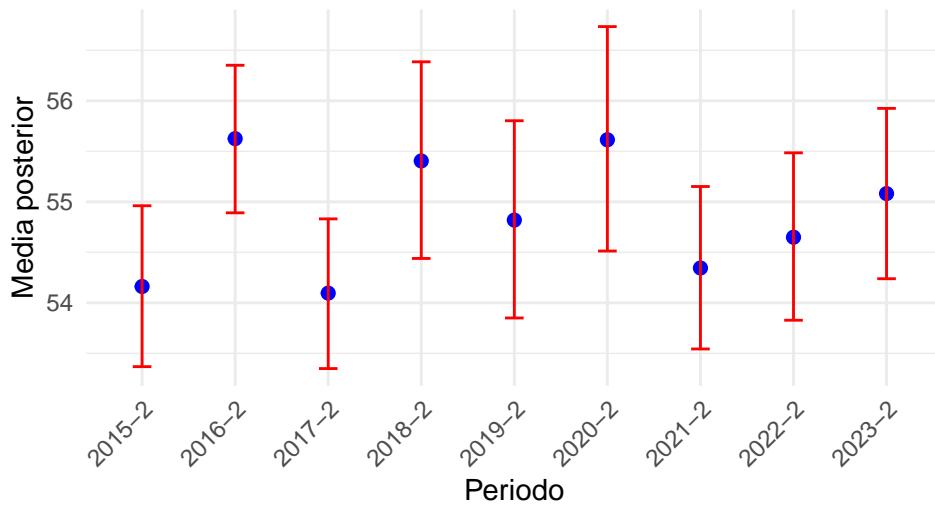


Figura 3: Media posterior de Bogotá y sus intervalos de credibilidad del 95% a través del tiempo.

5. Para Bogotá en todos los períodos ($t = 1, \dots, 9$), calcule la media posterior y el intervalo de credibilidad al 95% de confianza para el coeficiente de variación ζ . Presente los resultados mediante una visualización que muestre simultáneamente las medias posteriores y los intervalos de credibilidad.

Tabla 7: Inferencia sobre el coeficiente de variación del puntaje promedio de Bogotá a través del tiempo

	Estimación	CV	L. Inf	L. Sup
2015-2	22.362	0.025	21.304	23.497
2016-2	19.633	0.024	18.725	20.606
2017-2	20.678	0.024	19.733	21.690
2018-2	20.174	0.033	18.949	21.531
2019-2	20.549	0.033	19.288	21.940

2020-2	17.929	0.042	16.547	19.467
2021-2	20.834	0.026	19.773	21.938
2022-2	20.987	0.027	19.917	22.123
2023-2	21.776	0.026	20.673	22.963

Coeficiente de Variación del puntaje en matemáticas para Bogotá

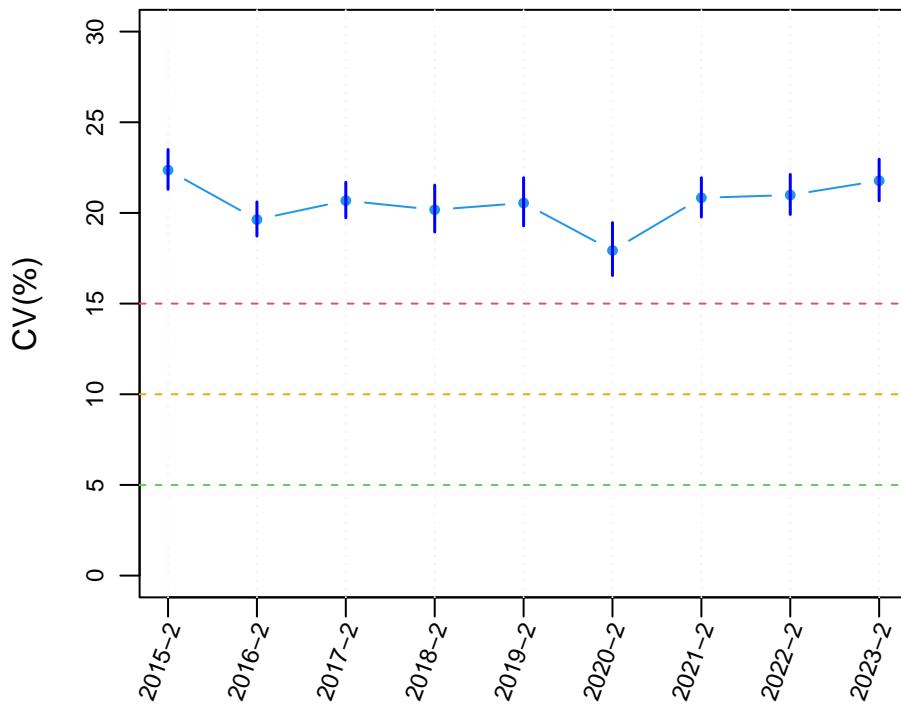


Figura 4: Estimación del coeficiente de variación para Bogotá y sus intervalos de credibilidad del 95% a través del tiempo.

6. Considere el parámetro $\eta_{t,k} = \theta_t - \theta_{t-k}$, para algún $k = 1, \dots, t-1$. Para Bogotá, calcule la media posterior y el intervalo de credibilidad al 95% de confianza para $\eta_{9,k}$, para cada $k = 1, \dots, 8$. Presente los resultados mediante una visualización que muestre simultáneamente las medias posteriores y los intervalos de credibilidad.

	Estimación	LI	LS
$\eta_{9,1}$	0.4319831	-0.7431620	1.5834279
$\eta_{9,2}$	0.7364352	-0.4046575	1.8897647
$\eta_{9,3}$	-0.5325337	-1.9151071	0.8658900
$\eta_{9,4}$	0.2623209	-1.0550157	1.5767503
$\eta_{9,5}$	-0.3232704	-1.6247022	0.9722876
$\eta_{9,6}$	0.9852949	-0.1400950	2.1056060

	Estimación	LI	LS
$\eta_{9,7}$	-0.5437484	-1.6539746	0.5732520
$\eta_{9,8}$	0.9195200	-0.2209856	2.0936569

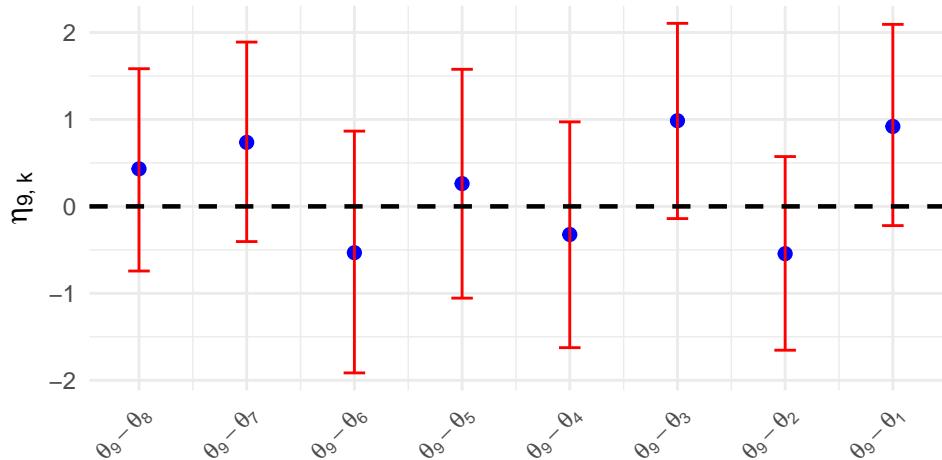


Figura 5: Diferencia de la media posterior de 2023-2 con respecto a los 8 períodos anteriores e intervalos de credibilidad (95%).

7. Considere el siguiente procedimiento Bayesiano de agrupamiento (clustering):

(a) **Construcción de la matriz $\Omega^{(b)}$:**

- Para cada iteración de las muestras de Monte Carlo ($b = 1, \dots, B$), construya la matriz:

$$\Omega^{(b)} = \begin{bmatrix} \theta_1^{(b)} & \sigma_1^{(b)} \\ \theta_2^{(b)} & \sigma_2^{(b)} \\ \vdots & \vdots \\ \theta_m^{(b)} & \sigma_m^{(b)} \end{bmatrix},$$

donde:

- $\theta_i^{(b)}$: Media estimada para el departamento i en la iteración b .
- $\sigma_i^{(b)}$: Desviación estándar estimada para el departamento i en la iteración b .
- $m = 33$: Número total de departamentos, incluyendo Bogotá.

(b) **Agrupamiento con k -medias:**

- En cada iteración b , utilice la matriz $\Omega^{(b)}$ para aplicar el algoritmo de segmentación k -medias mediante la función `kmeans` de R.
- Este procedimiento agrupa los departamentos en función de los valores de $\theta_i^{(b)}$ y $\sigma_i^{(b)}$, para $i = 1, \dots, m$.
- Como resultado, obtenga las etiquetas de grupo $\xi_1^{(b)}, \dots, \xi_m^{(b)}$, donde $\xi_i^{(b)} = k$ indica que el departamento i pertenece al grupo k en la iteración b .

(c) **Construcción de la matriz de similitudes A:**

- Con las etiquetas de grupo $\xi_1^{(b)}, \dots, \xi_m^{(b)}$ generadas en cada iteración, construya la matriz de similitudes de dimensión $m \times m$, denotada como $\mathbf{A} = [a_{ij}]$, donde:

$$a_{ij} = \mathbb{P}(\xi_i = \xi_j \mid \text{datos}) = \frac{1}{B} \sum_{b=1}^B I(\xi_i^{(b)} = \xi_j^{(b)}), \quad i, j = 1, \dots, m.$$

- Aquí:
 - $I(\cdot)$: Función indicadora que toma el valor 1 si los departamentos i y j pertenecen al mismo grupo en la iteración b , y 0 en caso contrario.
 - a_{ij} : Representa la probabilidad posterior de que los departamentos i y j pertenezcan al mismo grupo.
- Propiedades de \mathbf{A} :
 - Es una matriz simétrica.
 - Todos los elementos de la diagonal principal son iguales a 1.

(d) **Estimación puntual de la segmentación:**

- Utilice la matriz de similitudes \mathbf{A} como entrada para la función `Mclust` de la librería `mclust` en R.
- Esta función genera una segmentación final (estimación puntual) teniendo en cuenta las probabilidades contenidas en la matriz \mathbf{A} .

Para todos los períodos ($t = 1, \dots, 9$), aplique el procedimiento descrito previamente para obtener una segmentación de los 33 departamentos. Para cada período, presentar los resultados mediante un mapa de Colombia, utilizando una escala de colores adecuada en la que los departamentos que pertenezcan al mismo grupo se representen con el mismo color. La visualización debe consistir de 9 paneles dispuestos en un arreglo de 3×3 , todos en una misma página, para facilitar la comparación entre períodos.

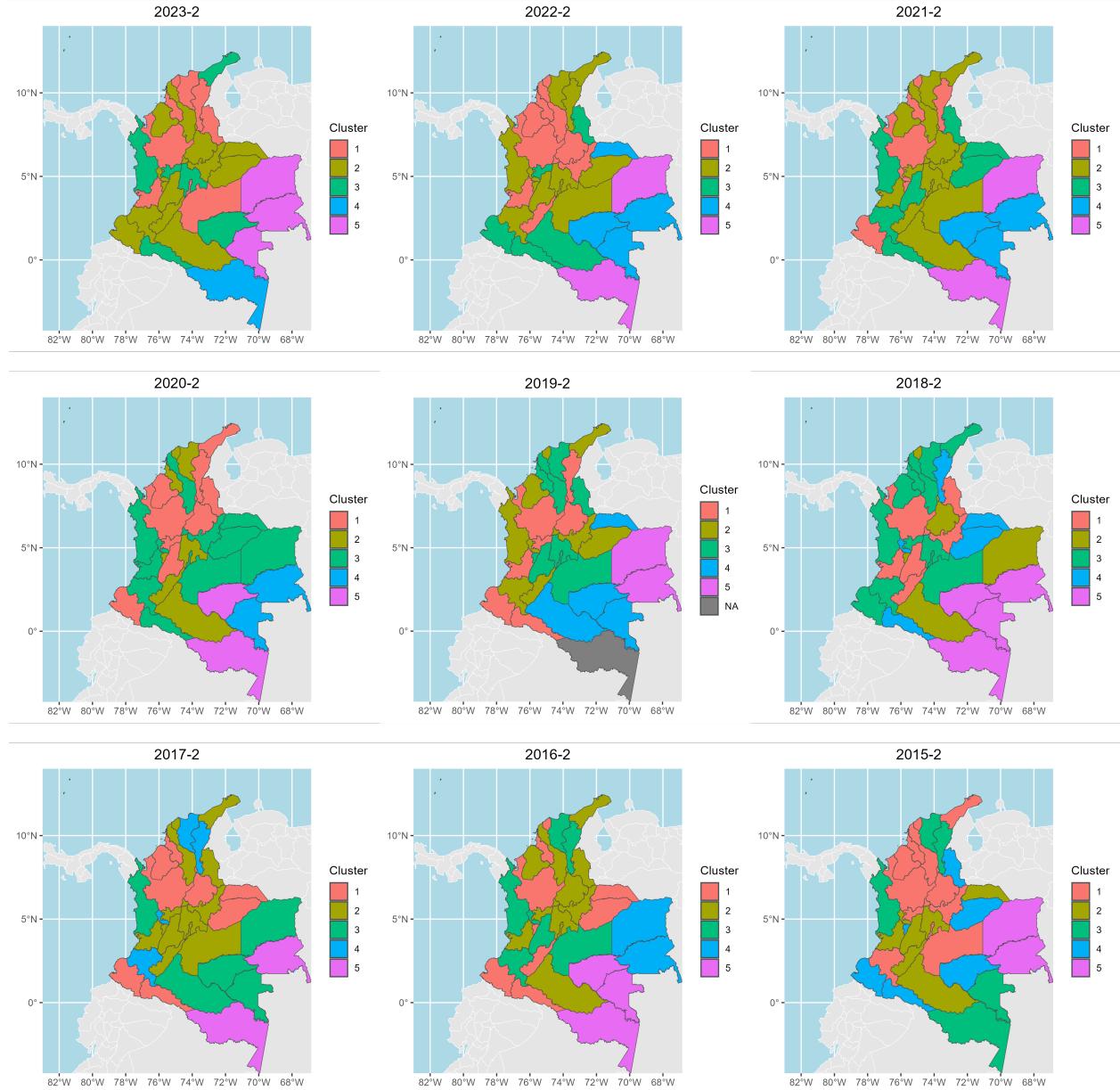


Figura 6: Agrupamiento dado por Mclust a través de los 9 períodos

8. Un **diagrama de Sankey** es una excelente herramienta para visualizar la evolución de un proceso de clustering, ya que puede mostrar cómo los elementos cambian de un grupo a otro a lo largo de diferentes etapas o períodos.

Representar la evolución del proceso de clustering obtenido en el numeral anterior mediante un diagrama de Sankey, mostrando cómo los elementos se redistribuyen entre los diferentes grupos a lo largo de los períodos.

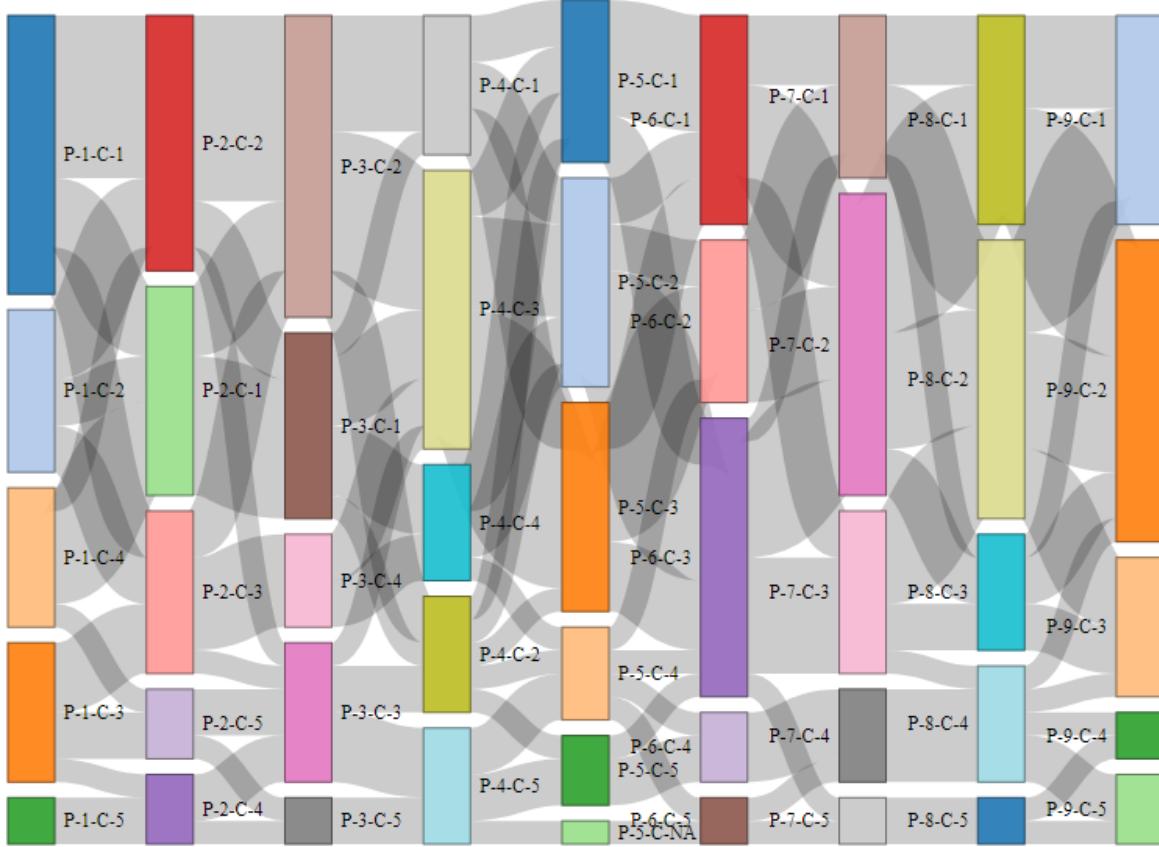


Figura 7:Diagrama de Sankey

9. Interprete **todos** los resultados obtenidos (máximo dos mil palabras).

Punto 1. El análisis, a través de la inferencia bayesiana para Bogotá en el período 2023-2 proporciona información clave sobre el rendimiento académico en Matemáticas. La media posterior se estima en **55.082 puntos**, con un intervalo de credibilidad al 95% que varía entre **54.239** y **55.925**. Esto indica que Bogotá supera el puntaje promedio esperado de la prueba, diseñado para ser **50 puntos**. Este resultado sugiere que el desempeño académico promedio de los estudiantes de Bogotá es significativamente superior al estándar establecido.

La desviación estándar (σ), estimada en **11.994**, refleja la dispersión de los puntajes individuales alrededor de la media. Este valor es ligeramente mayor que la desviación estándar esperada de la prueba, que es de **10 puntos**, lo que indica una mayor variabilidad en los resultados de Bogotá en comparación con el diseño inicial de la prueba, lo cual también se evidencia en el intervalo de credibilidad para σ , que va desde **11.424** hasta **12.607**.

El coeficiente de variación (ζ) del **21.776%** cuantifica la relación entre la desviación estándar y la media, lo que permite interpretar la dispersión relativa de los datos. Con un intervalo de credibilidad entre **20.673%** y **22.963%**, Bogotá muestra una dispersión moderada en relación con su promedio, sugiriendo que, aunque el desempeño es alto, existe una variabilidad relativa que merece atención para identificar áreas de mejora.

En conjunto, los resultados indican que Bogotá tiene un desempeño promedio sólido en Matemáticas, superando el estándar esperado de la prueba tanto en promedio como en rango de confianza. Sin embargo,

la desviación estándar mayor a la esperada sugiere la posibilidad de disparidades dentro del grupo de estudiantes, lo que podría ser atribuido a diferencias en factores como acceso a recursos educativos, infraestructura o preparación docente. Reducir esta variabilidad podría ser un objetivo clave para lograr un sistema educativo más equitativo y consistente.

Punto 2.

En el período 2023-2 ($t = 9$), se identificaron los cinco departamentos con los mejores puntajes promedio en la prueba de Matemáticas. El departamento con el mejor desempeño fue **Santander**, con una media posterior estimada en **55.403 puntos**, seguido por **Bogotá, D.C.** con **55.082 puntos**, **Huila** con **54.678 puntos**, **Boyacá** con **54.025 puntos**, y **Cundinamarca** con **53.848 puntos**. Todos ellos presentan puntajes superiores al promedio esperado de la prueba, diseñado en **50 puntos**, lo que refleja un rendimiento sobresaliente en comparación con el estándar nacional.

Es importante destacar que estos departamentos se encuentran ubicados principalmente en la región central del país. Este patrón geográfico sugiere que las áreas cercanas al núcleo económico y político de Colombia, donde se encuentran Bogotá y Cundinamarca, podrían beneficiarse de mejores recursos educativos, infraestructura, y políticas públicas más eficaces en la promoción de la educación.

En términos de variabilidad relativa, representada por el coeficiente de variación (CV), los cinco departamentos muestran una dispersión baja, lo que indica una alta consistencia en los puntajes dentro de cada región. **Bogotá, D.C.** exhibe la mayor estabilidad relativa con un CV de **0.008**, mientras que **Huila** presenta una dispersión ligeramente mayor, con un CV de **0.021**.

Los intervalos de credibilidad al 95% para la media posterior son estrechos para todos los departamentos, lo que denota un alto nivel de certeza en las estimaciones. Estos resultados sugieren que los cinco departamentos destacados tienen un desempeño consistente y sólido en la prueba de Matemáticas. Sin embargo, se resalta la necesidad de entender las ventajas de la región central y evaluar cómo podrían extenderse estas condiciones favorables a otras áreas del país para reducir las brechas educativas.

Punto 3.

En el período 2023-2 ($t = 9$), se identificaron los cinco departamentos con los puntajes promedio más bajos en la prueba de Matemáticas. El departamento con el desempeño más bajo fue **Vichada**, con una media posterior estimada en **38.024 puntos**, seguido por **Chocó** con **38.376 puntos**, **Amazonas** con **45.319 puntos**, **San Andrés** con **45.454 puntos**, y **La Guajira** con **45.851 puntos**. Todos ellos presentan puntajes significativamente inferiores al promedio esperado de la prueba, diseñado en **50 puntos**, lo que evidencia desafíos importantes en el desempeño académico de estas regiones.

Es importante destacar que estos departamentos se encuentran en zonas geográficamente alejadas y dispersas del territorio nacional, como la Amazonía, los Llanos Orientales, el Caribe insular y el Pacífico colombiano. Estas áreas suelen enfrentar limitaciones estructurales, como el acceso reducido a recursos educativos, infraestructuras escolares insuficientes, y menor presencia de programas de apoyo pedagógico. Esto podría contribuir a las brechas significativas observadas en los resultados.

En términos de variabilidad relativa, representada por el coeficiente de variación (CV), los resultados muestran diferencias notables entre los departamentos. **Chocó** exhibe la mayor estabilidad relativa con un CV de **0.037**, mientras que **Vichada** presenta la mayor dispersión relativa, con un CV de **0.147**. Esto último indica una mayor heterogeneidad en los puntajes de los estudiantes de Vichada, posiblemente debido a desigualdades internas en el acceso a recursos educativos.

Estos resultados subrayan la necesidad de implementar políticas públicas diferenciadas para las regiones más alejadas y vulnerables del país, buscando reducir las barreras geográficas y mejorar la equidad educativa en todas las zonas del territorio nacional.

Punto 4.

El análisis de la evolución del puntaje promedio en Matemáticas para Bogotá a lo largo de los períodos 2015-2 a 2023-2 muestra un rendimiento consistentemente superior al promedio esperado de la prueba. En todos los períodos, la media posterior (θ) se mantiene por encima de este estándar, oscilando entre **54.096** (2017-2) y **55.625** (2016-2). Esto refleja un desempeño sólido y sostenido en la capital.

El coeficiente de variación (CV) para todos los períodos es extremadamente bajo, con valores que oscilan entre **0.007** y **0.010**, lo que indica una alta consistencia en los puntajes de los estudiantes de Bogotá. Este nivel de estabilidad sugiere que las diferencias individuales entre los puntajes son mínimas, consolidando la capital como una de las regiones con menor dispersión en el desempeño académico.

Los intervalos de credibilidad al 95% para la media posterior son estrechos, lo que denota un alto nivel de certeza en las estimaciones.

Punto 5.

El análisis del coeficiente de variación (ζ) del puntaje promedio en Matemáticas para Bogotá a lo largo de los períodos 2015-2 a 2023-2 muestra estimaciones que oscilan entre **17.929%** (2020-2) y **22.362%** (2015-2). Estos valores se mantienen cercanos al coeficiente de variación esperado de **20%**, diseñado para la prueba, lo que indica una estabilidad en la dispersión relativa de los puntajes en Bogotá.

Según los criterios del DANE, un coeficiente de variación se interpreta de la siguiente manera:

- **Hasta el 7%:** la estimación es **precisa**.
- **Entre 7% y 15%:** la estimación tiene una **variación moderada**.
- **Mayor al 15%:** la estimación presenta una **alta variabilidad**.

Aplicando estos criterios, los valores de ζ observados en Bogotá se clasifican dentro de la categoría de **alta variabilidad**. Es notable que, aunque los puntajes promedio de Bogotá superan consistentemente el estándar nacional de **50 puntos**, la dispersión relativa de los resultados es significativa.

El período **2020-2** presenta el coeficiente de variación más bajo con **17.929%**, coincidiendo con el año de mayor impacto de la pandemia del COVID-19. Esta reducción en la variabilidad podría reflejar una homogeneización en los resultados, posiblemente debido a la transición a la educación virtual, la uniformidad en las condiciones de aprendizaje durante ese período y el hecho de que en ese año, debido a esta emergencia sanitaria se realizó la prueba en una sola sesión, a diferencia de los otros años en los que la prueba se realizaba en 2 sesiones.

Punto 6.

Este análisis permite evaluar la evolución de los puntajes promedio a lo largo del tiempo, considerando las diferencias respecto al período más reciente.

- **Estimaciones positivas y negativas:**

- Para los períodos más cercanos ($\eta_{9,1}$ y $\eta_{9,2}$), las estimaciones son **positivas** (0.432 y 0.736, respectivamente). Esto indica que los puntajes promedio en 2023-2 fueron mayores en comparación con 2022-2 y 2021-2.
- En cambio, para períodos más alejados ($\eta_{9,3}$, $\eta_{9,5}$ y $\eta_{9,7}$), las estimaciones son **negativas**, lo que sugiere que los puntajes en 2023-2 fueron más bajos respecto a esos años.

- **Intervalos de credibilidad amplios:**

- Los intervalos de credibilidad al 95% muestran que para varios períodos (por ejemplo, $\eta_{9,3}$ y $\eta_{9,5}$), incluyen tanto valores positivos como negativos, indicando una alta incertidumbre respecto a la dirección del cambio. Esto sugiere que las diferencias podrían no ser estadísticamente significativas.

- **Períodos con cambios notables:**

- $\eta_{9,6}$ muestra la mayor estimación positiva (0.985), lo que podría reflejar un cambio significativo hacia un desempeño más alto en 2023-2 comparado con 2017-2.
- Por otro lado, $\eta_{9,7}$ presenta una diferencia negativa (-0.544), indicando un desempeño inferior en comparación con 2016-2.

- **Tendencia general:**

- Aunque hay fluctuaciones, no se observa una tendencia clara de mejora o deterioro sistemático en los puntajes promedio. Esto podría estar relacionado con variaciones en factores educativos, socioeconómicos o metodológicos en los periodos evaluados.

En resumen, el análisis de $\eta_{9,k}$ evidencia que, si bien Bogotá mantiene cierta estabilidad en su desempeño, los cambios respecto a períodos anteriores no son siempre consistentes ni estadísticamente significativos. Este comportamiento subraya la importancia de continuar monitoreando las tendencias a largo plazo para identificar patrones relevantes en el rendimiento académico.

Punto 7.

- **Custer 1:** Este grupo incluye departamentos del suroeste del país, como Nariño, Cauca y Valle del Cauca, en varios períodos. Dicho comportamiento es una señal de similitud en el comportamiento del puntaje promedio y su variabilidad.
- **Cluster 2:** Aparece predominantemente en el norte y centro del país, incluyendo departamentos como Santander, Antioquia y Atlántico. Estos departamentos se caracterizan por ser centros urbanos importantes y zonas económicas más desarrolladas, lo cual podría reflejar mejores condiciones educativas y socioeconómicas que resultan en mejores puntajes y una menor variabilidad al presentarse mayor cantidad de información.

-Cluster 3: En varios períodos, incluye departamentos del sureste amazónico y Orinoquía, como Guainía, Vaupés y Vichada. Al ser estos departamentos aislados y con dificultades para el acceso a la educación se sospechan puntajes por debajo del promedio y alta variabilidad dada la poca información disponible (pocas muestras).

-Cluster 4: Incluye departamentos costeros y del Caribe, como Bolívar, Córdoba y Magdalena. En general destaca la diversidad en resultados educativos en el Caribe.

-Cluster 5: No presenta un patrón definido, en varios períodos agrupa departamentos dispersos, como Amazonas o San Andrés. Estos departamentos son excepciones geográficas o administrativas, lo que podría influir en su agrupamiento.

También resalta en 2019-2 la no presencia de valores para Amazonas.

Punto 8.

Se presentan flujos significativos donde departamentos de un cluster migran a varios clusters en el siguiente período. Esto da indicios de cambios en las características de los departamentos o fluctuaciones en los datos.

Para algunos períodos, un cluster se divide en varios clusters en el siguiente período, lo que indica diversificación en las características de los departamentos (por ejemplo, cambios en resultados de puntajes). Por ejemplo, el Cluster 4 en el período 5 (“P-5-C-4”). Este cluster se diversifica, enviando departamentos a varios clusters en el período 6 (“P-6-C-2”, “P-6-C-3”, etc.) y para el Cluster 3 en el período 7 (“P-7-C-3”) se divide y distribuye en tres clusters en el período 8 (“P-8-C-2”, “P-8-C-4”, “P-8-C-5”).

El diagrama refleja lo anteriormente mencionado respecto al cluster 5, pues se evidencia que tienen menos conexiones, lo que podría reflejar que los departamentos en ellos tienen características únicas o atípicas.

Referencias

- [1] Juan Sosa. *Estadística Bayesiana*. Google Sites, 2024. Disponible en: <https://sites.google.com/view/juansosa/bayesian-statistics>.
- [2] Peter Hoff. *A first course in Bayesian statistical methods*, volumen 580. Springer, 2009.
- [3] Schmidt, M. (2008). *The sankey diagram in energy and material flow management: part II: methodology and current applications*. *Journal of Industrial Ecology*, 12(2), 173–185.