

# Notas de Investigación

Modelo de estimación de  
áreas pequeñas para la  
prueba de Matemáticas de  
Saber 9° 2022

**Investigadores del proyecto**

Hanwen Zhang  
Juan José Rubio  
Nelson Rodríguez

Diciembre 2025

## Notas de Investigación

Edición No. 16  
Diciembre, 2025  
Edición digital

**Instituto Colombiano para la Evaluación de la Educación**  
Oficina de Gestión de Proyectos de Investigación  
Calle 26 N. 69-76, Edificio Elemento, Torre II, piso 18, Bogotá,  
D. C.  
Teléfono: (601) 4841410  
[proyectosinvestigacion@icfes.gov.co](mailto:proyectosinvestigacion@icfes.gov.co)  
<https://www.icfes.gov.co>

### Directora general

Elizabeth Blandón Bermúdez

### Autores

Hanwen Zhang  
Juan José Rubio  
Nelson Rodríguez

### Advertencia:

El contenido de este documento es el resultado de investigaciones y obras protegidas por la legislación nacional e internacional. No se autoriza su reproducción, utilización ni explotación a NINGÚN tercero. Solo se autoriza su uso para fines exclusivamente académicos. Esta información no podrá ser alterada, modificada o enmendada.

### Citar este documento en estilo APA así:

Zhang, H., Rubio, J., Rodríguez, N. (2025). *Modelo de estimación de áreas pequeñas para la prueba de Matemáticas de Saber 9° 2022*. Nota de Investigación No. 16). Instituto Colombiano para la Evaluación de la Educación (Icfes).  
<https://www.icfes.gov.co/web/guest/saber-investigar>

## **Modelo de estimación de áreas pequeñas para la prueba de Matemáticas de Saber 9° 2022<sup>1</sup>**

*Hanwen Zhang<sup>2</sup>*

*Juan José Rubio<sup>3</sup>*

*Nelson Rodríguez<sup>2</sup>*

### **Resumen**

Las pruebas Saber 3°, 5°, 7° y 9° son un conjunto de instrumentos que entregan información sobre el desarrollo de las competencias de los estudiantes. Estas pruebas no son censales, sino muestrales, puesto que se aplica un diseño muestral multietápico para la selección de estudiantes, el cual permite obtener estimaciones precisas del nivel de desempeño de diferentes pruebas a nivel de departamentos, regiones, zona, sector, pero no se puede obtener estimaciones a nivel de municipios. En este trabajo, liderado por la Subdirección de Estadísticas del Icfes, se aplica un modelo de estimación de áreas pequeñas para obtener estimaciones a nivel de municipios para la prueba de Matemáticas aplicada al grado noveno. Específicamente se aplica un modelo de unidad EB (empirical Bayes) bayesiano para establecimientos, y como covariables se utilizó información disponible en el DANE y el Ministerio de Educación Nacional. Una vez estimado y validado el modelo, se calcula la predicción del puntaje para todos los establecimientos no muestreados, y los valores agregados se calculan a nivel municipal después de aplicar el proceso de Benchmarking. De esta forma, este trabajo entrega insumos a nivel de municipios que permitirán la formulación de políticas públicas específicas para la mejora de los desempeños de los estudiantes.

---

<sup>1</sup> Las ideas, opiniones, tesis y argumentos expresados son de propiedad exclusiva de los autores y no representan el punto de vista del Icfes. Este proyecto forma parte del proyecto "Mapa con estimaciones por municipio para examen Saber 3°, 5°, 7° y 9°, basado en metodología de estimación en áreas pequeñas" del laboratorio de innovación en evaluación del instituto - IcfesLab

<sup>2</sup> Instituto Colombiano para la Evaluación de la Educación – Icfes, [estadisticapqrs@icfes.gov.co](mailto:estadisticapqrs@icfes.gov.co)

<sup>3</sup> Departamento Administrativo Nacional de Estadística - DANE, [jjrubiom@dane.gov.co](mailto:jjrubiom@dane.gov.co)



**Palabras claves:** estimación de áreas pequeñas; modelo de unidad; prueba Saber

## **Small Area Estimation Model for the Saber Grade 9 Mathematics Test 2022**

*Hanwen Zhang<sup>4</sup>*

*Juan José Rubio<sup>5</sup>*

*Nelson Rodríguez<sup>4</sup>*

### **Abstract**

The Saber exams for 3rd, 5th, 7th and 9th grades are a set of instruments that provide information on the development of student competencies. These tests are not census-based, but sample-based, since a multistage sample design is applied for the selection of students, which allows obtaining precise estimates of the competencies of different tests at the department, region, zone and sector levels, but estimates cannot be obtained at the municipal level. In this work, led by the Statistics Subdirectorate of Icfes, a small area estimation model is applied to obtain estimates at the municipal level for the mathematics test applied to the ninth grade. Specifically, a Bayesian EB (empirical Bayes) unit model is applied for schools, and information available from DANE and the National Ministry of Education was used as covariates. Once the model is estimated and validated, the score prediction is calculated for all non-sampled schools, and the aggregate values are calculated at the municipal level after applying the Benchmarking process. In this way, this work provides inputs at the municipal level that will allow the formulation of specific public policies to improve student performance.

**Keywords:** small area estimation; unit model; Saber test

---

<sup>4</sup> Instituto Colombiano para la Evaluación de la Educación – Icfes, [estadisticapqrs@icfes.gov.co](mailto:estadisticapqrs@icfes.gov.co)

<sup>5</sup> Departamento Administrativo Nacional de Estadística - DANE, [jjrubiom@dane.gov.co](mailto:jjrubiom@dane.gov.co)

## **Contenido**

1.	Introducción .....	4
2.	Revisión de literatura y estado del arte.....	5
3.	Marco teórico .....	6
3.1.	Modelo de calificación .....	7
3.2.	Diseño muestral Saber 3º, 5º, 7º y 9º .....	8
3.3.	Modelo SAE de unidad .....	9
3.4.	Selección del modelo.....	10
3.5.	Predicciones .....	11
3.6.	Benchmarking .....	12
3.7.	Incorporando la variabilidad de los puntajes.....	13
4.	Información y datos.....	14
4.1.	Información de EDUC .....	14
4.2.	Información de SIMAT .....	15
1.1.	Información de otras pruebas Saber .....	15
4.3.	Información de municipios.....	16
4.4.	Consolidación de la información .....	16
5.	Resultados .....	16
6.	Conclusiones .....	20

## 1. Introducción

Las pruebas Saber 3°, 5°, 7° y 9° son un conjunto de instrumentos que buscan aportar a la construcción de una educación de calidad para Colombia, entregando información sobre el desarrollo de competencias de los estudiantes, según lo propuesto en los estándares básicos de competencias del Ministerio de Educación Nacional. El examen también suministra insumos acerca de las habilidades socioemocionales y factores asociados que puedan incidir en el desarrollo de los aprendizajes.

El estudio principal de las pruebas Saber 3°, 5°, 7° y 9° se encuentra limitado a los estudiantes matriculados en los grados tercero, quinto, séptimo y noveno de educación básica y secundaria impartida en los establecimientos educativos oficiales y no oficiales, urbanos y rurales del país. La aplicación de la prueba se lleva a cabo por medio de un diseño muestral probabilístico estratificado multietápico, y tiene como objetivo principal la generación de estimaciones insesgadas y precisas del desempeño de los estudiantes en las áreas de interés para algunas agregaciones como son: zona, sector y población en condición de discapacidad, a nivel nacional y departamental. Específicamente, en la primera etapa se seleccionan los establecimientos educativos por medio de un muestreo estratificado con selección sistemática, usando probabilidades de inclusión proporcionales al número de estudiantes matriculados en los grados de interés; en la segunda etapa se seleccionan las sedes con un muestreo aleatorio simple, y finalmente en la última etapa, a todos los estudiantes se le asigna los cuadernillos de Matemáticas y Lectura, sin embargo, las demás áreas se asignan de manera aleatoria pero sistemática. Los detalles sobre el diseño muestral se encuentran en el documento Informe técnico para la selección de la muestra 3°, 5°, 7°, 9° - 2022, lápiz y papel (Icfes, 2022).

Los resultados de las pruebas Saber 3°, 5°, 7° y 9° están publicados en el Informe Nacional de los Resultados de las Pruebas Saber 3°, 5°, 7° y 9° - 2022, donde se encuentran las estimaciones de los puntajes promedios por zona (rural / urbano), sector (oficial / no oficial), y sexo (hombre / mujer), y también a nivel de regiones como a nivel de departamentos. A nivel municipal, no se encuentran estimaciones precisas de los

puntajes, ya que el diseño muestral no contempla la representatividad municipal puesto que en primer lugar la muestra no cubre todos los municipios del territorio nacional, además en los municipios donde sí se seleccionaron sedes tampoco se cuentan con un tamaño muestral suficiente para producir estimaciones con un error de estimación bajo. De allí surge la importancia de aplicar los **modelos de estimación de áreas pequeñas** (SAE, por sus siglas en inglés *Small Area Estimation*) para lograr predecir el puntaje promedio en la mayoría de los municipios del país, en este documento se presenta su aplicación al puntaje promedio de Matemáticas del grado 9º.

## 2. Revisión de literatura y estado del arte

Los modelos SAE son utilizados ampliamente en los años recientes para producir estadísticas oficiales en niveles de desagregación no planeados en el diseño de un estudio por muestreo, lo anterior, en estadísticas de pobreza, salud, indicadores socio económicos, etc. Por mencionar algunos, el Departamento Administrativo Nacional de Estadística de Colombia en el 2024 estimó la Escala de Experiencia de Inseguridad Alimentaria (FIES) a nivel de municipios con base en los datos de la Encuesta Nacional de Calidad de Vida (ECV) de Colombia de 2022; Casas-Cordero Valencia, C., Encina, J. y Lahiri, P. (2016) estimaron la incidencia de pobreza en las comunas de Chile con base en los datos de la Encuesta de Caracterización Socioeconómica Nacional de Chile (CASEN) de 2009; también se puede encontrar aplicaciones demográficas tal como lo indica en M. Sutton, A., W. Almquist, Z. (2023).

Los modelos SAE combinan la información proveniente de encuestas (que generalmente no tienen representatividad para niveles finos de desagregación) con información de censos y registros administrativos para producir estimaciones con bajos errores de estimación para dominios "pequeños". Cabe destacar que la palabra "pequeño" no hace referencia únicamente a las áreas geográficas sino a dominios donde el tamaño muestral es nulo o insuficiente para producir estadísticas confiables, y pueden corresponder a municipios, grupos poblacionales de ciertas características, etc.

Las bases teóricas de los modelos SAE combinan en general la teoría de muestreo y los modelos mixtos, y se clasifican en los modelos de área y los modelos de unidad.

Los modelos de área en general son menos restrictivos ya que utilizan solamente datos agregados para las áreas o dominios de estimación, el famoso modelo de Fay-Herriot (FH) pertenece a este grupo (Fay & Herriot, 1979). Por otro lado, los modelos de unidad se establecen para cada individuo de la población, y por tanto el ajuste de estos modelos requiere disponer de datos individuales de la variable respuesta y las variables auxiliares. Para mayores detalles sobre un panorama general de los modelos SAE, se recomienda consultar a Molina (2019).

Ahora, en el contexto de evaluación de la educación no se encuentran estudios o documentos a nivel internacional sobre el uso de estos modelos en datos de evaluaciones educativas que provengan de un modelo de Teoría de Respuesta al Ítem (TRI). Por lo tanto, este trabajo busca articular metodologías que sean válidas para producir indicadores de la calidad de la educación para áreas pequeñas (municipios de Colombia), de esta forma se suple la necesidad de contar con resultados oficiales de pruebas no censales (Saber 3º, 5º, 7º y 9º) a nivel municipal, ya que en este momento no se publican dichos resultados pues están asociados con errores muestrales muy grandes o al no disponerse de ningún elemento en la muestra no es posible la generación de ningún agregado. De esta forma, esta investigación entregará nuevos insumos e información que permitirán una mirada más profunda al panorama nacional en cuanto a la calidad educativa.

### **3. Marco teórico**

En esta sección se describe el marco teórico que fundamenta el ajuste del modelo de áreas pequeñas. Para ello se hace énfasis en que nuestra variable de interés, el puntaje promedio de Matemáticas del grado 9º, no es una característica directamente observable, sino que proviene de la estimación de un modelo de Teoría de Respuesta al Ítem ya que corresponde a la habilidad de los estudiantes que es una característica latente, de esta forma, la variable de interés viene con una incertidumbre debido a estos modelos. Por otro lado, las pruebas Saber 3º, 5º, 7º y 9º no son pruebas censales, sino que proviene de un diseño muestral probabilístico, por lo tanto, también introduce cierta incertidumbre en el puntaje de los estudiantes. Usando propiedades de la teoría de

varianza condicional se encuentra que la varianza del puntaje de un estudiante viene dada por la varianza debida al modelo TRI más la varianza inducida por el diseño muestral.

Una vez se abordan estas características, se profundiza en el modelo SAE de unidad que se utilizará para predecir los puntajes a nivel de municipios, considerando a los establecimientos educativos como las unidades. La razón de optar por un modelo de unidad, y no uno de área se debe a que se cuenta con información detallada de registros administrativos para los establecimientos provenientes de entidades públicas como el Ministerio de Educación Nacional y el Departamento Administrativo Nacional de Estadística. Para un mejor entendimiento del modelo SAE de unidad, se describe el modelamiento cuando la variable de interés es una constante, y posteriormente se describe el proceso computacional para incorporar la varianza de la característica de interés.

### **3.1. Modelo de calificación**

Tanto en las evaluaciones internacionales como en las nacionales, se suelen emplear metodologías estadísticas para el análisis de datos. Los modelos de la Teoría de Respuesta al Ítem buscan representar la habilidad de cada evaluado, considerada como una variable latente que no es directamente observable. Para la aplicación de estos modelos, se asume que los evaluados con mayor habilidad en un área específica tendrán una mayor probabilidad de responder correctamente a los ítems evaluados.

En el caso de la prueba Saber 3º, 5º, 7º y 9º y en particular en las pruebas que en este se aplican, se optó por usar el modelo logístico de 2 parámetros, conocido también como 2PL. Con este modelo se asume que la relación que existe entre la habilidad y la probabilidad de contestar correctamente se puede explicar mediante 2 parámetros asociados a los ítems, la expresión que define el modelo se puede expresar como:

$$\Pr(X_{ij} = 1 | \theta_j, a_i, b_i) = \frac{1}{1 + e^{-D a_i (\theta_j - b_i)}}$$

donde los subíndices  $j$  e  $i$  corresponden al j-ésimo evaluado y el i-ésimo ítem respectivamente.  $X_{ij}$  es una variable dicótoma que toma el valor 1 cuando el j-ésimo evaluado responde correctamente al i-ésimo ítem y 0 si es incorrecta la respuesta; la constante  $D$  puede tomar los valores de 1 ó 1.702, dependiendo de si se usa la escala logística o la aproximación normal;  $\theta_j$  es el rasgo latente (la habilidad) del individuo  $j$ ;  $a_i$  es el parámetro de discriminación el cual está relacionado con la capacidad del ítem de diferenciar evaluados de distintos niveles de habilidad, y la dificultad del ítem es denotada por  $b_i$ . La habilidad estimada de los evaluados se transforma en puntajes por medio del proceso de escalamiento.

### **3.2. Diseño muestral Saber 3°, 5°, 7° y 9°**

Para poder medir la calidad de la educación para los diferentes grados de interés es necesario realizar una muestra aleatoria de todos los estudiantes de dichos grados a nivel nacional. Para esto desde la Subdirección de Estadística se estableció una metodología que permite seleccionar establecimientos, sedes y estudiantes, y así obtener estimaciones precisas a nivel departamental, desagregado por zona y sector. Se plantea un diseño muestral estratificado por departamento, oferta<sup>6</sup>, zona y sector, dentro de cada estrato se realiza una selección de 2 etapas, en la primera se seleccionan los establecimientos por medio de un diseño  $\pi$ -PT, en la segunda etapa se realiza una selección aleatoria de sedes y dentro de las sedes se hace censo de estudiantes.

Dado el diseño anterior y de acuerdo con el Informe técnico para la selección de la muestra 3°, 5°, 7°, 9°- 2022, la población de estudio para las pruebas del noveno grado corresponde a 10392 establecimientos, 12751 sedes y 702978 estudiantes. La muestra para el grado noveno corresponde a 458 establecimientos en los 33 departamentos y 234 municipios en el territorial nacional, de los cuales 297 se ubican en la zona urbana y 161 en la zona rural, donde para determinar si un establecimiento pertenece a la zona rural o urbana se toma en cuenta la sede principal.

---

<sup>6</sup> Oferta: Establecimientos solo primaria y establecimientos con grados completos, para asegurar la muestra en todos los grados.

### 3.3. Modelo SAE de unidad

La variable de interés  $y$  es el puntaje por establecimiento de la prueba Matemáticas del grado 9º, asumiendo que ésta es una característica fija, se tiene que el modelo de unidad para el puntaje del establecimiento educativo  $i$  dentro del área (municipio)  $d$  corresponde a:

$$y_{di} = \mathbf{x}_{di}'\boldsymbol{\beta} + u_d + e_{di}$$

para  $i = 1, \dots, N_d$ ,  $d = 1, \dots, D$ , donde  $\boldsymbol{\beta}$  es el vector de coeficientes de las variables auxiliares, común para todas las áreas,  $u_d$  es el efecto aleatorio del área  $d$  y  $e_{di}$  es el error a nivel de individuo. Los efectos aleatorios representan la heterogeneidad no explicada de los valores  $y_{di}$  a través de las áreas; estos se consideran independientes de los errores. Adicionalmente se asume que  $u_d \sim_{iid} (0, \sigma_u^2)$  y  $e_{di} \sim_{iid} (0, \sigma_e^2 k_{di}^2)$ , siendo  $k_{di}$  constantes conocidas que representan la posible heteroscedasticidad.

Nuestro interés es predecir el puntaje promedio de  $y$  en las áreas, esto es,  $\bar{y}_d$ , para eso se tiene que el mejor predictor empírico insesgado lineal (EBLUP) está dado por:

$$\widetilde{\bar{y}_d} = \frac{1}{N_d} \left( \sum_{i \in s_d} y_{di} + \sum_{i \in r_d} \tilde{y}_{di}^{BLUP} \right)$$

donde  $s_d$  denota la muestra en el área  $d$ , y  $r_d$  su complemento en el área  $d$ . Y

$$\tilde{y}_{di}^{BLUP} = \mathbf{x}_{di}'\tilde{\boldsymbol{\beta}} + \tilde{u}_d$$

$$\tilde{u}_d = \gamma_d (\bar{y}_{da} - \bar{\mathbf{x}}'_{da} \tilde{\boldsymbol{\beta}})$$

$$\gamma_d = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2/a_d}$$

Donde  $\bar{y}_{da} = a_d^{-1} \sum_{i \in s_d} a_{di}$  y  $\bar{\mathbf{x}}_{da} = a_d^{-1} \sum_{i \in s_d} a_{di} \mathbf{x}_{di}$  son las medias muestrales ponderadas de la variable respuesta  $y$  y las variables auxiliares, respectivamente, con pesos  $a_{di} = k_{di}^2$  y  $a_d = \sum_{i \in s_d} a_{di}$

Por otro lado, el estimador de mínimos cuadrados ponderados de  $\beta$  está dado por

$$\widehat{\beta} = \left( \sum_{d=1}^D X_{ds} V_{ds}^{-1} X'_{ds} \right)^{-1} \sum_{d=1}^D X_{ds} V_{ds}^{-1} y_{ds}$$

donde  $X_{ds}$  es la matriz de variables auxiliares en la muestra del área  $d$ ,  $y_{ds}$  es el vector de la variable observada en la muestra del área  $d$ , y  $V_{ds}$  corresponde a la parte particionada de la muestra de la matriz  $V_d = \sigma_u^2 \mathbf{1}_{N_d} \mathbf{1}'_{N_d} + \sigma_e^2 A_d$  con  $A_d = \text{diag}(k_{di}^2; i = 1, \dots, N_d)$ .

Dado lo anterior, el proceso se centra en el cálculo de los  $\tilde{y}_{di}^{BLUP}$  para todo  $i$  y  $d$ , es decir, la predicción del puntaje para todos los establecimientos educativos de la población que no hayan sido seleccionados en la muestra, lo anterior, combinando con el puntaje de los establecimientos muestrados, permite construir un pseudo-universo donde tenemos el puntaje para todos los establecimientos de la población de estudio. Este pseudo-universo nos permite calcular los puntajes promedios en las áreas de interés, que en primer lugar corresponde a los municipios, pero también permite desagregar a otros niveles combinando características de los establecimientos como zona, sector, entre otros.

### 3.4. Selección del modelo

En esta sección se describe brevemente los pasos que se siguen para encontrar un modelo idóneo para nuestra característica de interés.

Considerando que posiblemente pueden existir relaciones no lineales entre las variables auxiliares y el puntaje de los establecimientos, en primer lugar, se toma el cuadrado a todas las variables auxiliares numéricas, y se aplica un procedimiento tipo *stepwise* para seleccionar las variables con base en el criterio de información Akaike y se verifica que no haya multicolinealidad entre las variables explicativas. Posteriormente se ajusta el modelo de unidad en su versión inicial, esto para poder detectar los establecimientos que se consideran como observaciones atípicas por medio de criterios

como la puntuación  $z$ , el rango intercuartílico y la distancia Mahalanobis. Una vez eliminados los establecimientos atípicos, se ajusta el modelo de unidad definitivo.

Se opta por el modelamiento bayesiano donde se utilizan distribuciones previas no informativas, ya que la estimación de la varianza de predicción es mucho más eficiente en comparación con el modelo frecuentista. Posterior al ajuste, se examina la convergencia del modelo bayesiano observando el valor de R.hat que debe ser cercano a 1 para comprobar que las cadenas hayan convergido, también se observan los valores estimados de los parámetros a través de las iteraciones, así como también sus densidades posteriores estimadas. Luego, se procede a validar el modelo, que consiste en verificar los siguientes supuestos:

- Normalidad de los residuales
- Normalidad de los efectos aleatorios
- Desempeño del modelo para predecir los puntajes para los establecimientos dentro de la muestra, esto se logra comparando (i) los valores de puntajes observados y los puntajes predichos, y (ii) la densidad de los puntajes observados y de los puntajes predichos.

### **3.5. Predicciones**

Una vez el modelo esté validado, se procede a predecir los valores del puntaje promedio para todos los establecimientos que no hayan sido seleccionados en la muestra, los cuales, en combinación con los valores muestrales, constituyen el pseudo-universo de los establecimientos del país. Dado que el modelo de unidad se estimó dentro del contexto bayesiano, se debe obtener la esperanza de la distribución predictiva del puntaje de cada establecimiento. Siguiendo la metodología estándar de simulación Monte Carlo para obtener la distribución predictiva, se calcula la predicción de los puntajes de los establecimientos para cada conjunto de parámetros simulados de su distribución posterior. Suponiendo que de las distribuciones posteriores se extraen  $G$  valores para cada parámetro, para cada establecimiento se tendrán  $G$  predicciones, y estas predicciones son realizaciones de la distribución predictiva del puntaje del

establecimiento. De esta forma, el pseudo-universo de los establecimientos tendrán  $G$  réplicas en forma de una matriz de tamaño  $N \times G$ , donde  $N$  denota el número de establecimientos en la población, esta matriz la denotaremos como  $[y_{m,g}]$  con  $g = 1, \dots, G$  y  $m = 1, \dots, N$ .

El cálculo de agregados en diferentes niveles, incluyendo los municipios, se lleva a cabo por medio de promedios ponderados por el número de estudiantes del grado noveno, así: para el municipio  $d$  y para cada iteración  $g$ , el puntaje Saber 9º Matemáticas promedio se calcula como:

$$y_{d,g} = \frac{\sum_{m \in d} y_{m,g} * n_m}{\sum_{m \in d} n_m}$$

donde  $y_{m,g}$  es el puntaje del establecimiento  $m$  en la iteración  $g$  y  $n_m$  es el número de estudiantes del noveno grado del establecimiento  $m$ . De esta forma, para cada municipio, se tiene  $G$  valores de puntaje promedio, calculando el promedio y la varianza sobre estos  $G$  valores permite obtener finalmente la predicción del puntaje promedio para el dominio  $d$  y su error de predicción.

### **3.6. Benchmarking**

Ahora, los agregados que se calculan con el pseudo-universo generado para los niveles de desagregación donde la encuesta sí tenga representatividad, esto es, a nivel de región, zona, sector y departamental, no coincidirán con las cifras oficiales publicadas por el Icfes en el Informe Nacional de los Resultados de las Pruebas Saber 3º, 5º, 7º y 9º -2022, por lo tanto, se debe realizar un ajuste adicional llamado *Benchmarking* que consiste en hacer uso de la calibración para forzar que los agregados coincidan con los agregados publicados oficialmente. Esto se realiza directamente a nivel de los establecimientos, es decir, al pseudo-universo descrito en el paso anterior. Una vez se termine el proceso de *Benchmarking*, se puede calcular los agregados para los dominios de interés tal como se describe en la subsección anterior. Cabe destacar que el paso anterior también constituye un procedimiento para validar los resultados de predicción,

ya que los agregados calculados con el pseudo-universo no deben estar muy alejados de los agregados oficiales publicados.

### **3.7. Incorporando la variabilidad de los puntajes**

Ahora, tal como se mencionó anteriormente, los puntajes de los establecimientos no corresponden a observaciones directas, sino que son productos de un modelo de Teoría de Respuesta al Ítem y un diseño muestral probabilístico, por lo tanto, el anterior proceso de ajuste del modelo y la generación del pseudo-universo debe ser replicada para poder incorporar la variabilidad de los puntajes. Para eso debemos tener en cuenta que para cada establecimiento  $m$ , el puntaje  $Y_m$  es el resultado de dos procesos generadores de datos, a continuación se describe el proceso para incorporar ambos procesos en el análisis.

En primer lugar, del proceso de ajuste del modelo de Teoría Respuesta al Ítem a partir de datos que provienen de un diseño muestral complejo, tenemos que  $Y_m \sim N(\mu_m, \tau_m^2)$ , denotaremos con  $P_1$  la medida de probabilidad asociada a este proceso. Esta fuente de variabilidad se tiene en cuenta al simular  $J$  valores de esta distribución normal para cada establecimiento muestreado, y así conformar  $J$  pseudo-muestras.

Por otro lado, para cada pseudo-muestra se ajusta el modelo de unidad presentado anteriormente, y se obtienen  $G$  réplicas para cada establecimiento no muestreado, denotaremos con  $P_2$  la medida de probabilidad asociada a este proceso, (cabe recordar que, para los establecimientos seleccionados en la muestra, el puntaje de las  $G$  réplicas será constante igual al puntaje calculado). De esta forma, para cada una de ellas se ajusta el modelo de unidad y se predice el puntaje promedio para cada establecimiento no seleccionado en la muestra. Teniendo en cuenta que se opta por el enfoque del modelamiento bayesiano, los resultados obtenidos corresponderán a  $J$  pseudo universos para todos los establecimientos con  $G$  réplicas:  $[y_{m,g}^{(j)}]_{g=1,\dots,G, j=1,\dots,J, m=1,\dots,N}$ . Finalmente,  $\xi_{obs}$  denotará toda la información observada, y definimos  $\mu_{m,j} = \frac{1}{G} \sum_{g=1}^G y_{m,g}^{(j)}$  y  $\sigma_{m,j}^2 = \frac{1}{G-1} \sum_{g=1}^G (y_{m,g}^{(j)} - \mu_j)^2$ .

De esta forma, nuestro objetivo consiste en encontrar  $E(Y_m)$  y  $Var(Y_m)$ , los cuales se puede obtener haciendo uso de los momentos condicionales:

$$E(Y_m) = E_1(E_2(Y_m | \xi_{obs})) \approx \frac{1}{J} \sum_{j=1}^J \mu_{m,j} = \bar{\mu}_m$$

$$Var(Y_m) = E_1(Var_2(Y_m | \xi_{obs})) + Var_1(E_2(Y_m | \xi_{obs})) \approx \frac{1}{J} \sum_{j=1}^J \sigma^2_{m,j} + \left(1 + \frac{1}{J}\right) \frac{1}{J-1} \sum_{j=1}^J (\mu_{m,j} - \bar{\mu}_m)^2$$

Así, la predicción del puntaje del establecimiento  $m$  está dada por  $E(Y_m)$ , y la medida de precisión viene dada por  $Var(Y_m)$  o su raíz cuadrática.

## 4. Información y datos

Teniendo en cuenta que se ajusta un modelo SAE de unidad, se requiere que los establecimientos cuenten con información auxiliar disponible, sin que sea necesario realizar filtros previos para la estimación del modelo. A continuación, se describen las principales fuentes de información.

### 4.1. Información de EDUC

Una de las fuentes estadísticas más importantes en el país es el Departamento Administrativo Nacional de Estadística (DANE). En términos de estadísticas de educación, en la página <https://microdatos.dane.gov.co/index.php/catalog/EDU-Microdatos> se encuentran los indicadores de educación generados por el DANE, donde se reportan los establecimientos oficiales y privados de educación formal que existen en el país. En particular, en este trabajo, se utilizaron los datos de Educación Formal 2021, disponibles para descargar en la página <https://microdatos.dane.gov.co/index.php/catalog/752>. Esta información corresponde al censo de establecimientos educativos que ofrecen educación básica formal en preescolar, básica primaria, básica secundaria y media. Se incluye también Ciclos Lectivos Especiales Integrados (CLEI) y modelos educativos flexibles. Dicha información se denominará EDUC en este documento.

La fuente EDUC contiene datos de diferentes aspectos de máximo 56.349 sedes educativas del país, y contiene información sobre: (i) indicadores de matrícula (número de estudiantes de diferentes características tales como sexo, raza, grupo etario, etc.), (ii) indicadores de docente (número de docentes según escalafón, máximo nivel educativo, sexo, etc.), (iii) indicadores de desempeño académico (tasa de aprobación y reprobación, etc.), (iv) indicadores relacionados con el TIC, y finalmente (v) información sobre jornada, ubicación, horario, etc. Estas variables a nivel de sedes se agregaron convenientemente para los establecimientos educativos.

Después de evaluar detalladamente la información disponible en EDUC, se decide considerar las siguientes variables pues se esperan que afecten en el rendimiento de los resultados de la prueba Saber: región, zona, sector, departamento, porcentaje de estudiantes indígena / desplazados / con discapacidad / con acceso a computador, información de los docentes según tipo de vinculación y nivel educativo, y características propias de los establecimientos.

#### **4.2. Información de SIMAT**

El Sistema de Matrícula Estudiantil (SIMAT) del Ministerio de Educación Nacional es el sistema integrado de matrícula que permite organizar y controlar el proceso de matrícula en todas sus etapas, así como tener una fuente de información confiable y disponible para la toma de decisiones. Por lo tanto, se extraen para cada establecimiento la información de: número de sedes, porcentaje de estudiantes femeninos / masculinos, porcentaje de estudiantes en cada uno de los 6 estratos socioeconómicos, edad promedio y el número de estudiantes.

##### **1.1. Información de otras pruebas Saber**

Otra fuente de información proviene del mismo Icfes, ya que se cuenta con resultados de la prueba Saber 11, que es una prueba censal, es decir, una prueba que se aplica a todos los estudiantes de todos los establecimientos. En este caso utilizamos los puntajes de Saber 11 correspondiente al año 2022.

#### **4.3. Información de municipios**

Teniendo en cuenta que las áreas de interés para esta investigación son los municipios, es importante incluir información a nivel de estas áreas, para eso se hizo uso de las siguientes fuentes de información: Censo Nacional de Población y Vivienda 2018, específicamente se toma la información a nivel de hogares y se calculan los agregados a nivel municipal. La información utilizada corresponde a condiciones de los hogares en temas de acceso a servicios públicos, condiciones de la vivienda, conformación del hogar, etc. Adicionalmente, se recurrió a las siguientes entidades para obtener información adicional que contribuya a lograr un mejor ajuste del modelo: DANE (composición demográfica), Evaluaciones Agropecuarias Municipales (oferta productiva agropecuaria), Instituto Nacional de Vías (INVIAS) del año 2022, Índice de Riesgo de Calidad de Agua para el Consumo Humano (IRCA) del 2018, Medición de Desempeño Municipal (MDM) del 2021, SHP del 2022, Sistema Integral de Información de la Protección Social (SISPRO) del 2019, Unidad de Planificación Rural Agropecuaria (UPRA) del 2022.

#### **4.4. Consolidación de la información**

Consolidando la información descrita anteriormente, se cuenta en total 147 variables clasificadas así: marco muestral (9 variables), EDUC (23 variables), Saber 11 (5 variables), SIMAT (13 variables), información de municipios (97 variables).

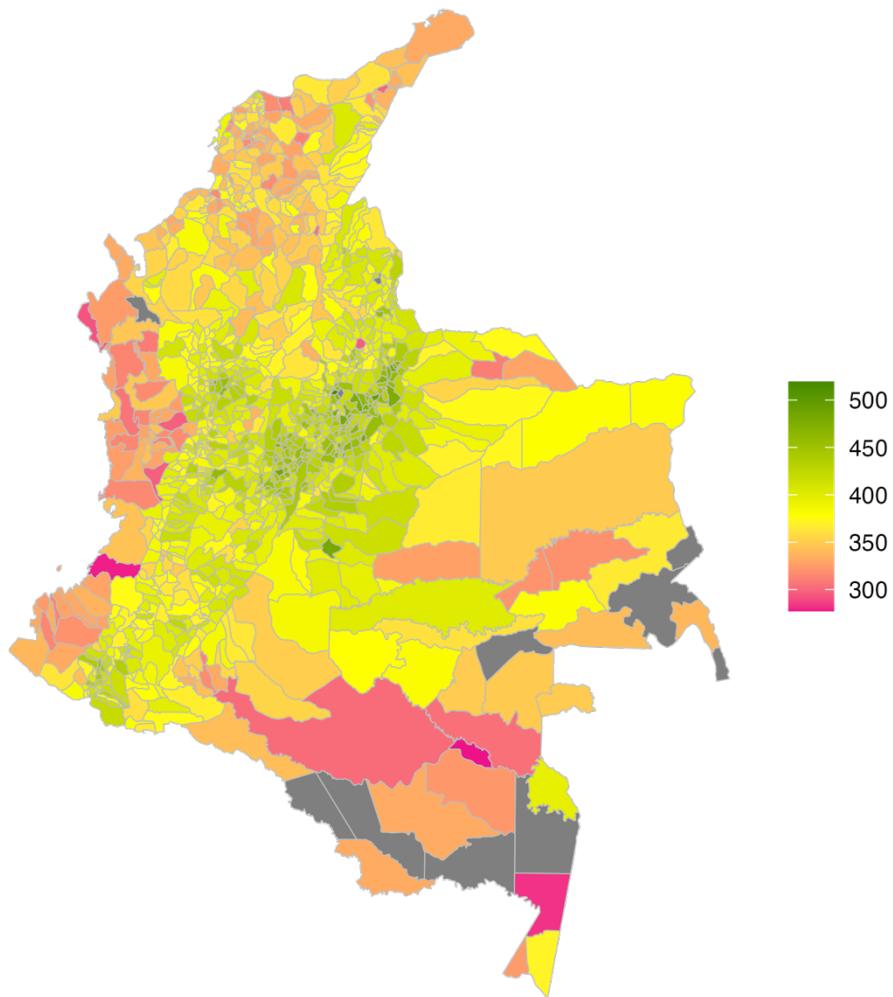
En términos de la información de la muestra de establecimientos, inicialmente se cuenta con 458 establecimientos, pero 17 de ellos no tienen información en algunas de las variables auxiliares, por lo que el modelo se ejecutará para los 441 establecimientos con información completa.

### **5. Resultados**

Es importante tener en cuenta que se debe revisar medidas de calidad de las predicciones de los puntajes, teniendo en cuenta que el puntaje es una característica continua, se utiliza el coeficiente de variación definido como el error estándar dividido por el puntaje.

A continuación se muestra los resultados de los puntajes de municipios donde el coeficiente de variación es menor al 30%, ya que valores más grandes implican que las cifras no son confiables. De esta forma, se cuentan en total predicciones con coeficientes de variación menores al 30% para 1111 municipios.

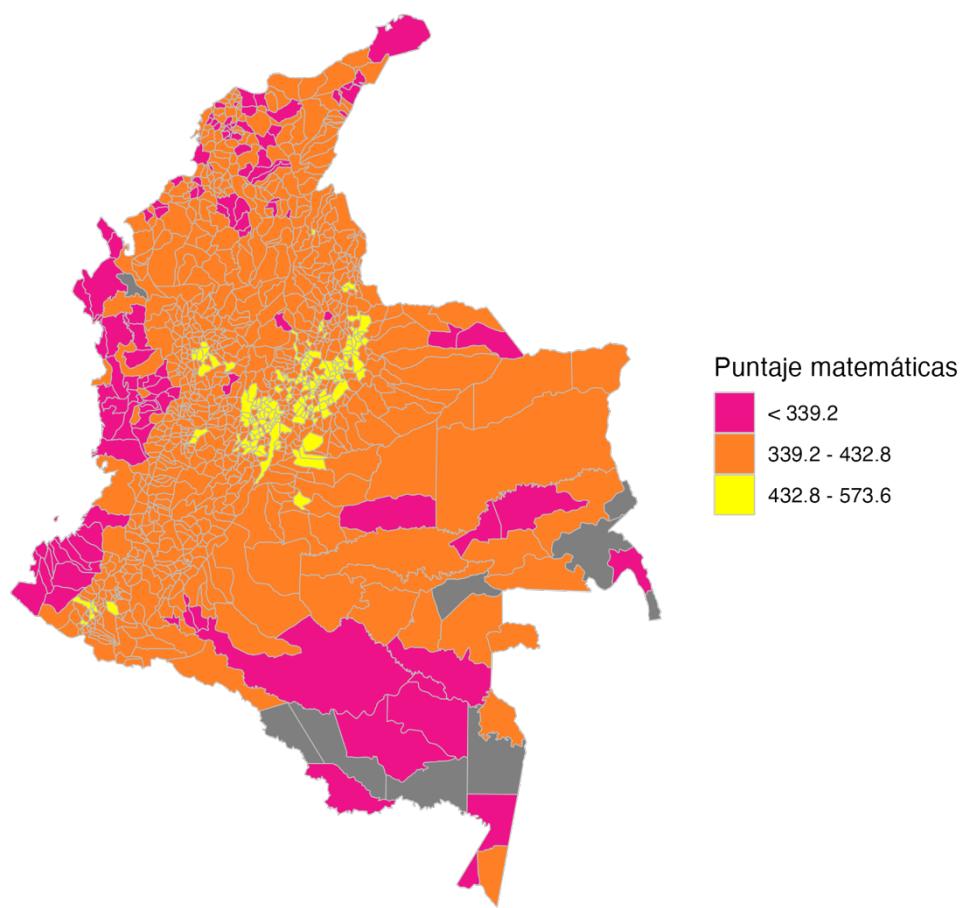
**Figura 1. Predicción de puntaje promedio de la prueba Saber 9° Matemáticas**



Fuente: Elaboración propia.

También se puede observar los diferentes niveles de desempeño para los municipios del país, tomando los puntos de corte 339.2, 432.8, 573.6 para el puntaje promedio de establecimientos para la prueba de Matemáticas del grado 9.

**Figura 2. Predicción de nivel de desempeño de la prueba Saber 9°  
Matemáticas**

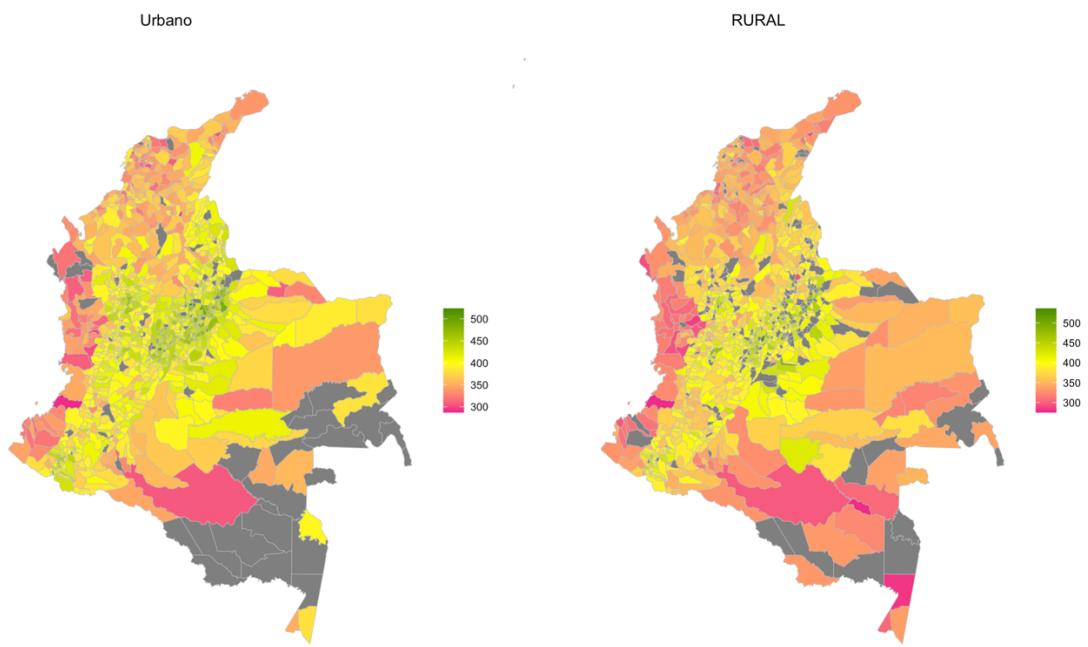


Fuente: Elaboración propia.

Adicionalmente, al contar con predicciones de puntaje a nivel de establecimientos, se puede obtener un panorama de los municipios solo considerando establecimientos de determinadas características, por ejemplo, en la siguiente gráfica

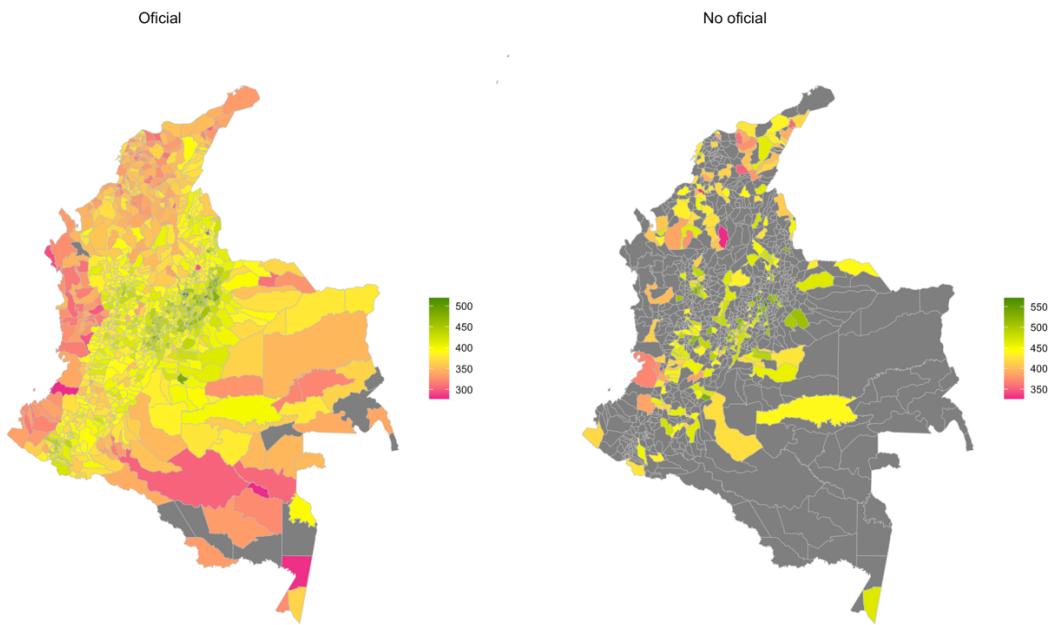
se muestran los puntajes promedios de los establecimientos por municipio según si pertenece a zona rural o urbana.

**Figura 3. Predicción de puntaje promedio de la prueba Saber 9° Matemáticas por zona**



De forma similar en la siguiente gráfica se muestran los puntajes promedios de establecimientos oficiales / no oficiales por municipio.

**Figura 4. Predicción de puntaje promedio de la prueba Saber 9° Matemáticas por sector**



## 6. Conclusiones

En este trabajo se presenta la metodología de estimación de áreas pequeñas para obtener la predicción del puntaje de Matemáticas para el grado noveno. Este mismo procedimiento se puede aplicar a las demás pruebas, como Lenguaje y Ciencias. En general, la metodología propuesta es adecuada para la estimación de puntajes que provienen de modelos de Teoría de Respuesta al Ítem, estimados a partir de datos recolectados mediante un diseño muestral complejo.

La aplicación de esta metodología en el contexto y bajo las características de las evaluaciones educativas estandarizadas a gran escala resulta novedosa y relevante, ya que permite generar información para dominios donde no se realizó aplicación muestral directa. Este enfoque amplía el alcance de los resultados y complementa las mediciones tradicionales. No obstante, la metodología no pretende por el momento generar

estadísticas oficiales, ni reemplazar las aplicaciones territoriales, que ofrecen información más detallada de agregados específicos y con mayor precisión, y cuya riqueza no siempre puede ser capturada por las covariables incluidas en el modelo de estimación.

En este sentido, la implementación del modelo de unidad constituye un primer paso hacia el fortalecimiento de estrategias de estimación de áreas pequeñas en el ámbito de la evaluación educativa. Su uso abre la posibilidad de profundizar en metodologías que contribuyan a mejorar la disponibilidad de información para la toma de decisiones de política pública. Asimismo, queda abierta la oportunidad de explorar otros modelos y enfoques que puedan complementar los resultados obtenidos y potenciar su utilidad.

### **Agradecimientos**

Los autores agradecen al Icfes por proporcionar el acceso a los datos de las pruebas Saber 3°, 5°, 7° y 9° y Saber 11°. Este proyecto contó con el apoyo del equipo de la Subdirección de Estadísticas del Icfes, el cual, en cumplimiento de su misión institucional, desarrolla proyectos de investigación en métodos psicométricos y estadísticos. Asimismo, los autores agradecen a Cristian Montaño y Alexander Calderón, por los valiosos comentarios realizados sobre versiones previas del manuscrito.

## Referencias

**Casas-Cordero Valencia, C., Encina, J. and Lahiri, P. (2016).** Poverty Mapping for the Chilean Comunas, In M. Pratesi (Ed.), Analysis of Poverty Data by Small Area Estimation, New York: Wiley.

**DANE - Departamento Administrativo Nacional de Estadística. (2024).** Boletín técnico - Inseguridad alimentaria a partir de la escala FIES - 2023.

**Fay, R.E. y Herriot, R.A. (1979).** Estimation of Income for Small Places: An Application of James-Stein Procedures to Census Data, Journal of the American Statistical Association, 74, 269–277.

**Icfes – Instituto Colombiano para la Evaluación de la Educación (2022).** Informe Nacional de los Resultados de las Pruebas Saber 3º, 5º, 7º y 9º -2022.  
<https://www.icfes.gov.co/web/guest/informe-nacional-2022>

**Icfes – Instituto Colombiano para la Evaluación de la Educación (2022).** Informe técnico para la selección de la muestra 3º, 5º, 7º, 9º - 2022, lápiz y papel. Subdirección de Estadísticas.

**Isabel Molina (2019).** Desagregación de datos en encuestas de hogares: metodologías de estimación en áreas pequeñas, Series Estudios Estadísticos, No 97, (LC/TS.2018/82/Rev.1), Santiago, Comisión Económica para América Latina y el Caribe, (CEPAL)

**Sutton, A. M., & Almquist, Z. W. (2023).** Small Area Estimation for Demographic Analysis. IntechOpen. doi: 10.5772/intechopen.1001842