

Comparación entre los modelos de unidad y área para la estimación en áreas pequeñas del puntaje de matemáticas en las entidades territoriales certificadas para la prueba Saber 11

Comparison between area-level and unit-level models for the estimation in the small areas of the mathematics score of the Saber 11 in the territorial tentities certified for the test Saber 11

José Andrés Flórez Gutiérrez^a
joseflorez@usantotomas.edu.co

Hugo Andrés Gutiérrez Rojas^b
hagutierrezro@gmail.com

Resumen

En los últimos años se ha evidenciado la necesidad de tener información confiable posterior a la aplicación de encuestas, sin embargo, dado este procedimiento hay áreas que no son posibles de representar debido a que su tamaño de muestra es reducido o nulo, generando una limitación a la hora de desarrollar análisis de dicha información. La estimación en áreas pequeñas es capaz de generar resultados para este tipo de áreas con el soporte de información auxiliar. Entre los tipos de información auxiliar se encuentran registros administrativos o datos históricos. Con el objetivo de obtener estimaciones con coeficientes de variación aceptables. En este artículo se desarrolló una aplicación de los modelos a nivel de área e individuo con los que se estimaron el puntaje promedio de la prueba de matemáticas a nivel de Entidades Territoriales Certificadas (ETC), de la PRUEBA SABER 11 del año 2017 realizada por el Instituto Colombiano para la Evaluación de la Educación (Icfes).

Palabras clave: Estimación en áreas pequeñas, Fay-Herriot, Battese-Harter-Fuller, estadísticas oficiales, evaluación a gran escala en educación.

Abstract

In the last few years, we have evidenced a necessity to obtain reliable information after the application of surveys, however in this process there are areas that are not possible to represent because they have small or null sample size wich generates a limitation at the moment to develop analysis of the information. For this areas small area estimation is capable to generate results with the support of auxiliary information, found in administrative records or history data. With the objective to obtain acceptable coefficients of variation. In this article we developed an application of area-level and unit-level models where it was estimated the average of the score in the test of mathematics at the level of Entidades Territoriales Certificadas (ETC), at the PRUEBA SABER 11 done by Instituto Colombiano para la Evaluación de la Educación (Icfes) in 2017 year.

Keywords: Small Area Estimation, Fay-Herriot, Battese-Harter-Fuller, officials statistics, Large-scale Assessments in Education.

^aEstudiante

^bDirector

Introducción

La información que se obtiene a partir de la utilización de herramientas estadísticas, es una base fundamental para el planteamiento, ejecución y evaluación de las políticas públicas de acuerdo con el Ministerio de Educación Nacional, (MEN 2014). Esta información permite comprender la realidad que tiene el país actualmente en el campo de la evaluación de la educación, por este motivo se debe monitorear constantemente la prestación de este servicio a través de instrumentos que permitan conocer el desempeño de los estudiantes y observar el proceso que se tiene al aplicarlos. Por este motivo, en los años sesenta y de acuerdo a los lineamientos del gobierno en esta época, se crea una entidad pública comprometida con el fomento de la educación y su evaluación, para medir el comportamiento de la educación.

Así es como el Icfes, una entidad adscrita al Ministerio de Educación Nacional, ofrece los servicios para evaluar la educación en todos los niveles, a través de exámenes de estado e investigaciones de factores asociados que puedan incidir en la calidad educativa, desde las instituciones hasta a los estudiantes, que puedan generar un valor agregado en la generación de nuevas políticas de educación. Entre las pruebas de estado desarrolladas por el Icfes actualmente se encuentran la prueba SABER 3, 5 y 9, la prueba SABER 11, la prueba SABER TyT (Técnicos y Tecnólogos) y la prueba SABER PRO, entre otras.

Este artículo se centra en la utilización de los resultados de la prueba SABER 11, aplicada en el segundo período de los años 2014, 2015, 2016 y 2017. La cual se aplica a los estudiantes que pertenecen a las instituciones educativas de calendario A y que se encuentran finalizando su grado 11, a parte de aquellas estudiantes que se presentan de forma individual. La información utilizada, (bases de datos de los puntajes anonimizados), se encuentra de manera pública y abierta, en forma de bases de datos para investigadores asociados y no asociados del Icfes, y que se pueden encontrar en el sitio web, (ICFES 2010). No se tuvo en cuenta los colegios pertenecientes al calendario B, dado que sus resultados representan una proporción reducida de la situación que actualmente tiene la educación en el país, al ser bilingües de base y su con puntajes medios muy altos que difieren de la media del puntaje total comparado con las instituciones de calendario A, puesto que la mayoría de estos colegios son de carácter privado y de mayor nivel económico.

La prueba SABER 11 se compone de cinco pruebas genéricas y 2 específicas que evalúan las competencias desarrolladas por los estudiantes y la deben presentar como requisito de grado todos los estudiantes de grado once, tanto aquellos que cumplen con su año escolar como los estudiantes que se encuentran en un proceso de validación, o que ya hayan obtenido el grado de bachiller (estudiantes individuales).

Los resultados de esta prueba se encuentran a nivel individual por cada estudiante, sin embargo, tiene la ventaja de ofrecer la capacidad de agregarse en diferentes niveles como lo son: establecimientos, municipios, departamentos y Entidades Territoriales Certificadas (ETC).

La agergación de interés para este artículo son las ETC, las cuales tienen como función conforme a lo establecido por la ley 715 de 2001 en MEN (2015), la competencia de administrar el servicio educativo en su jurisdicción, lo cual se fundamenta en una adecuada prestación del servicio, cumpliendo tres condiciones especiales: cobertura, calidad y eficiencia. Las ETC deben planificar, organizar y distribuir de forma clara todo tipo de recursos humanos y financieros garantizando transparencia, para que un mayor número de niños y jóvenes puedan acceder a la educación y así fortalecer la cobertura en todo el país.

Teniendo en cuenta la definición anterior de las ETC, obtener resultados a este nivel nos permite medir y diseñar políticas públicas con el objetivo de garantizar el acceso a la educación, sin embargo, actualmente dichas decisiones se pueden ver afectadas por tamaños reducidos de población en las divisiones con las que cuentan las 95 ETC (departamentos, distritos, municipios y territorios indígenas) a nivel nacional. Esto genera una baja confiabilidad en los resultados de dichas ETC. Cubrir esta limitante implicaría realizar un despliegue de recursos más alto y extra, tanto humano como económico, que no necesariamente se dispone en los recursos públicos después de cada aplicación de la prueba.

Al igual que la situación anterior, hay ocasiones donde no es posible obtener estadísticas confiables debido a la limitación de la disponibilidad de datos en el estudio, siendo allí donde la estimación en áreas pequeñas suplen esta dificultad. Según Hidirglou (2016) algunos ejemplos de los dominios pueden

ser regiones geográficas entre las que se incluyen provincias o municipios, además de grupos a nivel demográfico, (género o raza). En este artículo se definieron como dominios las ETC dentro de las cuales se encontraron dominios planificados y no planificados. El lector debe tener en cuenta que en los dominios no planificados no es posible obtener estimaciones directas fiables según Garcia (n.d.), debido a su tamaño de muestra reducido.

De acuerdo con Ferreira (2011) un dominio planificado es aquel que se puede identificar de manera previa al diseño de la muestra, lo que nos permite calcular el tamaño de muestra y poder controlarlo porque es contenido en el marco de muestreo. Un dominio no planificado no tuvo una definición en el diseño muestral, ya sea por falta de información o conocimiento al momento de realizar la muestra, con la cual se puedan identificar los individuos dentro de los dominios porque son requeridos de manera posterior a la aplicación de la muestra.

En vista de este inconveniente una posible solución es aumentar los tamaños poblacionales en los dominios no planificados, pero esto tiene un costo muy alto para conseguir la información faltante; por otro lado, también se podría cambiar el diseño muestral a fin de transformar los dominios no planificados en planificados. Por este motivo para realizar los análisis respectivos de las estimaciones se deben emplear estimadores indirectos en lugar de los estimadores directos, ya que si estos se utilizarán en la estimación para los dominios no planificados se obtendría “estimadores con una varianza muy grande” (Bouzas 2012).

Como medida de ajuste en el desarrollo de este artículo se utilizó el error cuadrático medio, MSE (por sus siglas en inglés), para diferenciar las estimaciones con errores más pequeños y adicionalmente el uso de modelos que generen una inclusión de variables que contengan información auxiliar de datos de tipo transversal, temporal o de relación espacial para generar las estimaciones en los dominios no planificados.

Considerando las definiciones anteriores se tuvieron en cuenta para la estimación en áreas pequeñas (SAE, *Small Area Estimation*) por sus siglas en inglés, los métodos basados en los modelos, que toman los datos que son generados por un modelo para apoyar las inferencias a partir de estos resultados. Adicionalmente una ventaja con la que se cuenta en el SAE es la posibilidad de utilizar información auxiliar, con el fin de optimizar los resultados basados en los modelos siendo más precisos a través de realizar una selección y diagnóstico del modelo adecuado. Al ajustar el modelo indicado con la información auxiliar el SAE nos permite, proporcionar estimadores con errores cuadráticos medios más pequeños en comparación con los que se basan en el diseño muestral.

El objetivo de este trabajo es estimar la media del puntaje de la prueba de matemáticas de 2017 para las ETC en Colombia a través del estimador directo propuesto por (Horvitz & Thompson 1952), y los estimadores indirectos que son el estimador de (Fay & Herriot 1979) y el estimador de (Battese et al. 1988). De esta forma comparar su eficiencia al tener dominios con tamaños muestrales reducidos a través de sus errores cuadráticos medios y sus coeficientes de variación, los cuales se desarrollarán más adelante en el artículo.

En la sección 1 se hace una referencia al marco teórico que incluye la descripción de cada uno de los estimadores; en la sección 2 se tiene en cuenta la presentación de la descripción de la información; la sección 3 presenta los análisis y resultados para los modelos utilizados; la sección 4 detalla las conclusiones obtenidas, la discusión a partir de los resultados y las investigaciones futuras.

1. Marco Teórico

1.1. Estimador Directo Horvitz-Thompson

En Horvitz & Thompson (1952) se introduce un estimador insesgado, denotado en la literatura como π , para un universo U , donde se quiere estimar el total poblacional t_y y la media poblacional $\bar{Y}_d$, del dominio. Para la media el cálculo se define como la sumatoria de todos los valores que tiene la variable objetivo multiplicados por cada uno de sus pesos muestrales. En los cuales, tendremos que cada subconjunto s de la población tiene una probabilidad de inclusión de $p(s)$ bajo un diseño de muestreo probabilístico

determinado.

El estimador insesgado para el total poblacional es:

$$\hat{t}_{y\pi} = \sum_{k=1}^n \frac{y_k}{\pi_k} = \sum_{k=1}^n w_k y_k, \quad (1)$$

dónde π_k es la probabilidad de inclusión del k-ésimo elemento y w_k hace referencia al factor de expansión o peso de muestreo, el cual corresponde al inverso de la probabilidad de inclusión (Gutierrez 2009).

De otra parte el estimador para la media se escribe de la siguiente forma:

$$\widehat{\bar{Y}}^{dir} = \frac{1}{\widehat{N}} \sum_{k=1}^n w_k y_k \quad \text{donde,} \quad \widehat{N} = \sum_{k=1}^n w_k. \quad (2)$$

El estimador directo, en la ecuación (2), utiliza solamente la información del dominio y período considerados al momento del estudio. Es un estimador centrado respecto de la distribución que se tiene en el diseño muestral pero con un incremento de su varianza en las áreas pequeñas.

Formalmente, los dominios d , en la estimación de áreas pequeñas se definen como tamaños de muestra pequeños o nulos, que se clasifican como sub-grupos poblaciones que contienen características similares y que no se hayan identificado previamente en el diseño de muestreo, pueden ser geográficos o demográficos como se explicó en la introducción. Para los cuales tenemos estimación del total y de la media, teniendo en cuenta a Zea & Ortiz (2018):

$$\hat{t}_{y\pi,d} = \sum_{k \in s_d}^s \frac{y_k z_{dk}}{\pi_k} = \sum_{k \in s_d}^s w_k y_k z_{dk}, \quad (3)$$

dónde

$$z_{dk} = \begin{cases} 1, & k \in d; \\ 0, & k \notin d. \end{cases}$$

De igual forma también hay una estimación para la media de los dominios expresada de la siguiente forma:

$$\widehat{\bar{Y}}_d = \frac{\hat{t}_{y\pi,d}}{\widehat{N}_d} = \frac{\sum_{k \in s_d}^s \frac{y_{dk}}{\pi_k}}{\sum_{k \in s_d}^s \frac{y_{dk}}{\pi_k}} = \frac{\sum_{k \in s_d}^s w_k y_{dk}}{\sum_{k \in s_d}^s w_k z_{dk}} \quad \text{donde} \quad y_{dk} = y_k z_{dk}. \quad (4)$$

Así como se describe en la ecuación (4), la media del estimador directo $\widehat{\bar{Y}}_d$, es posible definir el estimador insesgado de la varianza para el estimador de Horvitz-Thompson de $\widehat{\bar{Y}}_d$ teniendo en cuenta a Molina (2018), donde se indica que las probabilidades de inclusión deben ser estrictamente mayores a cero:

$$\widehat{var}_{\pi}(\widehat{\bar{Y}}_d) = N_d^{-2} \left\{ \sum_{k \in s_d}^s \frac{Y_{dk}^2}{\pi_{dk}^2} (1 - \pi_{dk}) + 2 \sum_{k \in s_d}^s \sum_{j \in s_d}^s \frac{Y_{dk} Y_{dj}}{\pi_{dk} \pi_{dj}} \left(\frac{\pi_{d,kj} - \pi_{dk} \pi_{dj}}{\pi_{d,kj}} \right) \right\} \quad (5)$$

Al momento de estimar la variable de interés si no se cuenta con la totalidad de información en el diseño muestral, si no por el contrario únicamente con los pesos muestrales w_{dk} y las probabilidades de inclusión de segundo orden $\pi_{d,kj}$, el estimador de la ecuación (5) no es posible de calcular.

Sin embargo, para los diseños de muestreo con probabilidades de inclusión de segundo orden $\pi_{d,kj} \approx \pi_{d,k}\pi_{d,j}$ para $j \neq k$, como lo indica Molina (2018), donde se cumple la igualdad y el segundo término en la ecuación (5), tiende a ser aproximadamente cero. Reemplazando los pesos muestrales $w_{di} = \pi_{di}^{-1}$ se obtiene el estimador de la varianza que no depende de las probabilidades de inclusión de segundo orden:

$$\widehat{var}_{\pi}(\widehat{Y}_d) = N_d^{-2} \sum_{k \in s_d}^s w_{dk}(w_{dk} - 1)Y_{dk}^2. \quad (6)$$

El estimador de Horvitz-Thompson realiza las ponderaciones individuales de la variable de interés Y_{dk} , a través de los pesos muestrales o inversos de probabilidades de inclusión en la muestra, $w_{dk} = \pi_{dk}^{-1}$, dado este proceso, es posible mitigar las selecciones muestrales en donde la probabilidad de seleccionar un individuo que tenga una alta relación respecto a la variable de interés sea alta, puesto que podría aparecer más frecuentemente en la muestra final teniendo un sesgo en la información seleccionada y en los análisis seleccionados; surge el mismo efecto en los individuos que cuentan con una baja representatividad en el diseño de muestreo puesto que probablemente escaseen en la muestra final.

Esto nos daría por resultado que, si realizamos las estimaciones dando el mismo peso a todas las observaciones de la muestra, de igual forma que en la media muestral básica, tendríamos un sesgo (infra-estimariamos o subestimariamos la variable de interés). Conociendo esta situación, es necesario brindar un menor peso de muestreo a las observaciones que cuentan con mayor probabilidad de selección de aparecer en la muestra y mayor peso a aquellas que tienen una menor probabilidad de estar en el diseño de muestreo.

Siendo el estimador de Horvitz-Thompson insesgado en el diseño muestral, su varianza puede dar resultados muy grandes cuando se tienen áreas o dominios pequeños; en este sentido se deben aplicar otro tipo de estimadores que cuenten con un menor sesgo y brinden confianza en sus estimaciones así como lo realiza el estimador de Hájek.

1.2. Estimador de Fay-Herriot: modelo a nivel de área

El modelo de Fay-Herriot (FH), es un modelo a nivel de área que fue introducido por Fay & Herriot (1979), para estimar los ingresos per cápita en áreas pequeñas de Estados Unidos. Este modelo en la actualidad es utilizado por la Oficina del Censo de EE. UU. (U.S. Census Bureau) en el programa "Small Area Income and Poverty Estimates (SAIPE)", en donde se realizan estimaciones de niños en estado de pobreza en edad escolar en sus respectivos condados. El modelo de FH nos permite estimar medias y totales de nuestra variable de interés en las áreas pequeñas, para generar estimaciones precisas cuando las estimaciones directas del estimador de HT no son confiables debido al tamaño reducido de las áreas.

Los procedimientos SAE basados en los modelos de área ajustan un modelo de datos en donde se denota a d , como área pequeña (dominio) $d = 1, \dots, D$, los x_i son las variables predictoras que provienen de la información auxiliar y está representada por un término lineal no aleatorio, los u_d son los efectos aleatorios del área, y e_i son errores aleatorios asociados al modelo. Es decir que $\bar{y}_d$ es la indicadora de interés para toda las áreas $d = 1, \dots, D$ y se asume que $x_d = (x_{d1}, \dots, x_{dp})'$ es el vector de valores para cada una de las p variables auxiliares del área d .

$$\bar{y}_d = x_d^t \beta + u_d + e_d \quad \text{con} \quad u_d \sim N(0, \sigma_u^2) \quad \text{y} \quad e_d \sim N(0, \psi_d), \quad (7)$$

donde β es un vector de dimensión p de coeficientes de regresión del modelo de las áreas d . El término u_d es el error modelo, también conocido como efecto aleatorio en cada una de las áreas, estos efectos

aleatorios u_d , representan la heterogeneidad de los indicadores de y_d a través de las áreas, que no es explicada por las variables auxiliares consideradas. En el modelo de regresión estos efectos aleatorios son independientes e idénticamente distribuidos (*iid*), con varianza σ_u^2 desconocida e independiente de e_d .

Donde e_d es el error de muestreo en el área d , el cual es independiente entre sí y también son independientes de los efectos aleatorios de las áreas, u_d , con media cero y varianza conocida ψ_d , además en la práctica dichas varianzas $\psi_d = \text{var}_\pi(\hat{y}_d^{DIR}|y_d)$ esta varianza es estimada con los datos muestreados. Por lo tanto, la media desconocida para el área d es $\theta_d = x_d^t \beta + u_d$.

El modelo de la ecuación (7) se obtuvo a partir de definir $\hat{y}_d^{DIR} = y_d + e_d$ con $d = 1, \dots, D$ y $y_d = x_d^t \beta + u_d$ con $d = 1, \dots, D$, con la definición del modelo y utilizando el método de multiplicadores de Lagrange, el estimador lineal insesgado de $\hat{y}_d$, $d = 1, \dots, D$ es insesgado bajo el modelo (7) minimiza el error cuadrático medio (MSE, por sus siglas en inglés, *Mean-Squared Error Estimation*) obtenemos el mejor predictor lineal insesgado (BLUP, por sus siglas en inglés *Best Linear Unbiased Predictor*) para la media de una área pequeña, bajo el modelo de Fay Herriot de $y_d = x_d^t \beta + u_d$. El estimador resultante se obtiene ajustando el modelo en la ecuación (7), el BLUP bajo el modelo de FH de y_d se define como:

$$\hat{y}_d^{FH} = x_d^t \tilde{\beta} + \tilde{u}_d, \quad (8)$$

donde $\tilde{u}_d = \gamma_d(\hat{y}_d^{DIR} - x_d^t \beta)$, es el BLUP de u_d , siendo $\gamma_d = \sigma_u^2 / (\sigma_u^2 + \psi)$ y donde $\tilde{\beta}$ es el estimador de mínimos cuadrados ponderados de β bajo el modelo (7), dado por:

$$\tilde{\beta} = \left(\sum_{d=1}^D \gamma_d x_d x_d' \right)^{-1} \sum_{d=1}^D \gamma_d x_d \hat{y}_d^{DIR}. \quad (9)$$

El BLUP es un promedio ponderado del estimador directo de la encuesta y el estimador sintético de regresión, con ponderaciones que dependen de la varianza de los efectos aleatorios de cada área. Si β , σ_u^2 y ψ , son conocidos para el modelo, el BLUP de u_d es $\tilde{u}_d = \gamma_d(u_d + e_d)$, donde $\gamma_d = (\sigma_u^2 + \psi)^{-1} \sigma_u^2$.

El BLUP para la media desconocida θ_d es:

$$\hat{Y}_d^{BLUP} = \begin{cases} x_d^t \beta + \gamma_d(\bar{y}_d - x_d^t \beta) & \text{si } d \in A; \\ x_d^t \beta & \text{si } d \notin A, \end{cases} \quad (10)$$

donde A es el conjunto de las áreas pequeñas en el que y_d es observada, además se debe tener en cuenta que ψ hace referencia a las varianzas que provienen de los datos estimados en la muestra o de otras variables auxiliares y se asumen conocidas, teniendo en cuenta esto a menudo no se tiene en cuenta para el cálculo del MSE. Sin embargo si β y σ^2 se asumen conocidos, el (EBLUP, por sus siglas en inglés *Empirical Best Linear Unbiased Predictors*), es:

$$\hat{Y}_d^{FH} = \begin{cases} x_d^t \hat{\beta} + \hat{\gamma}_d(\bar{y}_d - x_d^t \hat{\beta}) & \text{si } d \in A; \\ x_d^t \hat{\beta} & \text{si } d \notin A. \end{cases} \quad (11)$$

De acuerdo con Mukhopadhyay & McDowell (2011) la representación del EBLUP para las áreas pequeñas observadas se puede expresar como una combinación lineal convexa del estimador directo $\hat{Y}_d$ y el estimador sintético $x_d^t \hat{\beta}$.

1.2.1. Estimador del error cuadrático medio (MSE) para el modelo de Fay-Herriot

Bajo normalidad Prasad & Rao (1990), presentaron de forma analítica una aproximación de segundo orden (con error $o(D_{-1})$ para un número de áreas D grande) al MSE del EBLUP para el modelo de Fay-Herriot :

$$MSE(\hat{y}_d^{FH}) = g_{1d}(\sigma_u^2) + g_{2d}(\sigma_u^2) + g_{3d}(\sigma_u^2), \quad (12)$$

dónde:

$$g_{1d}(\sigma_u^2) = \gamma_d \psi_d, \quad g_{2d}(\sigma_u^2) = (1 - \gamma_d)^2 x_d' \left(\sum_{d=1}^D \gamma_d x_d x_d' \right)^{-1} x_d,$$

y

$$g_{3d}(\sigma_u^2) = (1 - \gamma_d)^2 x_d' (\sigma_u^2 + \psi_d^2) \overline{var}(\hat{\sigma}_u^2)$$

Teniendo en cuenta a Molina (2018), $\overline{var}(\hat{\sigma}_u^2)$ es la varianza asintótica del estimador $\hat{\sigma}_u^2$ de σ_u^2 dependiente del método de estimación que se utilice, $g_{1d}(\sigma_u^2)$ es el error debido a la predicción de los efectos aleatorios del área u_d , de orden $O(1)$ cuando D crece, por lo que no tiende a cero. $g_{2d}(\sigma_u^2)$ es el error que se obtiene del vector de coeficientes de regresión β . Finalmente $g_{3d}(\sigma_u^2)$ nos indica el error de la estimación de la varianza de σ_u^2 , donde los últimos dos términos de la ecuación tienden a cero a medida que D aumenta en orden $O(D^{-1})$.

En este sentido $g_{2d}(\sigma_u^2)$ y $g_{3d}(\sigma_u^2)$ tienden a desaparecer para un D de gran tamaño, mientras que $g_{1d}(\sigma_u^2)$ no desaparece, deben tenerse en cuenta los tres términos para evitar una infraestimación del MSE para tamaños moderados de D .

Existen dos métodos habituales para la estimación $\hat{\sigma}_u^2$, los cuales son máxima verosimilitud (en inglés, *maximum likelihood*, ML) y máxima verosimilitud restringida o residual (en inglés, *restricted/residual* ML, REML). El método REML presenta una corrección del estimador ML de la varianza de σ_u^2 por los grados de libertad que se obtienen a partir del vector de coeficientes del modelo β que presenta un estimador menos sesgado para tamaño de muestra pequeños o finitos.

Lo que conlleva que $\hat{\sigma}_u^2$ se define como el estimador REML y la varianza asintótica se obtiene como el inverso de la información de Fisher $J(\sigma_u^2)$.

$$\overline{var}(\hat{\sigma}_u^2) = J^{-1}(\sigma_u^2) = 2 \left(\sum_{d=1}^D (\sigma_u^2 + \psi_d)^{-2} \right)^{-1}.$$

Por lo que se puede decir que $g_{2d}(\hat{\sigma}_u^2)$ y $g_{3d}(\hat{\sigma}_u^2)$ son los estimadores respectivos de $g_{2d}(\sigma_u^2)$ y $g_{3d}(\sigma_u^2)$ insesgados de segundo orden, lo que significa que el sesgo es $o(D^{-1})$ por lo tanto tiende a cero mas rapido que D^{-1} cuando D crece. Se debe tener en cuenta que $g_{1d}\hat{\sigma}_u^2$ tienen un sesgo no despreciable como estimador de $g_{1d}\sigma_u^2$ que da como resultado $-g_{3d}(\sigma_u^2) + o(D^{-1})$. Al tener esta situación se debe sumar dos veces a $g_{3d}(\hat{\sigma}_u^2)$ y así corregir el sesgo de $g_{1d}(\hat{\sigma}_u^2)$, generando finalmente el estimador insesgado de segundo orden propuesto por Prasad & Rao (1990) y desarrollado en Molina (2018), como:

$$mse_{PR}(\hat{y}_d^{FH}) = g_{1d}(\hat{\sigma}_u^2) + g_{2d}(\hat{\sigma}_u^2) + g_{3d}(\hat{\sigma}_u^2).$$

Este estimador de Prasad & Rao (1990) del MSE es estable (o también puede llamarse eficiente) e insesgado bajo el diseño cuando se promedia a lo largo de un número grade de áreas, obteniendo resultados con menor sesgo frente a la utilización de estimadores directos.

1.3. Estimador de Battese-Harter-Fuller: modelo a nivel de unidad

Este estimador fue propuesto por Battese et al. (1988) y aplicado a un problema en el campo de la agricultura, proponiendo un modelo mixto para estimar las extensiones de ciertos cultivos. Estos autores

desarrollaron un modelo lineal mixto para estimar extensiones de determinados cultivos. Es por esto que si se tiene la posibilidad de consultar información auxiliar a nivel de unidad, es posible obtener muy buenas estimaciones para los dominios de interés.

Así mismo, el enfoque que se tiene en este modelo, considera la variable de interés como una variable aleatoria en un modelo poblacional, relacionando los valores que se tienen de la variable de estudio con el valor de la covariable medida en cada unidad. Se debe tener en cuenta que este estimador se basa en el tamaño muestral total con lo que se puede obtener una mayor eficiencia que el modelo a nivel de área, si y solo si, se tiene que las variables auxiliares se encuentran a nivel de cada individuo y que se encuentren relacionadas con la variable respuesta.

De acuerdo con Molina & Marhuenda (2015) en el modelo poblacional se tiene que para cada unidad de los dominios d , se encuentra un vector de información Y_{dk} , que es el puntaje promedio de la prueba de matemáticas que pertenece a cada una de las ETC y x_{dk} es un vector de información auxiliar obtenido a través de las covariables del modelo:

$$y_{dk} = x'_{dk}\beta + u_d + e_{dk} \quad \text{con} \quad k = 1, \dots, N_d \quad \text{y} \quad d = 1, \dots, D, \quad (13)$$

con $u_d \sim N(0, \sigma_u^2)$ y $e_{dk} \sim N(0, \sigma_e^2)$ donde u_d corresponde a los efectos de área, e_{dk} son los errores individuales, σ_u^2 y σ_e^2 son las varianzas de u_d y e_{dk} ; se asume que los efectos de área y el error a nivel de individuo son independientes. La covarianza entre e_{dk} y e_{dl} es igual a 0, $k \neq l$. En particular se observa que los individuos dentro de los dominios se encuentran correlacionados.

En el caso que σ_u^2 y σ_e^2 son conocidas el BLUP para el área pequeña de $\bar{Y}_d$, está dado por:

$$\hat{\bar{Y}}_d^{BLUP} = N_d^{-1} \left(\sum_{k \in s_d} y_{dk} + \sum_{k \notin s_d} \tilde{y}_{dk}^{BLUP} \right), \quad (14)$$

donde $\tilde{y}_{dk} = x'_{dk}\tilde{\beta} + \tilde{u}_d$ es el BLUP para Y_{dk} ; además, $\tilde{u}_d$ es el BLUP de u_d que corresponde al efecto aleatorio del dominio y n_d es el tamaño muestral del área d y está dado por:

$$\tilde{u}_d = \gamma_d(\bar{y}_{dk} - \bar{x}'_{dk}\tilde{\beta}), \quad \text{con} \quad \bar{y}_{dk} = \frac{1}{n_d} \sum_{k \in s_d} y_{dk}, \quad \bar{x}_{dk} = \frac{1}{n_d} \sum_{k \in s_d} x_{dk} \quad \text{y} \quad \gamma_d = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2 / n_d). \quad (15)$$

De acuerdo a la anterior ecuación se puede observar que el BLUP de, $\bar{Y}_d$ depende de las varianzas desconocidas σ_u^2 y σ_e^2 ; el EBLUP, $\hat{\bar{Y}}_d$, se puede escribir sustituyendo las varianzas del BLUP por $(\hat{\sigma}_u^2$ y $\hat{\sigma}_e^2)$,

$$\hat{\bar{Y}}_d = N_d^{-1} \left(\sum_{k \in s_d} y_{dk} + \sum_{k \notin s_d} \hat{y}_{dk} \right) \quad (16)$$

donde $\hat{Y}_{dk} = x'_{dk}\hat{\beta} + \hat{u}_d$ es el EBLUP para y_{dk} , $\hat{u}_d$ es está dado por:

$$\hat{u}_d = \hat{\gamma}_d(\bar{y}_{dk} - \bar{X}'_{dk}\hat{\beta}) \quad \text{con} \quad \hat{\gamma}_d = \hat{\sigma}_u^2 / (\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / n_d). \quad (17)$$

Es importante notar que el EBLUP requiere para sus cálculos el vector de medias de la población de las variables auxiliares $\bar{x}_d$ y sus tamaños poblacionales, N_d , aparte de los datos muestreados. Obteniendo de esta forma el estimador EBLUP:

$$\hat{\bar{Y}}_d^{EBLUP} = N_d^{-1} \left(\sum_{k \in s_d} y_{dk} + \sum_{k \in r_d} \hat{y}_{dk}^{EBLUP} \right) \quad (18)$$

El BLUP es insesgado teniendo en cuenta el modelo (13), y es óptimo en el sentido de minimizar el error cuadrático medio de los estimadores lineales en la muestra e insesgados. El EBLUP aumenta en gran consideración la eficiencia con respecto a los estimadores directos e incluso al estimador de Fay-Herriot, ya que en su proceso de estimación utiliza información con mayor detalle y de forma mas eficiente, sin llegar a reducir sus resultados a medias o totales.

1.4. Estimador del error cuadrático medio para un modelo de Battese-Harter-Fuller

De acuerdo con Molina (2018) el error cuadrático medio que tiene el EBLUP $\hat{\bar{Y}}_d^{EBLUP}$ de $\bar{Y}_d$, tiene la posibilidad de aproximarse mediante una fórmula analítica a un estimador de segundo orden a través de una fórmula similar a la de Prasad y Rao que se describió al momento de definir el MSE del estimador Fay-Herriot. Sin embargo otra opción disponible no requiere un número de áreas o dominios D grandes, esta opción se refiere a una técnica de remuestreo conocida como *Bootstrap* paramétrico para poblaciones finitas propuesto por W. González-Manteiga (2008), que se encuentra adaptado a la estimación de las medias de las áreas $\bar{Y}_d$.

El procedimiento *Bootstrap* conocido como (b) es el siguiente:

1. Ajustar el modelo (13) a los datos de la muestra $Y_s = (y'_{1s}, \dots, y'_{Ds})'$ y obtener los estimadores de los parámetros del modelo $\hat{\beta}$, $\hat{\sigma}_u^2$ y $\hat{\sigma}_e^2$ por medio del EBLUP de $\bar{y}_d$.
2. Generar los efectos de las áreas de la forma $u_d^{(b)} \sim^{iid} N(0, \hat{\sigma}_u^2)$, con $d = 1, \dots, D$.
3. Generar independientemente errores *Bootstrap* para las unidades de la muestra en el área $e_{di}^{(b)} \sim^{iid} N(0, \hat{\sigma}_e^2)$, $i \in S_d$.
4. Generar las medias poblacionales de los errores en las áreas $\bar{E}_d^{(b)} \sim^{iid} N(0, \frac{\hat{\sigma}_e^2}{N_d})$, con $d = 1, \dots, D$.
5. Calcular las verdaderas medias *Bootstrap* de las áreas, $\bar{Y}_d^{(b)} = \bar{X}_d' \hat{\beta} + u_d^{(b)} + \bar{E}_d^{(b)}$, con $d = 1, \dots, D$. Obsérvese que el computo de la media $\bar{Y}_d^{(b)}$ no requiere los valores individuales x_{dk} , para cada unidad fuera de la muestra del área $i \in r_d$.
6. Usando los vectores de los valores de las variables auxiliares para las unidades de la muestra x_{di} , $i \in S_d$, generar las variables respuesta para las unidades de la muestra a partir del modelo:

$$\bar{Y}_{dk}^{(b)} = \bar{x}_{di}' \hat{\beta} + u_d^{(b)} + \bar{e}_{dk}^{(b)}, \quad i \in S_d, \quad \text{con } d = 1, \dots, D \quad (19)$$

7. Para la muestra original $s = s_1 \cup \dots \cup s_D$, sea $y_s^{(b)} = ((y_{1s}^{(b)})', \dots, (y_{Ds}^{(b)})')'$ el vector *Bootstrap* de valores en la muestra. Ajustar el modelo 13 a los datos *Bootstrap* $y_s^{(b)}$ y calcular los EBLUPs *Bootstrap* $\hat{\bar{Y}}^{EBLUP(b)}$, con $d = 1, \dots, D$.
8. Repetir los pasos 2 y 6, para $b = 1, \dots, B$ y obteniendo las medias verdaderas $\bar{y}_d^{(b)}$ y los correspondientes EBLUPs $\hat{\bar{Y}}^{EBLUP(b)}$ para cada repetición *Bootstrap*, expresada como (b). Los estimadores *naive Bootstrap* del MSE de los EBLUPs $\hat{\bar{Y}}^{EBLUP(b)}$, obtenidos mediante el *Bootstrap* paramétrico son:

$$MSE_b\left(\hat{Y}^{EBLUP(b)}\right) = \frac{1}{B} \sum_{b=1}^B \left(\hat{Y}^{EBLUP(b)} - \bar{Y}_d^{(b)}\right)^2 \quad \text{con } d = 1, \dots, D \quad (20)$$

El estimador (20) no es insesgado de segundo orden sino de primer orden, lo que quiere decir que su sesgo no decrece más rápido que D^{-1} cuando el número de áreas D aumenta. Hay una gran variedad de correcciones de sesgo en la literatura sin embargo estas pueden tener comportamientos al tomar valores negativos después de su estimación o valores positivos que no son insesgados de segundo orden, de acuerdo con Molina (2018), y junto con esto, este tipo de correcciones aumentan la varianza en el MSE, por este motivo el estimador "*naiveBootstrap*" que no realiza la corrección del sesgo es la opción más deseable dentro de los estimadores no analíticos.

2. Descripción de la información

2.1. Diseño de muestreo

Como parte de la aplicación de la estimación en áreas pequeñas nuestro objetivo se encuentra focalizado en estimar el puntaje promedio de la prueba de matemáticas en la prueba SABER 11 del año 2017 a través de los estimadores desarrollados anteriormente. La población objetivo de la prueba pertenece a los establecimientos educativos oficial y no oficial, rural y urbano de Colombia.

Se excluyen de la población objetivo a aquellos estudiantes matriculados en establecimientos de educación para adultos, instituciones de educación especial o programas de educación alternativa no formal o no presencial y los estudiantes matriculados en jornadas nocturnas y sabatinas de los establecimientos educativos, en caso de que estos los ofrezcan, debido a que es posible que ya tengan un resultado previo en el examen en alguna aplicación anterior o porque no se encuentren dentro del programa que ofrecen los establecimientos educativos para grado 11 de manera regular.

Como parte fundamental del proceso de selección de una muestra y que represente de manera acorde la característica de interés de la población se debe definir de manera concreta los siguientes conceptos:

- **Unidades de muestreo:** establecimientos educativos y estudiantes.
- **Unidades de observación:** estudiantes.
- **Unidades de análisis:** establecimientos educativos.
- **Variable de interés:** puntaje de matemáticas de la prueba SABER 11 de 2017.

El diseño desarrollado es probabilístico donde las unidades de muestreo tienen una probabilidad de inclusión mayor a cero de ser incluidas en la muestra. La característica de estratificación se generó a través de la naturaleza (es decir si es oficial o no oficial) y la zona (es decir si es rural o urbano), obteniendo un resultado por zona/sector que proviene del marco muestral.

El marco muestral utilizado, contiene información a nivel de establecimiento educativo, estudiante y sus respectivos puntajes. Para el establecimiento se tiene clasificación según el puntaje promedio obtenido por sus estudiantes, su zona, sector. Para el estudiante se tiene información a nivel demográfico y sus respectivos puntajes por cada una de las pruebas presentadas, con el respectivo puntaje total.

El tamaño muestral que es representativo a nivel de los estratos fue diseñado a través del software estadístico R; el desarrollo de la selección proviene del marco de establecimientos educativos de la siguiente forma:

- **Primera etapa:** de la agrupación obtenida en la estratificación se realiza una selección de las instituciones que hacen parte de cada uno de los estratos realizando una selección πPT , donde se tuvo en cuenta el número de estudiantes que pertenecen a cada una.
- **Segunda etapa:** dentro de los establecimientos educativos (EE), obtenidos en la primera etapa se seleccionan dentro de las instituciones por medio de un muestreo aleatorio simple (MAS); en este trabajo todos los estudiantes al interior de los EE, cuentan con la misma probabilidad de ser seleccionados ya que cumplen con las características requeridas para presentar la prueba.

Por este motivo y con el objetivo de mejorar las estimaciones del parámetro de interés, se realiza una estratificación para disponer de una mayor homogeneidad en la variable que estamos considerando, los cuales describen las características principales de un EE, definidas por el Ministerio de Educación, (Gutierrez 2009).

Como se conoce a priori el marco de muestreo de cada uno de los estudiantes que presentaron la prueba para la segunda mitad del año 2017, es posible generar la estratificación puesto que estas características se encuentran como información auxiliar para cada individuo. De esta forma podremos obtener su comportamiento sin presentar un sesgo en los estratos que se generan y así poder representar de una manera inclusiva a todos los tipos de instituciones que se encuentran a nivel nacional.

Se evidencia que dentro de cada uno de los estratos se tiene un número de individuos que nos interesa medir, sin embargo este diseño de muestreo trae consigo un conjunto de probabilidades proporcionales al tamaño de una característica de información auxiliar (número de estudiantes). De esta forma el diseño de muestreo πPT con un tamaño de muestra fijo e igual a n , tiene su fundamento en una construcción de probabilidades de inclusión que tengan en cuenta la siguiente relación:

$$\pi_k = \frac{nx_k}{t_x}, \quad 0 < \pi_k < 1;$$

Adicionalmente, el resultado del diseño muestral a nivel de cada ETC se representa en forma de mapa donde se utiliza una partición de rupturas naturales de Jenks, de acuerdo con Smith (2006), realiza agrupaciones de manera natural a los datos que son similares y los límites se establecen en los puntos donde se generan diferencias considerables.

Es importante resaltar que en este trabajo se realizaron validaciones con tamaños de muestra diferente, con el objetivo de verificar el tamaño óptimo dentro de cada una de las ETC y así definir el modelo para aplicar los estimadores a nivel de área e individuo.

Esta validación indica que no se encuentra una variación porcentual de los coeficientes de variación de las validaciones comparándolos con los que se obtuvieron en la muestra inicial, de esta forma se corrobora la validez del diseño de muestreo. En el proceso de validación, adicionalmente se verificó con tamaños de muestra más grandes que dieron como resultado durante la estimación directa de los CVE presentaron una disminución, ya que dentro de las (ETC) que cuentan con un mayor tamaño muestral, por lo que realizando la estimación por medio de HT disminuían, sin embargo este análisis y su comportamiento están fuera del objetivo de este trabajo.

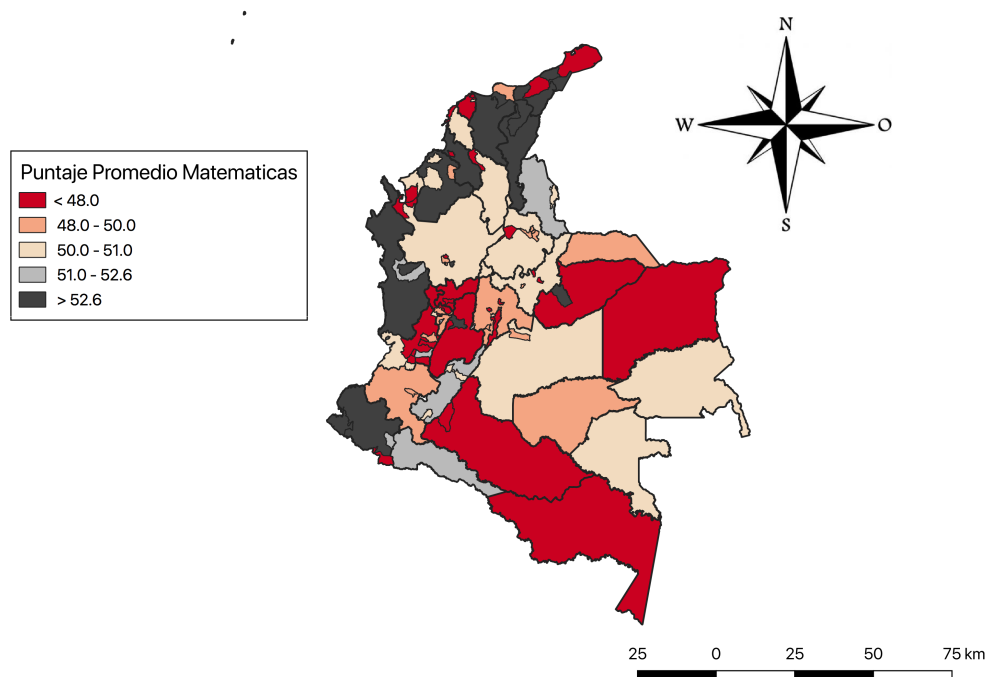
2.2. Comportamiento de la prueba de matemáticas

El Icfes brinda la posibilidad de obtener resultados de calidad a diferentes niveles de agregación a través de sus pruebas estandarizadas, como lo son a nivel Colombia, entidades territoriales certificadas, establecimientos privados y oficiales, que tienen división entre rurales y urbanos que generando esta agregación hace que pertenezcan a una ETC, estos establecimientos tienden a tener características socioeconómicas similares entre sí y diferente entre ellos.

De la página del DANE, obtuvimos el acceso a los microdatos que se encuentran en la investigación de la educación formal de la sección de educación. De allí podemos tener el registro de las entidades

territoriales que cuentan con el código de la institución; cabe aclarar que se tendrá información para aquellas instituciones que cuentan con este registro, el mismo se obtiene de la base de datos de la carátula única por sede educativa para las aplicaciones de la prueba mencionadas.

Al tener como insumo la carátula única y el código DANE de la sede se realizó una agrupación por cada una de las ETC desde la aplicación 2014 hasta 2017, obteniendo la media de los puntajes en cada una de las pruebas genéricas que se aplicaron durante cada prueba y así se generó un marco que fuera lo suficientemente robusto para seleccionar las variables auxiliares posteriormente. Además de esto se realiza una agrupación por los niveles de desempeño que se obtuvieron en cada una de las pruebas.



(a) Puntaje promedio prueba de matemáticas en 2014 por ETC

Figura 1: Resultados históricos por ETC de la prueba de matemáticas

La figura 1 muestra la distribución real de los puntajes promedios de la prueba de matemáticas a nivel de cada ETC, para la aplicación de 2014, como ejemplo de la información auxiliar con la que se cuenta para el desarrollo de este artículo, que equivalen a los datos teóricos, ya que contamos con la información completa a nivel de estudiante y de establecimiento, adicionalmente de la información auxiliar de los puntajes tenemos información correspondiente a los niveles de desempeño de los estudiantes y también se realizó una clasificación entre niveles de desempeño alto y bajo.

Bajo este tipo de información tenemos que las variables de puntaje de las pruebas anteriores, funcionan de manera de soporte ya que en ocasiones diferentes establecimientos no se presentan en las pruebas Saber, lo que ayuda a minimizar la pérdida de información para estos casos.

Con esta información es posible comparar los resultados obtenidos con los estimadores propuestos en este artículo y los resultados reales para identificar el ajuste que tienen al momento de tener información auxiliar que nos permita tener estimaciones en áreas con una muestra reducida.

3. Análisis y resultados

El resultado del diseño de muestreo aplicado se presenta el siguiente mapa que muestra la ubicación de las ETC de acuerdo a una escala de colores, que representa el tamaño muestral, con el cual se realizaron las estimaciones correspondientes del puntaje.

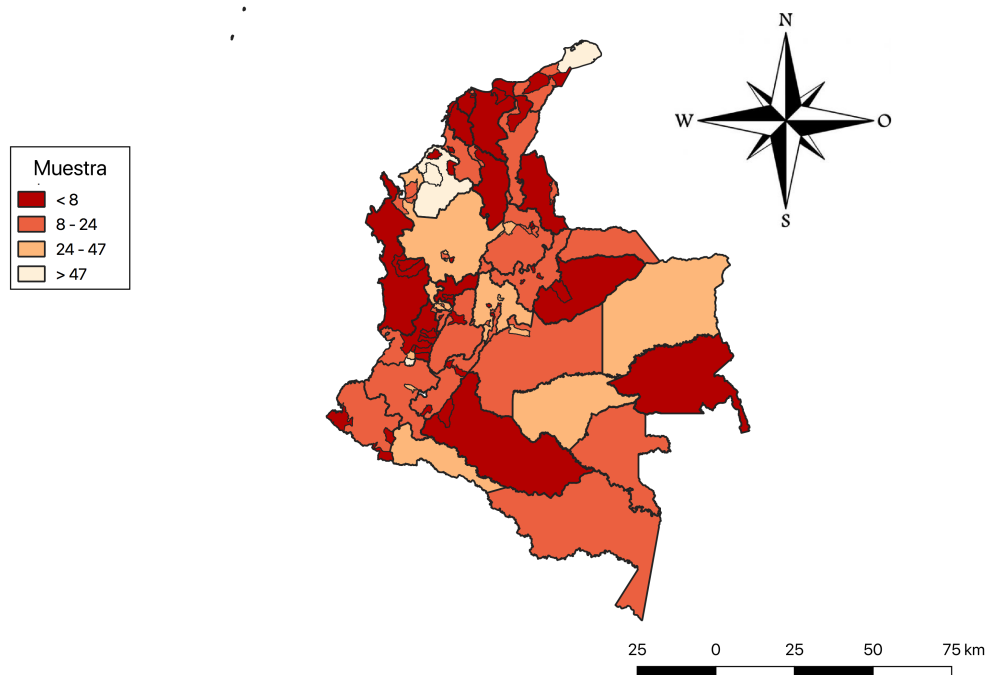


Figura 2: Escala del tamaño muestral a nivel de las ETC.

La escala de colores que se puede observar en la mapa 2 indica que las ETC que se encuentran con el color más oscuro, son aquellas que presentan el tamaño de muestra más reducido, así sucesivamente la escala va a ir aumentando con las ETC que cuentan con un tamaño de muestra más alto, siendo este último un color más claro.

En el intervalo de $n < 8$ se tiene el 38 % de los establecimientos educativos. En el siguiente intervalo se encuentran las ETC que cuentan con un tamaño muestral entre 8 y 24 correspondientes al 38 %, es decir que los dos primeros intervalos contienen el 76 % de las instituciones. Esto indica que son una mayor cantidad las ETC que cuentan con un tamaño de muestra pequeño.

El 24 % de las ETC cuentan con la mayor cantidad de muestra, por lo cual, para poder cubrir el 76 % de las ETC restantes podemos aplicar la estimación en áreas pequeñas.

Siendo una oportunidad de reducir costos y optimizar tiempos de análisis en pruebas que se aplican a gran escala, como las que aplica el Icfes, que por su magnitud logística, no se aplican anualmente por su alto costo en el despliegue de la misma.

Una posible solución es aplicarla de tipo muestral regularmente, esto tendría consecuencias positivas para expertos en educación y en evaluación de políticas públicas, ya que con esto se puede medir de manera frecuente y tener resultados en un período de tiempo más corto.

3.1. Resultados del modelo de Fay-Herriot

El modelo de Fay-Herriot se ajustó basado en un modelo lineal mixto donde se tuvo en cuenta el uso de variables auxiliares de acuerdo a la metodología dispuesta por la estimación en áreas pequeñas, que hacen referencia a la información de las pruebas Saber, en los períodos anteriores a la prueba de 2017. Con las cuales se realizó una exhaustiva selección revisando la significancia y aporte al modelo.

Las variables que se tuvieron en cuenta para generar el modelo de Fay-Herriot son los puntajes de las pruebas genéricas, las cuales fueron agregadas a nivel de área, para nuestro caso por ETC, para las aplicaciones de la prueba Saber 11 de 2014, 2015 y 2016, siendo: Matemáticas, Ciencias Naturales, Lectura Crítica, Sociales y Ciudadanas e Inglés. Estas variables se seleccionaron a través de dos metodologías; la primera de tipo temático que hace referencia al lado contextual y su alto contenido de asociación a los establecimientos educativos y una fuerte relación temporal dentro de cada uno de los EE.

Debido a que los EE deben presentar a sus estudiantes en cada prueba Saber 11, hay un impacto positivo para este trabajo, ya que es posible tener la información perteneciente a cada una de las instituciones a través del tiempo y además de ello a nivel de ETC. Es por esta razón, que no se utilizaron variables a nivel de estudiante, puesto que no cuentan con una relación a través del tiempo para ser medidos mediante esta prueba y así contar con información confiable. Para los estudiantes esta prueba es un evento de única presentación y validez (fuera de los casos donde se inscriben de forma individual, que se excluyeron al momento de realizar las estimaciones, esto se puede revisar en la sección de población objetivo con mayor detalle).

El método práctico aplicado para la selección de variables, se denomina *stepwise* y se utiliza frecuentemente para seleccionar modelos las variables explicativas óptimas y significativas. Esta práctica es una combinación de los métodos de *backward* y *forward*. El método *stepwise* remueve las variables no significativas del modelo a medida que incluye nuevas variables que se evalúan con el resto de variables que se encuentran incluidas con el objetivo de verificar si resultan significativas, que en tal caso se dejan en el modelo para evaluar las siguientes variables o por el contrario se vuelve a repetir el proceso excluyéndolas, hasta dar como resultado las variables con mayor significancia dentro del modelo que generan el mejor resultado. El modelo que presentó el mejor ajuste obtuvo un R^2 de 0.84 y un AIC de 128.16 más bajo frente a los demás modelos verificados, utilizándolo para la aplicación del modelo de Fay-Herriot.

Las variables seleccionadas fueron los puntajes promedios de Lectura Crítica, Sociales y Ciudadanas, Inglés de 2014, Matemáticas, Ciencias Naturales en 2015, Lectura Crítica, Matemáticas e Inglés de 2016 a nivel de ETC. El ajuste de este modelo se presenta en la sección de anexos con sus respectivas estimaciones de los coeficientes del modelo (β).

Con las variables seleccionadas, se aplicó la metodología de Fay-Herriot presentada en la sección 1.2, obteniendo como resultado la estimación del puntaje promedio de la prueba matemáticas para la prueba Saber 11 del 2017 para cada una de las ETC. Uno de los objetivos de este trabajo fue posterior a obtener los resultados por FH se compararon frente a su efectividad con el estimador directo de Horvitz y Thompson. Así mismo, como disponíamos del marco de muestreo completo, fue posible realizar comparaciones conjuntas del puntaje real para cada una de las ETC con las estimaciones obtenidas para los dos estimadores nombrados previamente.

En el cálculo del estimador se utilizó la función *eb lupFH* del paquete SAE, (Molina & Marhuenda 2015), en el cual, como insumo se utilizaron los estimadores directos obtenidos por el estimador HT (la media y sus correspondientes varianzas muestrales). Por medio del EBLUP se obtienen las estimaciones por cada una de las ETC del puntaje de la prueba de matemáticas y las estimaciones de los coeficientes β de las variables utilizadas.

Para visualizar las estimaciones que se obtuvieron a través del estimador FH, se realizaron por medio de la ubicación geográfica de cada una de las ETC en el país, utilizando la misma herramienta para los resultados de la estimación directa.

Cabe resaltar que los resultados de los MSE y CVE se presentarán en la sección de conclusiones a fin de

consolidar estos resultados de manera conjunta para los tres estimadores, (HT, FH y BHF).

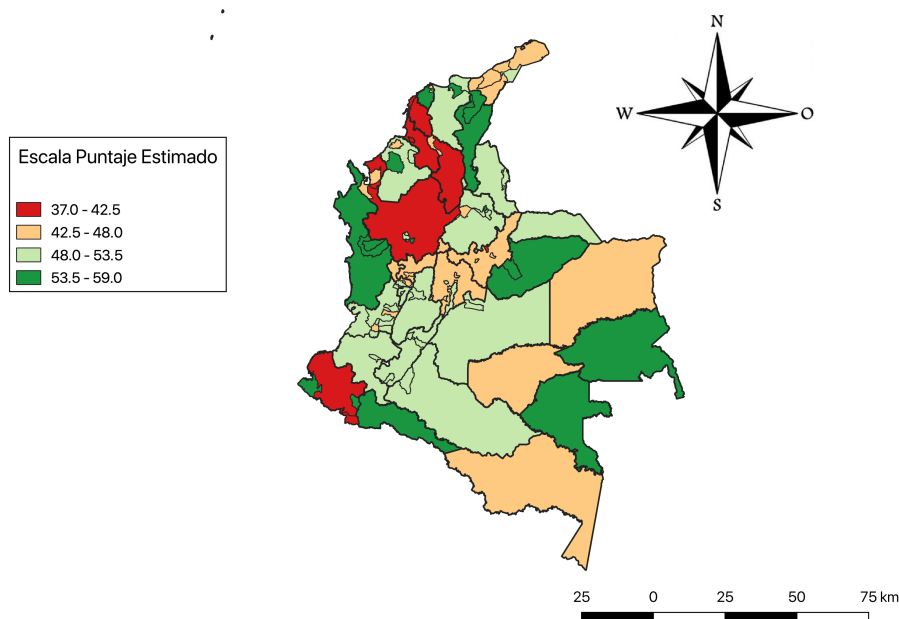


Figura 3: Horvitz-Thompson

Por medio de la ubicación geográfica en el país de cada una de las entidades territoriales se observa que los estimadores HT (Horvitz-Thompson) y FH (Fay-Herriot) de manera estimada se asemejan sus resultados al del puntaje real (los resultados de las estimaciones de los puntajes se encuentran en una tabla de resumen por ETC en la sección de Anexos), el cual se obtuvo con el total de la información en cada una de las pruebas. Con estos resultados podemos observar que la estimación FH tiende a ser más cercana a las estimaciones del puntaje real, lo que indica de primera instancia que tener variables auxiliares, implica una ganancia de precisión al momento de tener áreas pequeñas con poca información para un análisis confiable, como se puede observar en el estimador de HT, que no las incluye en su estimación.

Los resultados obtenidos por FH se ajustan en un 58 % a los resultados obtenidos por el estimador real, sobre un 25 % de los resultados de los puntajes estimados que se obtuvieron con el estimador HT. Además de esto en las ETC como Vichada y Casanare que tuvieron una muestra menor a 8 individuos, los resultados los resultados se mantuvieron a una desviación aceptable, no siendo tan disperso en su resultados. Entre el grupo de las ETC que tuvieron un tamaño de muestra pequeño están Magdalena, Chocó, Risaralda, Guainía y Norte de Santander tuvieron resultados similares a los del puntaje real al tener en cuenta la información auxiliar, sin embargo, las ETC: Florencia y Caquetá que tuvieron puntajes alejados de la realidad, teniendo la misma característica de área pequeña.

Para aquellas ETC que cuentan con tamaños de muestra más grandes como Cundinamarca, Vichada, Villavicencio, entre otras tuvieron resultados menos dispersos, en comparación con aquellas ETC que tuvieron poca muestra, esto significa que se tiene una mejora en las estimaciones donde hay disponibilidad de tener una cantidad de individuos representante dentro de las áreas pequeñas e información auxiliar, generando resultados precisos, debe tenerse en cuenta que estos resultados hacen referencia a la estimación FH. De esta forma es posible notar que las estimaciones de los puntajes respectivas al estimador FH, en un porcentaje del 60 % al 70 % representan de forma acertada los puntajes que se obtienen en el puntaje real a nivel nacional.

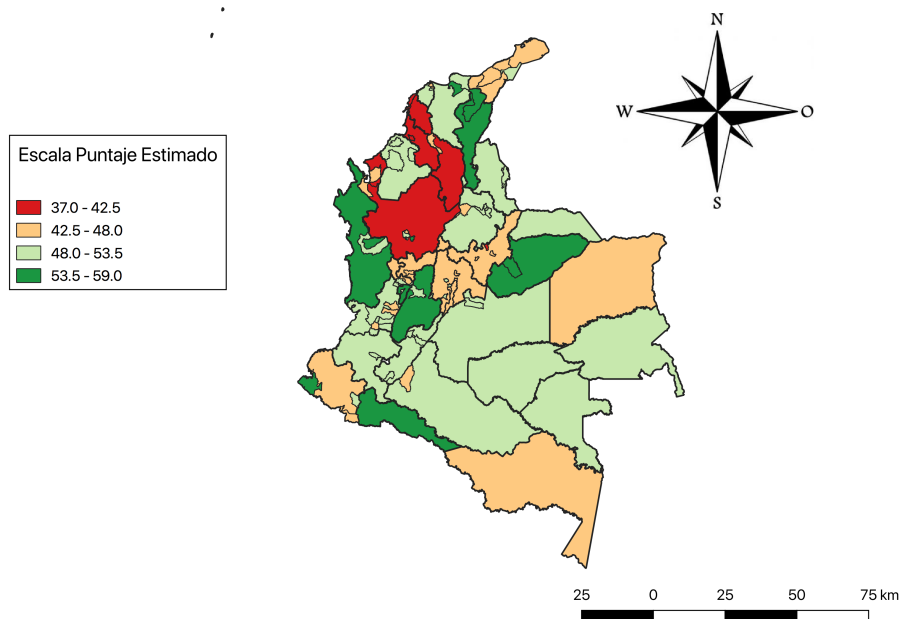


Figura 4: Fay-Herriot

Al visualizar la escala de la estimación promedio del puntaje de matemáticas por medio del estimador directo de HT y el puntaje de matemáticas real para las ETC, se puede ver que las zonas que tienen un desempeño mucho menor en las pruebas con promedios menores a 42.5 (representadas por el color rojo) cubren la parte norte del país; específicamente se encuentran dentro de los departamentos de Atlántico, Bolívar, Sucre y Antioquia a excepción de unas entidades territoriales que tienen un puntaje mayor dentro de estos. Sin embargo a este grupo de ETC se une Nariño, esto evidencia que la situación de educación que se está presentando en esta zona a través de las pruebas SABER evidencia un problema en su calidad.

Para tener una claridad del comportamiento que tiene el estimador directo se puede observar el mapa del puntaje real, con lo cual a través de las escalas de colores indica que hay ETC's que se encuentran con puntajes por arriba o abajo del puntaje real siendo estas un 72 %. En este sentido si se tiene como ejemplo Nariño, su puntaje real se encuentra un escalón más arriba del estimado, esto muestra que para una ETC con unos de los tamaños más altos de muestra, se presenta una mayor precisión dentro de sus estimaciones, resultando relativamente cercana a los reales.

Si se observan las ETC de Guinía y Vaupés, se puede ver que hay una estimación en el límite superior de la escala del estimador directo, es decir, las estimaciones tienen un puntaje mayor comparándolo con su puntaje real, por lo que al momento de realizar interpretaciones únicamente bajo este estimador, genera incertidumbre por la credibilidad que pueda ofrecer, aún con ETC que cuenten con individuos dentro de ellas. Por lo contrario para las ETC con menor tamaño de muestra hay resultados más alejados como sucede en la zona de Tolima.

3.2. Resultados del modelo Battese Harter Fuller

Para estimar el puntaje promedio de la prueba de matemáticas de la prueba Saber 11 del 2017, además de utilizar un modelo a nivel de área también se utilizó un modelo lineal mixto con errores anidados que se encuentra a nivel de individuo. En nuestro caso los individuos fueron los establecimientos educativos (EE)

que pertenecen a cada una de las 95 ETC, y que se encuentren registrados con el código de identificación del Ministerio de Educación.

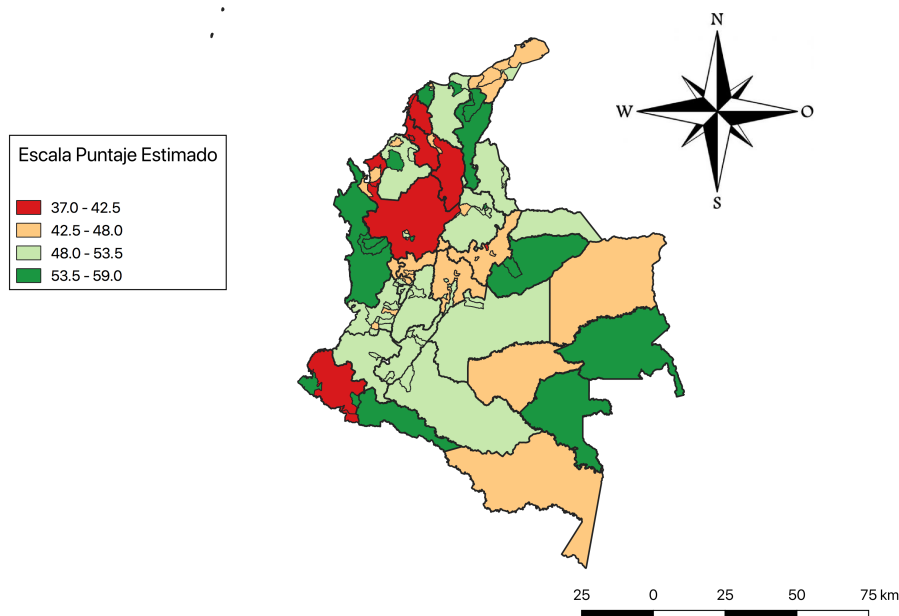


Figura 5: Horvitz-Thompson

Se tuvo en cuenta para la creación del modelo los puntajes medios de las pruebas Saber 11 en los años anteriores como variables de auxiliares refiriéndonos a los puntajes obtenidos por los establecimientos educativos, de igual forma los tamaños poblaciones y los dominios donde se estimó el puntaje de la prueba, en nuestro caso las ETC.

En la selección de variables se aplicó el procedimiento de *stepwise* y la selección por forma temática para aquellas variables que tienen una mayor relación al estudio y al contexto educativo. Como en el estimador de Fay-Herriot; de igual forma se tuvieron en cuenta las bases que contienen los resultados agregados a nivel de establecimientos de la prueba Saber 11 para 2014, 2015 y 2016 para posteriormente generar la agregación a las ETC.

La siguiente lista muestra las variables utilizadas a nivel de Establecimientos Educativos (EE) para este modelo: estas variables hacen parte del puntaje promedio en las pruebas de Ciencias Naturales (para la aplicación de la prueba Saber 11 en los años 2015 y 2016), Lectura Crítica, Matemáticas, Sociales y Ciudadanas e Inglés para la aplicación de la prueba Saber 11 del 2016.

Para obtener dichos resultados se debe tener en cuenta que la media de los puntajes en cada una de las ETC se descompone en la suma de los valores que se observaron en la muestra y los que no, siendo para nuestro caso las ETC con menor tamaño de muestra, teniendo en cuenta la ecuación (18), de igual forma se van a generar las estimaciones de los puntajes para las áreas que cuentan con un tamaño de muestra significativo para tener estimaciones confiables.

En el caso del modelo de Battese-Harter-Fuller es posible analizar que al comparar los resultados con el estimador directo, se observa que este último tiende a tener resultados diferentes, presentando puntajes medios más bajos que los obtenidos a través del estimador de BHF. El cual es más similar al puntaje real en las ETC, siendo esta situación similar a lo que se presentó con el estimador de FH y las estimaciones de los puntajes promedios que se obtuvieron.

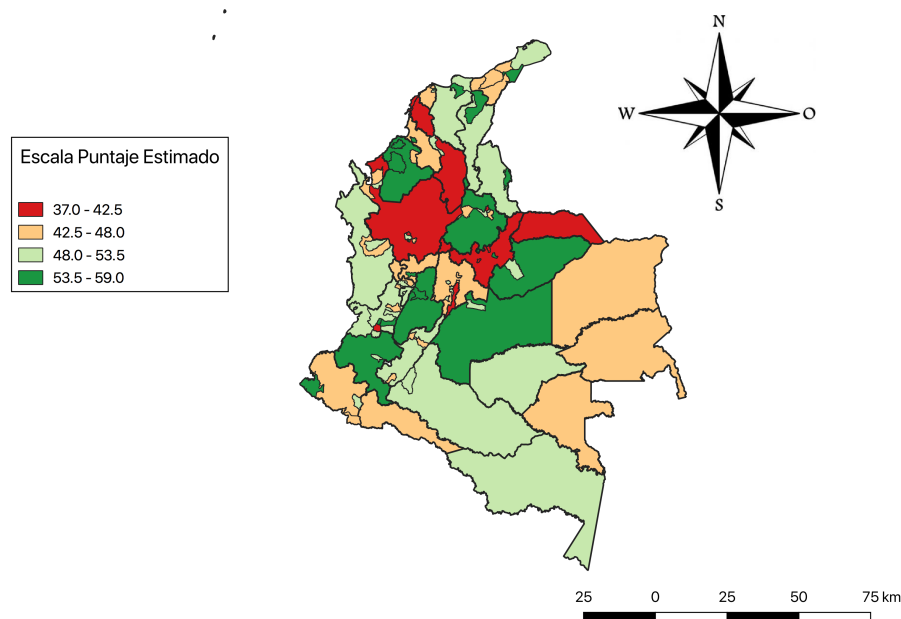


Figura 6: Battese-Harter-Fuller

Las estimaciones fueron realizadas con el paquete SAE del software R, y la función *eblupBHF*, cabe resaltar que para su realización, no se debe contar con la estimación de la varianza en comparación con el método de Fay-Herriot. Si tenemos en cuenta como se están comportando las estimaciones en las ETC, se puede observar en los mapas (Figura 5 y Figura 6) que hay unas áreas que cuentan con valores más cercanos a los puntajes reales, esto sobre todo en ETC donde se cuenta un mayor tamaño de muestra, sin embargo se tiene un porcentaje mayor de eficacia del estimador de 68 %, (porcentaje del puntaje cercano a los valores reales) en diferencia con el estimador de HT para las áreas que cuentan con un menor tamaño de muestra cercano a cero.

Dada la metodología *bootstrap*, para estimar el error cuadrático medio, se realizaron varias simulaciones, con la función *pbmseBHF*, donde se hicieron tres tipos de replicas, las cuales tomaron una variación de tiempo entre 10 a 12 minutos, a medida que se incrementaban las simulaciones el tiempo de igual forma aumentaba (se realizaron pruebas con 100, 200 y 300 replicas), evidenciando al final de la realización de cada replica si había una variación que nos indicara que se necesitaban realizar más o menos simulaciones, con lo que encontramos que no había una diferencia significativa dentro de los resultados del MSE.

Esto indica que se llegan a tener estimaciones con menores errores cuadráticos medios de áreas con tamaños reducidos que con el estimador directo, aunque desde la parte visible la escala indica que aquellos valores que presentan los resultados más bajos en promedio de la prueba de matemáticas para las ETC que se encuentran en rojo, ambos estimadores tienen una cobertura completa en las ETC.

Aunque se encuentre una diferencia en la escala para los valores promedios más altos de la prueba de matemáticas, obtenidos por el estimador de BHF de las ETC frente a su puntajes reales, puede llegar a tenerse en cuenta para aplicar en pruebas que cuenten con una identificación individual del Icfes y que su aplicación tenga un impacto de gran magnitud en cuanto a su despliegue a nivel nacional. Puesto que si es posible y brinda una ventaja tener una muestra con la que se incurre en menos gastos, un mejor control de la población y una actualización constante a nivel de estudiante e institución, se podrían aplicar esta metodología para generar estimaciones confiables dando a si la posibilidad a ser frecuentes

en su aplicación como actualmente sucede con la prueba Saber 3, 5, 9.

Adicionalmente con la muestra, se tendría un diseño representativo de la variable de interés, ya que por motivos logísticos, hay establecimientos que al momento de realizar la prueba presentan un tipo de inconsistencia de tipo locativo o logístico por el material, lo cual estaría cubierto por el desarrollo de la estimación en áreas pequeñas, reduciendo de este forma costos e información faltante que considerablemente se convierte en pérdidas a nivel de recursos públicos.

4. Conclusiones

Posterior a la realización de las estimaciones que se obtuvieron para la metodología de FH y BHF con el objetivo de tener resultados del puntaje promedio de matemáticas en las ETC y sobre todo en aquellas donde su tamaño de muestra no es representativo, fue posible obtener estimaciones confiables del puntaje con respecto a los valores reales que se tienen del marco muestral.

A partir de estos estimadores podemos resaltar en la figura 7 el comportamiento que tiene el CVE (coeficiente de variación) obtenido para cada uno de ellos en la estimación del puntaje de matemáticas para la prueba Saber 11 del 2017. Se puede resaltar que el coeficiente para el estimador directo por Horvitz y Thompson (identificado en la figura 7 de color verde) tiende a presentar valores altos en las ETC con menor muestra, en este sentido en promedio tuvimos un 29% del coeficiente estimado entre todas las ETC, lo que se considera como un nivel bajo de precisión, si se tiene en cuenta que hay una definición previa de un diseño de muestreo que tiene una representación consistente para cada una de las ETC.

Al observar los valores tan altos del coeficiente de variación obtenido por el modelo de HT, se procede a observar el comportamiento que tuvo el coeficiente para el estimador de FH (identificado en la figura 7 de color azul), el cual disminuye considerablemente a nivel de cada una de las ETC, en los cuales es posible percibir una mayor homogeneidad del mismo obteniendo para la mayoría de las ETC con un CVE de 5.72% en promedio, el cual refleja una disminución frente al resultado del estimador directo dando una mejor precisión, considerándose aceptable para las estimaciones del puntaje promedio, de igual forma se puede observar como hay una brecha considerable frente a estos dos primeros estimadores en la figura 7, lo que nos da paso a revisar los CVE para las estimaciones del modelo BHF.

De acuerdo al comportamiento que tuvieron el estimador de HT y el de FH, se realizó el ejercicio con el estimador de BHF (que se encuentra de color rojo en la figura 7) a nivel de unidad. Recordemos que nuestras unidades fueron los EE y estos a nivel de cada ETC obtuvieron CVE promedio del 3.04%, siendo más pequeño y óptimo que los dos métodos explicados previamente.

Las estimaciones tienden a ser mas eficaces al ajustar el modelo de Battese Harter Fuller dado que el número de observaciones que se tuvieron fueron el tamaño muestral del marco, lo cual es mucho mayor que el número de observaciones, número total de áreas, que en el modelo de Fay-Herriot, por este motivo los coeficientes de variación para las estimaciones del modelo de Battese.

Teniendo en cuenta que los modelos a nivel de área e individuo presentan unos resultados óptimos y confiables de aquellos dominios no observados o que tienen una representación muy baja en su tamaño muestral se puede tener en cuenta para ser utilizadas en la reducción de inversión de recursos para aplicar cada una de las pruebas Saber que tiene el Icfes. La metodología de estimación en áreas pequeñas ayuda a minimizar los errores (que se pueden ver a través de la figura 8), que se puedan presentar por la utilización de los estimadores directos (que se encuentra de color verde en la figura 8).

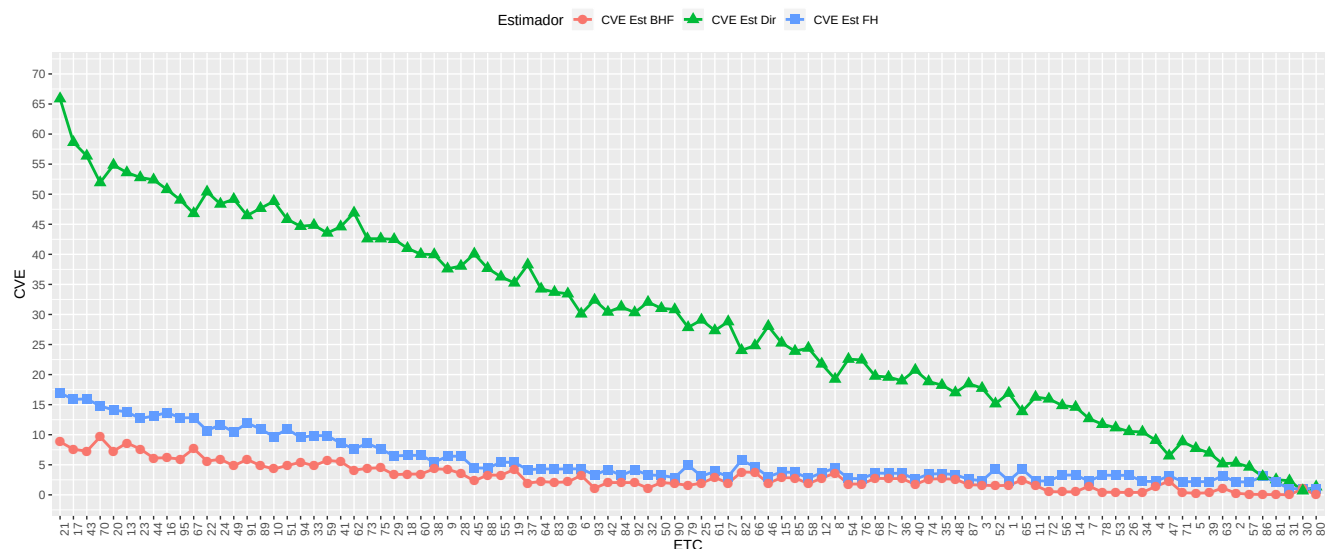


Figura 7: CVE por ETC. Estimadores: HT color verde, FH color azul, BHF rojo

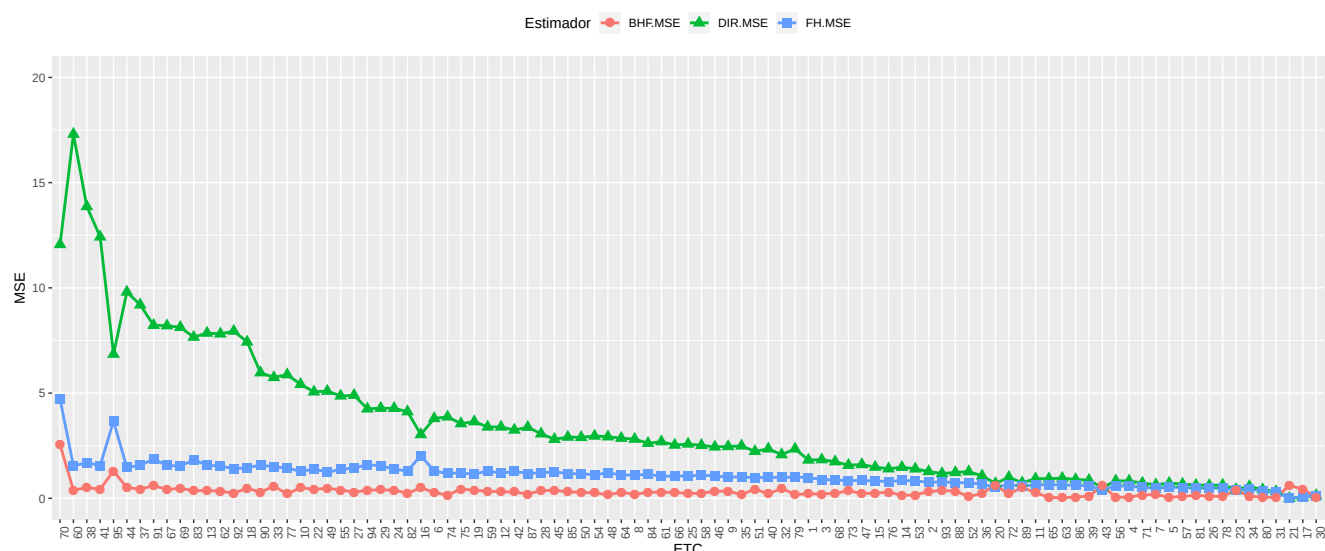


Figura 8: Error cuadrático medio (ó MSE) por ETC. Estimadores: HT color verde, FH color azul, BHF rojo

Además se puede observar como el estimador de FH minimiza el error cuadrático medio, MSE, (que se puede observar de color azul en la figura 8). Estos MSE's se encuentran orgnaizados de forma descendiente lo que nos indica que los EBLUP's del modelo a nivel de individuo aumenta considerablemente su precisión, siendo el que menor error produce comparado a través de los otros dos estimadores (que se encuentra de color rojo en la figura 8), dando estos resultados por utilizar información más detallada, además de esto se debe resaltar que el estimador de MSE para el modelo de BHF es un estimador estable bajo el diseño y es insesgado al promediarse a lo largo de los dominios.

El modelo de FH tiene como ventaja a nivel de aplicación que mantiene los datos anonimizados de los individuos utilizados al momento de realizar la agregación y su posterior desarrollo, de forma que cumple

correctamente con la política de privacidad de datos, en este caso de los estudiantes que presentan la prueba y concuerda con la misión del Icfes, por el contrario los datos utilizados a nivel de individuo en el modelo de BHF, tienden a ser más propensos a presentar este inconveniente y tengan que revisarse los resultados a fondo para no transmitir información de tipo confidencial.

El estimador de FH no se ve afectado en término de los datos atípicos dado que sus resultados se encuentran a nivel de agregación, es por esto que los puntajes que se encuentren en los extremos dentro de la muestra, no van a sesgar los resultados obtenidos mediante esta metodología.

Los resultados obtenidos por los modelos presentados identifican la capacidad de aplicabilidad de la estimación en áreas pequeñas en la prueba Saber 3, 5, 9 que al 2018 se presentaba cada 3 años de dos formas, una es de tipo censal y otra de tipo control. Esta última se desarrolla por medio de un diseño de muestreo para poder monitorear el comportamiento de la prueba que ya nos brinda el acceso a un diseño de muestreo óptimo y bien definido, ya que si se deseara ver su desarrollo para la aplicación censal los costos de desplazamiento de personal y de material estarían por fuera del presupuesto.

Por este motivo si es posible realizar este tipo mediciones a nivel muestral, anualmente, se reducirían costos y así mismo se podría obtener resultados confiables para las ETC y los establecimientos educativos por medio de los estimadores de áreas pequeñas, ya que si se llegasen a estimar los puntajes para establecimientos o estudiantes con el estimador directo se podrían tener interpretaciones deficientes de la calidad de educación en el país porque no se asemejarían a la realidad, como si lo han demostrado las estimaciones por la metodología de estimación en áreas pequeñas.

Además de esto si tenemos en cuenta que los impactos a nivel de la educación se ven reflejados en al menos 5 años después de la aplicación de las políticas públicas aplicadas, el monitoreo regular de la educación en el país daría efectos en un menor tiempo que si se continua realizando cada 3 años, recogiendo una mayor información y así los investigadores tendrían herramientas actuales para realizar sus respectivas evaluación de las políticas presentadas, siendo inminente un ahorro sustancial en los recursos que actualmente se utilizan para aplicar este tipo de pruebas, permitiendo tener la posibilidad de invertir en mejorar la calidad de educación (por ejemplo con infraestructura, acceso a tecnologías y mejor capacitación de los docentes), al disponer de resultados confiables que se pueden obtener en un tiempo no mayor a 10 años.

Como finalidad del trabajo y cumpliendo con el objetivo de obtener la estimación de puntajes promedios para la prueba de matemáticas de la prueba Saber 11, se procedió a verificar los puntajes que se obtuvieron por cada uno de los estimadores, todos siendo comparados con el puntaje real, visualizados en conjunto en la figura 12.

Se utilizó una metodología de para medir el ajuste de las estimaciones del puntaje para HT, FH y BHF del puntaje a través de un intervalo de confianza del 95 %, que toma en cuenta la variabilidad entre el puntaje real y el valor estimado.

El estimador directo HT tiene un 25 % de similitud en los resultados del puntaje promedio de matemáticas con respecto al puntaje real, en este sentido se tiene que el restante 75 % de las ETC presentaron una diferencia en las estimaciones del puntaje. Para el estimador FH, y siguiendo esta comparación se obtuvo un 40 % de estimaciones similares con respecto al puntaje real y para el estimador BHF 47 %, que aumenta la precisión sobre los estimadores previamente obtenidos.

Ya que el 38 % de las ETC contaron con un tamaño de muestra menor, se puede ver que los estimadores utilizados de áreas pequeñas cumplen con los objetivos de obtener mejores estimaciones cuando el tamaño de muestra es reducido, no solo siendo más precisos en estos casos, si no siendo también óptimo en las ETC con suficiente muestra.

Dentro de la estimación en áreas pequeñas se debe verificar que la disponibilidad de información auxiliar para poder realizar este tipo de aplicación, ya que en caso que no se cuenten con registros administrativos previos no podría desarrollarse. Para el caso de aplicación como lo es con las pruebas Saber del Icfes, este efecto se mitiga ya que, cuentan con suficiente información histórica a través del tiempo para todos los tipos de pruebas que ellos desarrollan. Es importante resaltar que aún teniendo una sola aplicación

previa, se pueden realizar la aplicación de la estimación de puntajes en establecimientos y estudiantes en los dominios que se definan, por medio de la estimación en áreas pequeñas.

4.1. Discusion

Si bien con el desarrollo de los estimadores propuestos (HT, FH y BHF), con sus respectivos modelos en el presente estudio, dando como resultado el conocimiento de la estimación del puntaje promedio de la prueba de matemáticas a nivel de las Entidades Territoriales Certificadas a nivel nacional, y poder incorporar técnicas estadísticas confiables que apoyen la formulación, desarrollo y aplicación de las pruebas aplicadas por el ICFES a gran escala y con esto la reducción de costos a través de que estas pruebas se puedan aplicar a un menor tamaño de población y que se puedan obtener resultados que nos indiquen el comportamiento de la educación, para diferentes tipos de agregaciones, de esta forma se optimizan recursos públicos en su desarrollo, aplicación, de tipo logístico y como punto a favor de reducción en los temas de calificación y procesamiento de los resultados, por lo cual se reducen tiempos y errores en esta última etapa del proceso generando así una ganancia de tipo global para el ICFES, el Ministerio de Educación y el gobierno en general.

Estos resultados de ser adoptados, serían de vital importancia para conocer de forma mas frecuente el estado de la educación y ampliar la cobertura en las muestras donde se aplique para tener una variación en términos de ubicación e identificación de zonas donde este derecho se encuentra resagado. Al poder diseñar políticas públicas que se basen en esta propuesta metodológica, podríamos llegar a impactar un porcentaje alto de niños y jóvenes que no pueden acceder a este derecho fundamental.

Para que esta metodología tenga éxito es importante que los diseños muestrales, la información auxiliar y la aplicación de la prueba deben garantizar que sean confiables. Para que se puedan tener las estimaciones de los puntajes que se desean. Siendo esta situación para el ICFES una ventaja ya que cuenta con toda información histórica a nivel de todas las pruebas a través de la política de datos abiertos para estudiantes e investigadores.

4.2. Investigaciones futuras

Como investigación futura para estimar el puntaje promedio de una de las pruebas que hagan parte de las pruebas Saber, como lo es la prueba Saber 11, es la aplicación del modelo de Fay-Herriot Espacio Temporal, que nos brinda la posibilidad de medir la correlación que tienen los dominios a partir de un efecto aleatorio que incluyó Salvati (2008) y la matriz de proximidad W generada por la conformación espacial de los dominios vecinos.

Lo cual permitiría estimar con mayor precisión a parte de la información auxiliar que se utilizó, indicar un resultado para las ETC que se encuentran porque podrían presentar cercanas comportamientos similares en sus estudiantes y así estimar el puntaje promedio de la prueba en común, adicionalmente poder desarrollar el modelo con variables de tipo económico en las ETC y de relación en cuanto a los recursos públicos que tiene disponibles, este tipo de información debe solicitarse al Ministerio de Educación.

Además de los estimadores presentados anteriormente una forma de representar los puntajes agregados para las pruebas del Icfes, es un método aplicado por el Banco Mundial para construir mapas de pobreza de Elbers et al. (2003), el cual una vez ajustado el modelo, se pueden realizar las estimaciones para sub áreas y sub dominios, teniendo la facultad de poder estimar a nivel de individuo. Un requisito para su utilización es que las variables auxiliares deben estar medidas de la misma forma en la que se hizo el estudio, es decir, que tener variables de puntajes de años anteriores facilitaría la aplicación de este método. Sin embargo debe tenerse en cuenta que puede verse alterado en sus resultados por datos atípicos.

Referencias

- Battese, G., Harter, R. & Fuller, W. (1988), 'An error-components model for prediction of county crop areas using survey and satellite data', *Journal of the American Statistical Association* **83**(401), 28.
- Bouzas, D. C. (2012), Estimación trimestral en áreas pequeñas de los ocupados de la encuesta de población activa según su actividad económica, Master's thesis, Universidad de Coruña.
- DANE (2015), 'Carátula única'.
*<http://microdatos.dane.gov.co/index.php/catalog/561>
- Elbers, C., Lanjouw, L. & Lanjow, P. (2003), 'Micro-level estimation of welfare', *World Bank*.
- Fay, R. E. & Herriot, R. A. (1979), 'Estimates of income for small places: An application of james-stein procedures to census data', *Journal of the American Statistical Association* **74**(366), 269–277.
- Ferreira, J. (2011), Estimación en dominios, Master's thesis, Universidad de la Republica Uruguay.
- Garcia, S. (n.d.), 'Estimación en áreas pequeñas'.
- Gutierrez, H. A. (2009), *Estrategias de Muestreo, Diseño de Encuestas y Estimación de Parámetros*, Ediciones de la U.
- Hidirglou, M. (2016), 'Small-area estimation: Theory and practice', *Statistical Innovation and Research Division, Statistics Canada*.
- Horvitz, D. G. & Thompson, D. J. (1952), 'A generalization of sampling without replacement from a finite universe', *Journal of the American Statistical Association* **47**(260), 663–685.
- ICFES (2010), 'Acceso a base de datos y diccionarios'.
*<ftp://ftp.icfes.gov.co/>
- ICFES (2018), *Guía de orientación saber 11 para instituciones educativas 2018-1*, Instituto Colombiano para la Evaluación de la Educación.
- Martín, A. P. (2008), Estimación en áreas pequeñas bajo modelos lineales mixtos con dos factores aleatorios anidados, PhD thesis, Universidad Miguel Hernández de Elche.
- MEN (2014), 'Sistema nacional de indicadores educativos para los niveles de preescolar, básica y media en colombia', *Sistema Nacional de Indicadores*.
- MEN (2015), 'Guía fondos de servicios educativos', *Documentos de trabajo*.
- Molina, I. (2018), *Desagregacion de datos en encuestas de hogares*, Vol. 97, CEPAL.
- Molina, I. & Marhuenda, Y. (2015), *R package sae: Methodology*.
- Morales, D. (2017), Estimación en áreas pequeñas basadas en el diseño muestral, in 'Centro de Investigación Operativa', Universidad Miguel Hernández de Elche.
- Mukhopadhyay, P. K. & McDowell, A. (2011), 'Small area estimation for survey data analysis using sas software', *SAS Institute Inc., Cary, NC*.
- Prasad, N. G. N. & Rao, J. N. K. (1990), 'The estimation of mean squared errors of small area estimators', *Journal of the American Statistical Association* (85), 163–171.
- Rao, J. (2003), *Small Area Estimation*, Wiley.
- Salvati, N. (2008), 'Small area estimation: the eblup estimator based on spatially correlated random area effects', *Statistical Methods and Applications* **17**(1), 113 – 141.

- Smith, M. (2006), ‘Univariate classification schemes’, *Geospatial Analysis—A Comprehensive Guide* .
- W. González-Manteiga, M. J. Lombardía, I. M. D. M. . L. S. (2008), ‘Bootstrap mean squared error of a small-area eblup’, *Journal of Statistical Computation and Simulation* **78**, 443–462.
- Warnholz, S. & Schmid, T. (2016), ‘Simulation tools for small area estimation: Introducing the r-package saesim’, *Austrian Journal of Statistics* **45**, 55.
- Zea, J. & Ortiz, F. (2018), ‘Small area estimation methodology (sae) applied on bogota multipurpose survey (emb)’, *Romanian Statistical Review* .

5. Anexos

ETC	Entidad	Est Dir MSE	FH MSE	BHF MSE
1	Amazonas	1.82	0.98	0.21
2	Antioquia	1.27	0.76	0.32
3	Medellin	1.84	0.89	0.20
4	Bello	0.83	0.58	0.05
5	Envigado	0.72	0.52	0.04
6	Itagui	3.80	1.29	0.27
7	Turbo	0.64	0.48	0.18
8	Arauca	2.81	1.10	0.19
9	Atlantico	2.46	1.01	0.34
10	Soledad	5.42	1.28	0.50
11	Barranquilla	0.93	0.64	0.27
12	Bogota D.C.	3.39	1.19	0.31
13	Bolivar	7.85	1.57	0.37
14	Magangue	1.48	0.85	0.14
15	Boyaca	1.48	0.82	0.24
16	Tunja	3.03	2.03	0.50
17	Duitama	0.06	0.05	0.44
18	Sogamoso	7.43	1.44	0.48
19	Caldas	3.64	1.14	0.38
20	Manizales	0.71	0.53	0.62
21	Caqueta	0.00	0.00	0.61
22	Florencia	5.06	1.39	0.43
23	Cartagena	0.42	0.39	0.38
24	Casanare	4.28	1.38	0.39
25	Cauca	2.58	1.05	0.24
26	Popayan	0.61	0.49	0.10
27	Cesar	4.90	1.44	0.29
28	Valledupar	3.07	1.21	0.37
29	Choco	4.28	1.55	0.40
30	Cordoba	0.14	0.13	0.05
31	Monteria	0.32	0.28	0.04
32	Lorica	2.08	1.00	0.47
33	Sahagun	5.74	1.49	0.57
34	Cundinamarca	0.56	0.44	0.11
35	Fusagasuga	2.49	1.01	0.18
36	Girardot	1.07	0.67	0.23
37	Soacha	9.19	1.58	0.44
38	Guainia	13.87	1.69	0.51
39	Guaviare	0.86	0.59	0.10
40	Huila	2.35	0.99	0.21
41	Neiva	12.42	1.55	0.43
42	La Guajira	3.25	1.28	0.30
43	Maicao	0.50	0.40	0.60
44	Magdalena	9.80	1.47	0.51
45	Cienaga	2.80	1.26	0.37

Figura 9: Estimación del error cuadrático medio por estimador

ETC	Entidad	Est Dir MSE	FH MSE	BHF MSE
46	Meta	2.44	1.05	0.34
47	Villavicencio	1.61	0.85	0.23
48	Narino	2.92	1.21	0.17
49	Pasto	5.09	1.24	0.48
50	Tumaco	2.90	1.16	0.29
51	Norte de Santande	2.23	0.99	0.44
52	Cucuta	1.27	0.75	0.08
53	Putumayo	1.40	0.83	0.16
54	Quindio	2.96	1.12	0.26
55	Armenia	4.87	1.38	0.40
56	Risaralda	0.83	0.58	0.05
57	Pereira	0.68	0.51	0.10
58	Dosquebradas	2.52	1.11	0.24
59	San Andres	3.39	1.30	0.34
60	Santa Marta	17.30	1.54	0.39
61	Santander	2.69	1.04	0.29
62	Bucaramanga	7.81	1.53	0.34
63	Barrancabermeja	0.96	0.63	0.03
64	Floridablanca	2.86	1.11	0.28
65	Giron	0.93	0.65	0.05
66	Sucre	2.53	1.07	0.28
67	Sincelejo	8.20	1.56	0.43
68	Tolima	1.74	0.87	0.24
69	Ibague	8.12	1.53	0.47
70	Valle del Cauca	12.06	4.71	2.54
71	Cali	0.72	0.52	0.16
72	Buenaventura	1.00	0.65	0.21
73	Buga	1.57	0.84	0.36
74	Cartago	3.87	1.19	0.16
75	Palmira	3.55	1.21	0.44
76	Tulua	1.40	0.80	0.30
77	Vaupes	5.88	1.44	0.21
78	Vichada	0.63	0.48	0.09
79	Quibdo	2.35	1.01	0.18
80	Uribia	0.41	0.35	0.05
81	Jamundi	0.62	0.48	0.14
82	Piedecuesta	4.12	1.30	0.23
83	Rionegro	7.66	1.82	0.37
84	Pitalito	2.62	1.14	0.28
85	Sabaneta	2.91	1.17	0.31
86	Zipaquirá	0.89	0.62	0.06
87	Mosquera	3.38	1.18	0.20
88	Ipiales	1.23	0.74	0.35
89	Facatativa	0.74	0.58	0.52
90	Apartado	5.97	1.59	0.29
91	Chia	8.22	1.86	0.60
92	Malambo	7.95	1.41	0.25
93	Riohacha	1.18	0.79	0.37
94	Yopal	4.25	1.59	0.39
95	Yumbo	6.85	3.66	1.35

Figura 10: Estimación del error cuadrático medio por estimador

ETC	Entidad	Prom Real	Prom Est Dir	Prom Est FH	Prom Est BHF
1	Amazonas	46	45	46	52
2	Antioquia	50	49	50	50
3	Medellín	49	49	49	67
4	Bello	44	44	44	42
5	Envigado	49	49	49	55
6	Itagüí	51	48	48	54
7	Turbo	47	47	47	44
8	Arauca	51	52	50	42
9	Atlántico	44	44	44	44
10	Soledad	45	48	44	45
11	Barranquilla	43	43	43	43
12	Bogotá D.C.	47	49	48	41
13	Bolívar	44	42	43	45
14	Magangüe	46	47	46	55
15	Boyacá	52	54	52	42
16	Tunja	48	42	46	48
17	Duitama	39	37	37	46
18	Sogamoso	46	45	45	46
19	Caldas	48	45	46	44
20	Manizales	46	43	45	46
21	Caquetá	49	47	47	49
22	Florencia	47	51	48	53
23	Cartagena	41	37	37	41
24	Casanare	43	39	41	57
25	Cauca	50	52	51	50
26	Popayán	52	51	51	52
27	Cesar	57	58	57	46
28	Valledupar	57	57	57	55
29	Chocó	58	59	59	58
30	Córdoba	55	55	55	64
31	Montería	52	54	53	57
32	Lorica	52	47	50	55
33	Sahagún	49	49	50	43
34	Cundinamarca	51	51	51	60
35	Fusagasugá	49	49	50	53
36	Girardot	53	52	53	42
37	Soacha	46	48	45	52
38	Guainía	51	53	51	51
39	Guaviare	48	47	48	48
40	Huila	53	56	54	53
41	Neiva	52	53	52	48
42	La Guajira	55	53	55	55
43	Maicao	56	52	53	57
44	Magdalena	55	58	56	45
45	Ciénaga	56	55	56	46

Figura 11: Estimación del puntaje promedio de matemáticas por estimador

ETC	Entidad	Prom Real	Prom Est Dir	Prom Est FH	Prom Est BHF
46	Meta	51	51	52	51
47	Villavicencio	51	51	51	53
48	Narino	58	57	56	58
49	Pasto	54	54	53	50
50	Tumaco	58	56	57	59
51	Norte de Santande	52	53	52	52
52	Cucuta	51	52	52	55
53	Putumayo	54	55	54	45
54	Quindio	50	48	49	50
55	Armenia	55	52	53	55
56	Risaralda	48	49	49	48
57	Pereira	49	50	50	52
58	Dosquebradas	44	46	45	52
59	San Andres	52	52	51	52
60	Santa Marta	50	52	50	50
61	Santander	52	49	51	54
62	Bucaramanga	52	57	52	52
63	Barrancabermeja	44	45	45	47
64	Floridablanca	51	52	51	51
65	Giron	51	52	51	51
66	Sucre	54	56	55	54
67	Sincedejo	41	42	41	57
68	Tolima	49	48	48	53
69	Ibague	54	52	53	55
70	Valle del Cauca	50	50	49	49
71	Cali	47	45	46	47
72	Buenaventura	52	51	51	50
73	Buga	48	48	48	48
74	Cartago	51	51	51	51
75	Palmira	53	52	53	54
76	Tulua	49	49	48	49
77	Vaupes	52	53	51	52
78	Vichada	48	46	46	52
79	Quibdo	53	54	53	46
80	Uribe	48	48	48	48
81	Jamundi	52	51	51	53
82	Piedecuesta	50	50	51	50
83	Rionegro	56	59	57	56
84	Pitalito	51	49	50	51
85	Sabaneta	43	44	43	43
86	Zipaquirá	46	43	46	46
87	Mosquera	52	53	52	49
88	Ipiales	45	41	43	45
89	Facatativa	52	52	52	50
90	Apartado	40	42	41	40
91	Chia	44	48	45	44
92	Malambo	53	42	53	53
93	Riohacha	48	45	46	48
94	Yopal	56	56	56	56
95	Yumbo	47	52	50	47

Figura 12: Estimación del puntaje promedio de matemáticas por estimador

Variables	Estimate	Std. Error	P-Valor
Intercepto	14.5727	9.4032	0.12487
Puntaje Lectura Crítica 20142	0.6146	0.3773	0.10695
Puntaje Sociales y Ciudadanas 20142	0.7455	0.4314	0.08758 .
Puntaje de Inglés 20142	-0.8528	0.2935	0.00466 **
Puntaje Matemáticas 20152	0.8224	0.3951	0.04036 *
Puntaje Ciencias Naturales 20152	-1.2275	0.4210	0.00453 **
Puntaje Lectura Crítica 20162	-0.7334	0.5103	0.15428
Puntaje Matemáticas 20162	0.9080	0.4151	0.03143 *
Puntaje Inglés 20162	0.4627	0.2843	0.10734

Tabla 1: Coeficientes de los parámetros del modelo FH

En la *tabla 1*, se presentan los coeficientes del modelo propuesto para la obtención de las estimaciones del puntaje promedio de la prueba de matemáticas.