UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Estimación del rendimiento medio en las pruebas PISA 2018. Un enfoque espacial desde la estimación en áreas pequeñas

Cupertino Jiménez Coley

Universidad Santo Tomás
Facultad de Estadística
División de Ciencias Económicas y Administrativas
Bogotá, D.C., Colombia
2023

Estimación del rendimiento medio en las pruebas PISA 2018. Un enfoque espacial desde la estimación en áreas pequeñas

Cupertino Jiménez Coley

Trabajo de grado presentado como requisito parcial para optar al título de:
Magister en Estadística Aplicada

Director:
Cristian Fernando Tellez Piñerez Ph.D.

Línea de Investigación:
Estimación en áreas pequeñas
Grupo de Investigación:
USTAdística

Universidad Santo Tomás
Facultad de Estadística
División de Ciencias Económicas y Administrativas
Bogotá, D.C., Colombia
2023

(Dedicatoria o un lema)

Dedicado a mis padres Elso Jimenez Rodelo y Lucy Coley Canchila, mis hermanos y demás familiares.

En general a todos aquellos que me apoyaron e impulsaron a seguir adelante para culminar un peldaño más de mi carrera.

"Nada llega después ni antes todo llega a su debido tiempo, ten paciencia sin desesperarse tus problemas se los lleva el tiempo"Diomedes Díaz.

"No vaya para dónde no lo inviten, no hable tampoco de lo que no sepa, un lápiz pesa menos que una pala, aunque la vida es la mejor escuela".Eden Muñoz.

"Nada es casualidad en la vida todo lo hizo mi Dios perfecto, todo el que sube y va pa arriba sabe que baja en cualquier momento"Diomedes Díaz.

Resumen

Estimar la habilidad de los estudiantes en las pruebas PISA se realiza utilizando valores plausibles obtenidos por medio de un modelo logístico de tres parámetros, con esta metodología solo se tiene en cuenta la información capturada en las pruebas y solo se pueden realizar estimaciones en los países y economías perteneciente a la OCDE (dominios) que fueron muestreados. Para mejorar la precisión de las estimaciones y poder realizar estimaciones en dominios no muestreados que cuenten con información auxiliar disponible y confiable, Tellez (2020) propuso utilizar la metodología de estimación en áreas pequeñas mediante un modelo Fay-Herriot. Esta metodología incorpora información auxiliar externa a la captada en la prueba y permite hacer estimaciones más precisas.

En esta investigación, se extiende la metodología propuesta por Tellez (2020) al incluir la variabilidad espacial en la técnica de estimación en áreas pequeñas. Esto se logra mediante el uso del modelo Fay-Herriot espacial, que tiene en cuenta la autocorrelación espacial entre los dominios. Este enfoque se aplica específicamente a las pruebas de lectura y matemáticas del PISA 2018. La información auxiliar utilizada en el modelo Fay-Herriot espacial incluye variables asociadas a la educación, ciencia, tecnología, características sociodemográficas, economía, infraestructura y desarrollo. Estas variables auxiliares proporcionan información adicional para mejorar la precisión de las estimaciones de habilidad de los estudiantes en los dominios

Palabras clave:. habilidad, dominios, modelo logístico de tres parámetros, datos plausibles, Estimación en áreas pequeñas, variabilidad espacial.

Abstract

Estimating students' abilities in PISA tests is done using plausible values obtained through a three-parameter logistic model. With this methodology, only the information captured in the tests is taken into account, and estimates can only be made for countries and economies belonging to the OECD (domains) that were sampled. To improve the precision of the estimates and be able to make estimates in unsampled domains that have available and reliable auxiliary information, Tellez (2020) proposed using the small area estimation methodology through a Fay-Herriot model. This methodology incorporates external auxiliary information captured in addition to the test and allows for more precise estimates.

In this research, the methodology proposed by Tellez (2020) is extended by including spatial variability in the small area estimation technique. This is achieved by using the spatial Fay-Herriot model, which takes into account the spatial autocorrelation among domains. This approach is specifically applied to the reading and mathematics tests of PISA 2018. The auxiliary information used in the spatial Fay-Herriot model includes variables associated with education, science, technology, sociodemographic characteristics, economy, infrastructure, and development. These auxiliary variables provide additional information to improve the precision of the estimates of students' abilities in the domains.

Keywords: ability, domains, three parameter logistic model, plausible data, small area estimation, spatial variability

Lista de Tablas

1 Introducción

La creciente demanda de generación de estadísticas de calidad en el menor tiempo posible ha impulsado la implementación de métodos de muestreo que permiten comprender las características de una población mediante un subgrupo representativo. Estos métodos ofrecen ventajas en términos de tiempo, costo y precisión (Gutiérrez 2016). Lo anterior es importante cuando los datos y los análisis estadísticos son la base para la toma de decisiones que afectan el quehacer de una organización o la asignación de recursos públicos a situaciones urgentes.

Un ejemplo de esto se da con las pruebas SABER 3°, 5°, 7° y 9° aplicadas por el ICFES (Instituto Colombiano para la Evaluación de la Calidad de la Educación), donde se eligen algunos estudiantes como muestra representativa, permitiendo obtener resultados en un periodo corto (ICFES 2022). Esto conlleva a cumplir los propósitos para los cuales fueron creadas las pruebas, en cuanto a servir como base en los planes de mejoramiento de cada institución de educacion y brindar información que sirva como referente estratégico para establecer políticas educativas a nivel nacional o territorial.

Ahora bien, los métodos clásicos de estimación basados en el diseño de muestreo, conocidos como estimadores Directos, utilizan solo la información de la muestra en cada dominio (Molina 2019). En la mayoría de los casos, estos son representativos en los niveles de desagregación planeados en el diseño muestral (Rao & Molina 2015), si se requiere una desagregación mayor de la planificada en el diseño muestral, estos estimadores pueden ser insuficientes en términos de precisión. Por lo tanto, han surgido otras alternativas, como la estimación en áreas pequeñas (SAE, por sus siglas en inglés) que generan estimaciones más precisas (Molina 2019).

La mejoría en la precisión de las metodologías SAE se debe a que se aprovecha la información de otras áreas y datos auxiliares externos, lo que permite predecir dominios con muestras pequeñas o nulas y obtener estimaciones confiables y precisas Pfeffermann (2002), Rao (2003). Longford (2005) añade que las áreas que tienen características similares tendrán estimaciones parecidas.

En este contexto, es crucial contar con información completa en la muestra del dominio para que se puedan realizar las estimaciones, ya que la falta de datos dificulta el uso de las metodologías SAE. El caso concreto que ocupa esta investigación tiene esta característica de datos

incompletos. Se trata de las pruebas PISA (sigla en inglés para Programa para Evaluación Internacional de Alumnos), realizadas por la OCDE (Organización para la Cooperación y el Desarrollo Económicos) para evaluar las habilidades en lectura, matemáticas y ciencias de estudiantes en su etapa final de formación escolar. El objetivo es proporcionar información que respalde la toma de decisiones y el diseño de políticas educativas en los países miembros de la OCDE y otros países asociados. Las pruebas se aplican a estudiantes de aproximadamente 15 años de edad y se utiliza la Teoría de Respuesta Ítem (TRI) para establecer una relación entre los resultados de la prueba y los rasgos latentes que se desean medir en los estudiantes OECD (2016).

La TRI, originalmente desarrollada en psicometría, permite construir instrumentos de medición con propiedades invariantes entre poblaciones Attorresi et al. (2009, p.180). Esta teoría asume que la habilidad del individuo y las características del ítem son variables latentes y propone modelos matemáticos para modelar su relación Embretson & Reise (2013), como el modelo logístico de tres parámetros (3PL) propuesto por Birnbaum (1958).

Hasta el año 2020, no se habían encontrado casos donde se combinara la metodología SAE, la TRI y la imputación múltiple. Sin embargo, dada la importancia de estas metodologías en la evaluación educativa y el diseño de políticas públicas, surge la metodología propuesta por Tellez (2020). Esta metodología estima la habilidad media de los estudiantes en la prueba PISA de matemáticas en 2015, que presenta datos faltantes en los ítems y dominios observados. Para abordar este problema, se utiliza la imputación múltiple a través de valores plausibles, siguiendo la propuesta de Rubin (1987). Otros trabajos similares también abordan estos desafíos, como los realizados por Triviño et al. (2020) y Tellez et al. (2021).

Para abordar el problema de datos faltantes, se utiliza la metodología propuesta por Tellez (2020), que se basa en la propuesta de Rubin (1987) para la aplicación de SAE en el contexto de datos faltantes. Esta metodología se fundamenta en la Teoría de Respuesta Ítem (por sus siglas en inglés TRI). La TRI es un enfoque utilizado en psicometría que permite establecer una relación entre los resultados de una prueba y los rasgos latentes o no observables que se desean estudiar en un individuo. Estos rasgos latentes podrían ser habilidades como la impulsividad, la extroversión o la habilidad de los estudiantes en el caso de las pruebas PISA. La TRI asume que la habilidad del individuo y las características del ítem son variables latentes y propone modelos matemáticos para modelar su relación Embretson & Reise (2013), como el modelo logístico de tres parámetros (3PL) propuesto por Birnbaum (1958).

En resumen, la metodología propuesta por Tellez (2020) combina SAE, TRI e imputación múltiple para abordar el problema de datos faltantes en las pruebas PISA. Esto permite estimar las habilidades de los estudiantes a partir de datos incompletos y brinda información valiosa para la evaluación de la educación y la formulación de políticas educativas.

En el presente estudio, se abordan los resultados de los estudiantes en las pruebas de lectura y matemáticas del programa PISA 2018, aplicadas por la OCDE, La metodología utilizada se basa en la propuesta de Tellez (2020) con la extensión del modelo Fay-Herriot espacial para la estimación del rendimiento medio en los dominios. Este modelo ha demostrado un mejor ajuste que el estimador Fay-Herriot en investigaciones recientes (Pratesi & Salvati 2008, Wawrowski 2016, Chandra & Salvati 2018).

Además, se incorpora información proveniente de áreas vecinas mediante la combinación de un estimador de áreas pequeñas y un modelo Simultáneamente Autorregresivo (SAR) o un modelo Autorregresivo Condicional (CAR), como se explicó en el trabajo de Petrucci & Salvati (2006). La selección de una muestra representativa de cada país y la respuesta de los estudiantes a un subconjunto de ítems relacionados con las temáticas de investigación son características clave del diseño de las pruebas PISA que cumplen con los requisitos necesarios para la aplicación de la metodología propuesta.

Dado que el rendimiento es una variable latente, la forma común de estimarlo es a través de modelos TRI, como el modelo logístico de tres parámetros (3PL). Sin embargo, debido a que los estudiantes solo responden a un subconjunto de ítems, los datos faltantes generan un error considerable en la medición del rendimiento individual. Para abordar este problema, se sigue el enfoque de imputación múltiple con K valores plausibles, basado en la metodología propuesta por Rubin (1987), siguiendo los ejemplos de Mislevy (1991), Mislevy, Beaton, Kaplan & Sheehan (1992) y Tellez (2020). Sin embargo, se reconoce que este procedimiento no es suficiente cuando se requiere generar estimaciones en dominios no observados, por lo que se propone el uso del modelo Fay-Herriot espacial, que permite obtener estimaciones precisas en esos casos.

La investigación se estructura en seis secciones, comenzando con el planteamiento del problema en la introducción. A continuación, se exponen los conceptos que fundamentan la metodología SAE y el modelo logístico de tres parámetros, así como los objetivos y la forma en que se pretenden alcanzar. Se incluye una simulación para estudiar las características de los estimadores a comparar, seguida de una aplicación real utilizando los datos de la prueba PISA 2018. Finalmente, se presentan las conclusiones y los compromisos adquiridos en el estudio.

2 Marco teórico

2.1. Población y muestra

En estadística, los conceptos de población y muestra son fundamentales para la construcción de conocimientos. Särndal et al. (2003) define población como como el conjunto completo de elementos o unidades de análisis que se desea estudiar. Esta puede ser finita o infinita, y puede estar definida por criterios geográficos, demográficos, socioeconómicos, u otros. A partir de esta población, se extrae una muestra que consiste en un conjunto de elementos representativos de la población o grupo al que pertenecen. El objetivo de la muestra es determinar las características que los identifican y luego trasladar esas características como ciertas a cualquier elemento de la población en su totalidad (Gutiérrez 2009).

Una muestra ideal debe cumplir con ciertas características, tales como tener un marco muestral completo y actualizado, ser seleccionada de manera aleatoria y ser lo suficientemente grande para obtener una precisión adecuada en las estimaciones Särndal et al. (2003).

2.1.1. Muestreo

El muestreo se refiere al proceso de seleccionar un subconjunto de unidades de una población con el propósito de hacer inferencias sobre la población en su conjunto Särndal et al. (2003). La selección de una muestra se realiza para ahorrar tiempo, dinero y recursos, y para evitar el costo y la dificultad de estudiar toda la población.

Muestra aleatoria

Gutiérrez (2016) define muestra aleatoria como aquel "subconjunto de la población que ha sido extraído mediante un mecanismo estadístico de selección". Dicha selección puede darse con un muestreo aleatorio simple, estratificado, sistemático, doble, conglomerado, submuestras interpenetrantes o por combinación de dos o más de los anteriores. Cualquiera que sea el procedimiento que se aplique, el objetivo de elegir una muestra con estas características es que sea representativa del universo que se estudia y se puedan obtener los estimadores de los parámetros.

Así las cosas, dado que lo que se quiere es que la muestra sea representativa de la población, se necesita que ésta a parte de ser aleatoria sea probabilística lo cual conlleva a que tenga

que cumplir unos requisitos para caracterizarse como probabilística. Särndal et al. (2003) y Gutiérrez (2009) establecen que en estos casos, siempre se puede definir el conjunto de muestra $S = s_1, \ldots, s_m$, a cada muestra posible, s, le corresponde una probabilidad de selección $P(s)$ conocida. Además, el proceso de selección garantiza que todo elemento de la población tenga una probabilidad de selección distinta de 0 y, finalmente, $\sum_{s \in Q} P(s) = 1$, donde Q es el conjunto de todas las posibles muestras.

Variable y parámetro de interés

Todas las investigaciones marcan en sus objetivos una característica de interés que motiva el estudio y sobre la que se calculan los datos estadísticos, esta se denomina variable de interés (Gutiérrez 2009). De dicha variable, se deriva una expresión numérica que resume los datos de estudio, asociados a cada individuo de la población y que se convierte en la finalidad del muestreo. A esta expresión numérica se le denomina parámetro de interés (Tellez 2020). Siendo U la población o universo y N el numero total de unidades en la población o universo, Algunos de los parámetros de uso más frecuente son:

1. Total poblacional: $t_y = \sum_{i \in U} y_i$
2. Media poblacional: $\overline{y}_U = \frac{t_y}{N}$
3. Varianza poblacional: $S^2_{yu} = \frac{\sum_{j \in U} (y_i - \overline{y}_U)^2}{N-1}$
4. Proporción poblacional: $P = \frac{X}{N}$, X es una variable binomial.

Dominio

En determinados casos, se requieren realizar estimaciones separadas del parámetro de interés para lo cual se define un subgrupo poblacional al que se le denomina como dominio (Gutiérrez 2009). Adicionalmente, debe cumplir con las siguientes condiciones:

1. $A_d \subset A$, tal que $A = \cup_{d=1}^{D} A_d$
2. Si $i \in A_l$, entonces $i \notin A_d$ para $d \neq l$
3. El número de elementos en el dominio A_d es N_d y es llamado **tamaño absoluto** del dominio.
4. La proporción de elementos en el dominio u_d con respecto al tamaño poblacional es $P_d = \frac{N_d}{N}$ y se conoce como **tamaño relativo** del dominio.

Estimador

Como se afirmó antes, para estimar los parámetros se usa una expresión numérica, por tanto, Un estimador es una regla o fórmula matemática que se utiliza para obtener una aproximación o estimación de un parámetro desconocido de una población, basándose en la información proporcionada por una muestra de la población. Wackerly, Mendenhall & Scheaffer (2014). Martínez (2012) aclara que es posible que muestras de la misma población tengan cuantificadores diferentes incluso usando el mismo estimador, por lo que siempre es necesario calcular el error de muestreo que es la diferencia entre el parámetro y el estimador. Particularmente en esta investigación se usarán estimadores de tipo Directo e Indirecto (Molina 2019).

Entre los estimadores directos, para este estudio se emplea el estimador de Horvitz-Thompson (HT), propuesto por Narain en 1951 y dado a conocer por quienes lo nombraron en 1952. Este estimador para el total de un dominio se define así:

$$\hat{t}_{dy,\pi}^{Dir} = \sum_{s} \frac{y_{di}}{\pi_{di}} = \sum_{s} f_{di} y_{di} \tag{2-1}$$

con $f_{di} = \frac{1}{\pi_{di}}$, conocido como el **factor de expansión**. Así mismo, la varianza del estimador (HT) está dada por:

$$Var(\hat{t}_{dy,\pi}^{Dir}) = \sum \sum_{U} \Delta_{dil} \frac{y_{di}}{\pi_{di}} \frac{y_{dl}}{\pi_{dl}} \tag{2-2}$$

con $\Delta_{dil} = \pi_{dil} - \pi_{di}\pi_{dl}$. Un estimador insesgado para $Var(\hat{t}_{dy,\pi}^{Dir})$ es:

$$\hat{Var}(\hat{t}_{dy,\pi}^{Dir}) = \sum \sum_{s} \frac{\Delta_{dil}}{\pi_{dil}} \frac{y_{di}}{\pi_{di}} \frac{y_{dl}}{\pi_{dl}} \tag{2-3}$$

con $\pi_{dil} > 0$ para todo i y $l \in s_d$ (Gutiérrez 2016). Existen otros estimadores Directos que se pueden consultar en Deville & Särndal (1992) los cuales son el estimador de Hájek, GREG, Calibrado entre otros.

Ahora bien, los estimadores indirectos incluidos para la presente investigacion son el estimador Sintético y el estimador Compuesto. En el Sintético, se asume que las subáreas resultantes de la división de la población tienen los mismos rasgos que la población total, supuesto que, se termina evidenciando en el sesgo del estimador ya que este es aproximadamente insesgado (Deville & Särndal 1992). De tal manera que, la recomendación es tener pocos estratos para que el volumen por estrato aumente y se incremente en la relación directamente proporcional, la eficiencia del estimador. Así, este estimador está definido por:

$$\hat{t}_{d}^{Sint} = \sum_{g} N_{dg} \hat{\bar{y}}_{g} \tag{2-4}$$

g es el grupo definido por la variable auxiliar, su valor esperado y sesgo son: $E[\hat{t}_d^{Sint}] = \sum_g N_{dg}\overline{y}_d$; $B[\hat{t}_d^{Sint}] \approx \sum_g N_{dg}(\overline{y}_g - \overline{y}_{dg})$, si $\pi_{il} \approx \pi_i\pi_l$ la varianza toma la forma:

$$Var(\hat{t}_d^{Sint}) = \sum_g N_{dg}^2 Var[\hat{\overline{y}}_d] \approx \sum_{g=1}^{G} \frac{N_{dg}^2}{N_g^2} \sum_{U_g} \frac{1-\pi_i}{\pi_i}(y_i - \overline{y}_d)^2 \tag{2-5}$$

$$\hat{Var}(\hat{t}_d^{Sint}) = \sum_{g=1}^{G} \frac{N_{dg}^2}{\hat{N}_g^2} \sum_{s_g} w_i(w_i - 1)(y_i - \hat{\overline{y}}_d)^2 \tag{2-6}$$

Igualmente, si se pretende es equilibrar el sesgo potencial del estimador Sintético y la inestabilidad del estimador Directo, se debe tomar una media ponderada de ambos y hallar el estimador Compuesto (Drew et al. 1982). Sea $\hat{\delta}_d^{Dir}$ un estimador Directo genérico de δ_d y $\hat{\delta}_d^{Sint}$ un estimador Sintético, el estimador Compuesto se define como:

$$\hat{\delta}_d^{Comp} = \phi_d\hat{\delta}_d^{Dir} + (1-\phi_d)\hat{\delta}_d^{Sint}, \quad 0 \leq \phi_d \leq 1 \tag{2-7}$$

siendo,

$$\phi_d = \left\{ \begin{array}{c} 1, si\hat{N}_d \geq \alpha N_d \\ \frac{\hat{N}_d}{\alpha N_d}, si\hat{N}_d < \alpha N_d \end{array} \right\}$$

Donde $\hat{N}_d = \sum_{i \in s_d} w_{di}$ y $\alpha > 0$ usualmente toma los valores de $\frac{2}{3}, 1, \frac{3}{2}$ y 2.

2.1.2. Estimación en áreas pequeñas

Según Molina (2019) cuando el error de muestreo del estimador directo es considerado como inaceptable para el indicador de interés, el área de estimación es pequeña. Pero la consideración de inaceptable es subjetiva, por lo que, en cada territorio, la entidad de estadística correspondiente fija ese límite de error de muestreo. Dicho esto, existe una demanda alta de información de grupos específicos de población, para áreas menores o grupos de desagregaciones no planeadas, en los que el tamaño muestral es pequeño y el nivel de confiabilidad de los resultados es baja e inhabilita su uso.

En cuanto a los métodos para hallar esta clase de estimaciones se basan en el modelo que integra la información con todas las áreas y que, además, proviene de una información auxiliar que se toma desde diversas fuentes. Estos métodos resultarán más realistas que un método indirecto y se recomiendan en casos donde existe la posibilidad de un sesgo potencial (Molina 2019). Uno de los modelos más usados en la estimación de áreas pequeñas es el de Fay & Herriot (1979). A continuación, se explica este modelo y el modelo Fay-Herriot espacial.

Modelo de Fay-Herriot

El modelo Fay & Herriot (1979) es uno de los métodos más utilizados y conocidos en la estimación de áreas pequeñas, se define como:

$$\delta_d = x_d'\beta + u_d,\ d = 1, \ldots, D \tag{2-8}$$

Donde β es el vector de coeficientes común en todas las áreas y u_d es el efecto aleatorio de área, u_d representa la heterogeneidad de los indicadores δ_d a través de las áreas no explicadas por las variables auxiliares consideradas. En el modelo más simple se asume u_d independiente e idénticamente distribuida con media cero y varianza σ_u^2 desconocidad Molina (2019). Sea $\hat{\delta}_d$ un estimador directo de δ_d:

$$\hat{\delta}_d = x_d'\beta + u_d + \varepsilon_d,\ d = 1, \ldots, D \tag{2-9}$$

Donde ε_d es el error de muestreo en el área d, estos errores se asumen independientes entre sí e independientes de los efectos de áreas u_d, tienen media 0 y varianza conocida ψ_d.
El mejor predictor BP(*best predictor*) de $\hat{\delta}_d$ que minimiza el error cuadrático medio se define como Rao & Molina (2015):

$$\hat{\delta}_d^{FH} = \phi_d\hat{\delta}_d^{Dir} + (1-\phi_d)x_d'\beta \tag{2-10}$$

El peso para el estimador directo se define:$\phi_d = \frac{\sigma_u^2}{\sigma_u^2+\psi_d} \in (0,1)$. Este peso depende del tamaño muestral del área a través de la varianza ψ_d del estimador Directo y de la bondad de ajuste del modelo sintético medido por σ_u^2 (en otras palabras, la heterogeneidad no explicada entre las áreas). Si σ_u^2 es conocida, β puede estimarse por medio de mínimos cuadrados ponderados.

$$\widetilde{\beta} = (\sum_{d=1}^{D}\phi_d x_d x_d')^{-1}\sum_{d=1}^{D}\phi_d x_d\hat{\delta}_d^{Dir} \tag{2-11}$$

Con esto se obtiene el mejor predictor lineal insesgado(BLUP):

$$\hat{\delta}_d^{BLUP} = \phi_d\hat{\delta}_d^{Dir} + (1-\phi_d)x_d'\widetilde{\beta} \tag{2-12}$$

Cuando σ_u^2 es desconocido, se estima con el cuadrado medio de error del modelo (2-9) por el método REML Molina (2019, p.47) y se obtiene el mejor predictor lineal insesgado empírico (EBLUP):

$$\hat{\delta}_d^{EBLUP} = \hat{\phi}_d\hat{\delta}_d^{Dir} + (1-\hat{\phi}_d)x_d'\hat{\beta} \tag{2-13}$$

donde $\hat{\phi}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2+\psi_d}$ y $\hat{\beta} = (\sum_{d=1}^{D}\hat{\phi}_d x_d x_d')^{-1}\sum_{d=1}^{D}\hat{\phi}_d x_d\hat{\delta}_d^{Dir}$. Para resumir se llamará FH al EBLUP basado en el modelo Fay-Herriot dado en (2-13) y la estimacion del error cuadratico medio para el estimador FH se realiza de la siguiente forma Prasad & Rao (1990):

$$\hat{MSE}(\hat{\delta}_d^{FH}) = g_{1d}(\hat{\sigma}_u^2) + g_{2d}(\hat{\sigma}_u^2) + 2g_{3d}(\hat{\sigma}_u^2) \tag{2-14}$$

donde, $g_{1d}(\hat{\sigma}_u^2) = \frac{\hat{\sigma}_u^2 D_d}{\hat{\sigma}_u^2 + D_d}$, $g_{2d}(\hat{\sigma}_u^2) = \frac{D_d^2}{(\hat{\sigma}_u^2 + D_d)^2} x_d (X' V^{-1} X)^{-1} x_d$ y $g_{3d}(\hat{\sigma}_u^2) = (\frac{2D_d^2}{m(\hat{\sigma}_u^2 + D_d)^3})(\hat{\sigma}_u^4 + 2\hat{\sigma}_u^2 \sum_{d=1}^{D} \frac{D_d}{D} + \sum_{d=1}^{D} \frac{D_d^2}{D})$, con $V = diag(\sigma_u^2 D_1, \ldots, (\sigma_u^2 D_d))$.

Modelo de Fay-Herriot espacial

El estimador Fay Herriot espacial propuesto por Pratesi & Salvati (2008) es una extensión del modelo clásico Fay & Herriot (1979). Al incluir el efecto espacial se puede obtener una mejora en el estimador EBLUP reduciendo tanto la varianza, como el sesgo del estimador (Pratesi & Salvati 2008). Este modelo se caracteriza por tener en cuenta la proximidad geográfica y tiene la siguiente forma:

$$\delta_d^{Dir} = x_d' \beta + Z(I - \rho W)^{-1} u_d + \varepsilon_d, \ d = 1, \ldots, D \tag{2-15}$$

Donde δ_d^{Dir} es el modelo de Fay-Herriot con efectos espaciales para el área d, I es la matriz identidad, ρ es el coeficiente de autocorrelacion espacial que mide la fuerza de las relaciones espaciales entre los efectos aleatorios en las áreas vecinas y W es la matriz de proximidades entre las áreas analizadas, los demás parámetros siguen siendo los mismos del modelo Fay-Herriot clásico. Wawrowski (2016) define el mejor predictor lineal empírico espacial (SEBLUP) como:

$$\hat{\delta}_d^{SEBLUP} = x_d'\hat{\beta} + b_d'\{\hat{\sigma}_u^2[(I-\hat{\rho}W)(I-\hat{\rho}W')]^{-1}\}Z' \times \{diag(\psi_d) + Z\hat{\sigma}_u^2[(I-\hat{\rho}W)(I-\hat{\rho}W')]^{-1}Z'\}^{-1}(\hat{\delta}_d^{Dir} - x_d'\hat{\beta}) \tag{2-16}$$

Donde $\hat{\delta}_d^{SEBLUP}$ es el estimador Fay-Herriot Espacial de δ_d, b_d' es un vector con valor 1 en la i-ésima entrada y 0 en las demás entradas y $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}\hat{\delta}_d^{Dir}$ con $V = diag(\psi_d) + Z\hat{\sigma}_u^2[(I - \hat{\rho}W)(I - \hat{\rho}W')]^{-1}Z'$. Asumiendo normalidad de los efectos aleatorio σ_u^2 y ρ pueden estimarse mediante procedimientos de máxima verosimilitud (ML) y de máxima verosimilitud restringida (REML); en la práctica frecuentemente se estiman por REML dado que el uso de estimadores de ML podría llevar a subestimar la aproximación de MSE Salvati (2004a), Salvati (2004b) y Pratesi & Salvati (2005). El error cuadratico medio para el estimador Fay-Herriot Espacial se puede estimar de la siguiente forma Prasad & Rao (1990):

$$\hat{MSE}(\hat{\delta}_d^{FHS}) \approx g_{1d}(\hat{\sigma}_u^2, \hat{\rho}) + g_{2d}(\hat{\sigma}_u^2, \hat{\rho}) + 2g_{3d}(\hat{\sigma}_u^2, \hat{\rho}) \tag{2-17}$$

donde, $g_{1d}(\hat{\sigma}_u^2, \hat{\rho}) = \mathbf{b}_d'\{\hat{\sigma}_u^2[(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}')]^{-1} - \hat{\sigma}_u^2[(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}')]^{-1}\mathbf{Z}' \times \{diag(\psi) + \mathbf{Z}\hat{\sigma}_u^2[(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}')]^{-1}\mathbf{Z}'\}^{-1}\mathbf{Z}\hat{\sigma}_u^2 \times [(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}')]^{-1}\}\mathbf{b}_d$,

$g_{2d}(\hat{\sigma}_u^2, \hat{\rho}) = (\mathbf{x}_d - \mathbf{b}_d'\hat{\sigma}_u^2[(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}')]^{-1}\mathbf{Z}' \times \{diag(\psi) + \mathbf{Z}\hat{\sigma}_u^2[(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}')]^{-1}\mathbf{Z}'\}^{-1}\mathbf{X}) \times (\mathbf{X}'\{diag(\psi) + \mathbf{Z}\hat{\sigma}_u^2[(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}')]^{-1}\mathbf{Z'}\}^{-1}\mathbf{X})^{-1} \times (\mathbf{x}_d - \mathbf{b}_d'[(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}')]^{-1}\mathbf{Z}' \times \{diag(\psi) + \mathbf{Z}\hat{\sigma}_u^2[(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}')]^{-1}\mathbf{Z}'\}^{-1}\mathbf{X})'$

$g_{3d}(\hat{\sigma}_u^2, \hat{\rho}) = tr\{[\mathbf{b}_d'(\mathbf{C}^{-1}\mathbf{Z}'\mathbf{V}^{-1} + \hat{\sigma}_u^2\mathbf{C}^{-1}\mathbf{z}'(-\mathbf{V}^{-1}\mathbf{Z}\mathbf{C}^{-1}\mathbf{Z}'\mathbf{V}^{-1}))\mathbf{b}_d'(\mathbf{A}\mathbf{Z}'\mathbf{V}^{-1} + \hat{\sigma}_u^2\mathbf{C}^{-1}\mathbf{Z}'$
$(-\mathbf{V}^{-1}\mathbf{Z}\mathbf{A}\mathbf{Z}'\mathbf{V}^{-1}))]\mathbf{V} \times [\mathbf{b}_d'(\mathbf{C}^{-1}\mathbf{Z}'\mathbf{V}^{-1} + \hat{\sigma}_u^2\mathbf{C}^{-1}\mathbf{z}'(-\mathbf{V}^{-1}\mathbf{Z}\mathbf{C}^{-1}\mathbf{Z}'\mathbf{V}^{-1}))\mathbf{b}_d'(\mathbf{A}\mathbf{Z}'\mathbf{V}^{-1} + \hat{\sigma}_u^2\mathbf{C}^{-1}\mathbf{Z}'$
$(-\mathbf{V}^{-1}\mathbf{Z}\mathbf{A}\mathbf{Z}'\mathbf{V}^{-1}))]'\bar{\mathbf{V}}(\hat{\sigma_u^2}, \hat{\rho})\}$

con $\mathbf{C} = [(\mathbf{I} - \hat{\rho}\mathbf{W})(\mathbf{I} - \hat{\rho}\mathbf{W}')]$, $\mathbf{A} = \hat{\sigma}_u^2[-\mathbf{C}^{-1}(2\hat{\rho}\mathbf{W}\mathbf{W}' - 2\mathbf{W})\mathbf{C}^{-1}]$ y $\bar{\mathbf{V}}(\hat{\sigma_u^2}, \hat{\rho})$ la matriz de covarianzas asintótica de $\hat{\sigma}_u^2$ y $\hat{\rho}$ Pratesi & Salvati (2008). $g_{1d}-$ componente encargada del error estimado de ψ, $g_{2d}-$ componente responsable del error estimado de β y $g_{3d}-$ componente encargada del error estimado de σ_u^2 en el modelo EBLUP o σ_u^2 y ρ en el modelo SEBLUP Wawrowski (2016). Para aplicaciones prácticas se obtiene un estimador de MSE siguiendo los resultados de Harville & Jeske (1992) y luego se extiende a modelos con covarianzas generalizadas por Zimmerman & Cressie (1992).

Esta sección del estimador Fay-Herriot espacial fue tomada del articulo Pratesi & Salvati (2008) en el cual pueden profundizar más detalles acerca de otros aspectos como el algoritmo iterativo y la modificación del MSE al considerar ML como método de estimación. Para poder utilizar la metodología SAE en la presente investigación primero se debe aplicar la teoría de respuesta al ítem la cual nos permite establecer una relación entre los resultados de la prueba y los rasgos latente que se quieren medir.

2.2. Teoría de respuesta al ítem

La Teoría de Respuesta al Ítem (TRI) pretende medir las características latentes a través de modelos matemáticos, en los que se asume la existencia de una relación lineal entre la respuesta de una persona y su habilidad para ser expresada en términos de probabilidad (Attorresi et al. 2009), por lo que los avances de los últimos años de esta teoría han permitido que se use en la evaluación de rendimiento académico y otras mediciones de competencia.

De acuerdo con Matas Terrón et al. (2010) la fortaleza de esta teoría se da bajo los supuestos de que existen uno o más rasgos latentes que son la explicación de la respuesta de una persona a una prueba o ítem, que dicho ítem mide un solo rasgo (unidimensionalidad) y, que la respuesta no influye en la respuesta que otro da (independencia local).

En concordancia con lo anterior, los modelos TRI se clasifican según el número de parámetros que se estiman entre la dificultad del ítem, la discriminación del ítem y el azar. Así, el modelo de Rasch (un parámetro) estima solo la dificultad del ítem, el modelo de dos parámetros estima ese y, además, la discriminación del ítem y, el modelo de Birnbaum (tres parámetros) los incluye todos. Este último es el modelo que se aplicó en esta investigación.

Modelo de Birnbaum o modelo de tres parámetros (3PL)

Birnbaum (1958) presenta el modelo de tres parámetros asumiendo que la función característica, la cual es una función matemática que describe la relación entre la probabilidad de una respuesta correcta en un ítem de un examen y el nivel de habilidad o rasgo que posee el examinado, está dada por un enlace logístico:

$$P_i(\varepsilon_{ij} = 1|\theta_i, a_j, b_j, c_j) = c_j + (1 - c_j)\frac{e^{Da_j(\theta_i - b_j)}}{1 + e^{Da_j(\theta_i - b_j)}} \tag{2-18}$$

Donde a_j mide la potencia discriminatoria del ítem $a_i \geq 0$, b_j mide la dificultad del ítem $(-\infty \leq b_j \leq \infty)$, c_j, mide el azar y D es un factor de escala constante para cada ítem y toma valores 1 o $\frac{17}{10}$, donde θ es el nivel de rasgo latente que se desea medir con el ítem $(-\infty \leq \theta \leq \infty)$ y ε_{ij} es una variable dicotómica que toma valor de 1 cuando el individuo i responde correctamente el ítem j y 0 si responde de forma incorrecta.

A medida que la dificultad tiende a infinito(∞) la probabilidad de respuesta correcta tiende a cero(0), lo que implica que si el ítem es extremadamente difícil es poco probable que los individuos respondan correctamente al ítem, independientemente de su nivel de habilidad y en caso contrario cuando la dificultad tiende a ($-\infty$)menos infinito la probabilidad de respuesta correcta tiende a 1 dado que el ítem se vuelve extremadamente fácil.

Para aplicar el modelo 3PL se aborda la metodología de valores plausibles, dado que por la naturaleza de las pruebas PISA las respuestas presentan datos faltantes y el modelo 3PL genera estimaciones con errores de medición altos en presencia de datos faltantes (Von Davier, Gonzalez & Mislevy 2009). En este estudio el modelo $3Pl$ se utilizó tan solo para estimar las habilidades de los estudiantes de cada institución o país seleccionado en la muestra y, a partir de este, implementar el modelo Fay-Herriot espacial y poder estimar $\hat{\delta}_d^{FHS}$.

2.3. Valores plausibles

Los valores plausibles se obtienen de una muestra aleatoria con distribución condicional en la que exista sujeción al error de medición al que se aplique un método de imputación múltiple (Rubin 1987). El caso particular de estudio que ocupa esta investigación tiene todas las características dadas para este tipo de valores.

Esto es, en las pruebas estandarizadas PISA existen algunos ítems que los estudiantes no alcanzan a responder, por lo que la medición individual del desempeño de los estudiantes está sujeta a error de medición, entonces para disminuir este error, se implementa un método de imputación múltiple del que se general K valores plausibles que representan la distribución del desempeño del estudiante. Estos valores se obtienen como una muestra aleatoria de la distribución condicional de la habilidad para cada estudiante, esto es:

$P(\theta_i|A_j, a_j, b_j, c_j, O_i, \alpha, \sigma^2)$ donde A son las respuestas del evaluado a los ítems j, (a, b, c) son los parámetros estimados para el modelo, O_i es la información auxiliar disponible para cada evaluado, α es la relación de la información auxiliar con la habilidad de los evaluados en la población y σ^2 es la varianza común de todos los evaluados (Rubin 1987).

Para definir la distribución condicional de la habilidad se necesita conocer los valores de α y σ^2 por lo que se incurre a un modelo lineal para la habilidad condicional dado por Dempster et al. (1977) en el cual los parámetros del modelo se estiman por medio de un algoritmo Esperanza-Maximización:

$$P(\theta_i|A_j, a_j, b_j, c_j, O_i, \alpha, \sigma^2) \approx P(A_j|a_j, b_j, c_j, O_i, \alpha, \sigma^2)P(\theta_i|O_i, \alpha, \sigma^2) \tag{2-19}$$

$$\approx P(A_j|\theta_i, a_j, b_j, c_j)P(\theta_i|O_i, \alpha, \sigma^2) \tag{2-20}$$

Para realizar las estimaciones basada en los datos plausibles se realizan estimaciones independientes de la estadística de interés para cada uno de los valores plausibles y se promedian:

$$\hat{\gamma} = \frac{1}{K}\sum_{k=1}^{K}\hat{\gamma}_k \tag{2-21}$$

El error estándar asociado a la estimación esta dado por:

$$EE(\hat{\gamma}) = \sqrt{\frac{1}{K}\sum_{k=1}^{K}Var_m(\hat{\gamma}_k) + (1+\frac{1}{K})\frac{1}{(K-1)}\sum_{k=1}^{K}(\hat{\gamma}_k - \hat{\gamma})^2} \tag{2-22}$$

$Var_m(\hat{\gamma}_k)$ es la varianza del estimador calculada con el K-ésimo conjunto de valores plausibles y de acuerdo con el diseño muestral (Von Davier et al. 2009).

3 Objetivos

3.1. Objetivo general

- Implementar una línea de análisis y evaluación de las pruebas PISA que incluya el factor espacial en la metodología de estimación de áreas pequeñas (SAE) y teoría de respuesta al ítem (TRI) donde se utiliza imputación múltiple a través de valores plausibles propuesta por Tellez (2020)

3.2. Objetivos específicos

- Incluir el factor espacial en el modelo Fay-Herriot para el estudio del rendimiento medio de los estudiantes en las pruebas PISA de 2018 en las áreas de lectura y matemáticas.
- Comparar las estimaciones obtenidas para las pruebas PISA 2018 con la metodología propuesta en este documento y la metodología Tellez (2020) en términos de error cuadrático y error estándar relativo.
- Producir estimaciones del rendimiento medio para las pruebas PISA 2018 de países no participantes que puedan servir como fundamento en el diseño de políticas públicas educativas en cada país.

4 Metodología

El objetivo principal de esta investigación es determinar el promedio de habilidad de los estudiantes en el rendimiento de las pruebas PISA, específicamente en las áreas de lectura y matemáticas. Definimos $y = h(\eta)$ como la representación de una estadística o parámetro de interés, donde la estimación de dicha habilidad en presencia de datos faltantes, se realizará a través del modelo 3PL basado en la metodología de valores plausibles, y las estimaciones en áreas pequeñas se hará con el modelo de Fay-Herriot espacial.

El modelo 3PL es ampliamente utilizado en la estimación de habilidades de los estudiantes. Sin embargo, presenta la particularidad de generar estimaciones con altos errores de medición cuando hay datos faltantes (Von Davier et al. 2009). En el contexto de las pruebas PISA, que son pruebas estandarizadas, es común que los estudiantes no respondan todos los ítems, lo que resulta en datos faltantes, lo que hace que la medición del desempeño individual esté sujeta a un error de medición.

Con el fin de reducir el error de medición y obtener estimaciones insesgadas a nivel de dominio, se implementa un método de imputación múltiple. Este método generará K valores plausibles que representen la distribución del desempeño de un estudiante, teniendo en cuenta la incertidumbre asociada a los datos faltantes.

Ahora bien, los valores plausibles se obtienen como una muestra aleatoria de la distribución condicional de las habilidades (θ) en el dominio, denotada por $P(\theta|A_{ij}, O)$. Tellez (2020) define probabilidad total como:

$$P(\theta|A_{ij}, O) = P(A_{ij}|O, \theta)P(\theta|O) \tag{4-1}$$

bajo el supuesto que los A_{ij} son condicionalmente independientes entre sí y los A_{ij} son condicionalmente independientes de O se tiene que:

$$P(\theta|A_{ij}, O) = P(A_{ij}|\theta)P(\theta|O) \tag{4-2}$$

Donde $P(A_{ij}|\theta)$ se obtiene de un modelo logístico de 3PL y $P(\theta|O)$ se aproxima por una distribución normal de la forma:

$$P(\theta|O) = \Phi(\theta, O\Delta, \Sigma) \tag{4-3}$$

Con θ el vector de rasgo latente de toda la población, $\Phi(*)$, la función de probabilidad de una distribución normal y Δ, Σ son parámetros que se estiman mediante el algoritmo de Esperanza-Maximización. Se generan K valores plausibles (realizaciones aleatorias) para cada individuo i en el dominio d, utilizando la ecuación (4-2). Estos valores plausibles se denotan como θ_{dik}^{vp}, con $k = 1, \ldots, K$.

A continuación, utilizando el k-ésimo valor plausible en el dominio d, se estima una estadística de interés $\hat{\delta}_{dk}^{Dir} = \frac{\theta_{dik}^{vp}}{\pi_{dik}}$ y su respectiva varianza mediante el estimador directo propuesto por Horvitz & Thompson (1952), utilizando las probabilidades de inclusión inducidas por el diseño $P(*)$.

Para obtener una estimación de δ_d se promedian las estimaciones resultante de los K modelos,

$$\hat{\delta}_d^{Dir} = \frac{1}{K} \sum_{k=1}^{K} \hat{\delta}_{dk}^{Dir} \tag{4-4}$$

Se calcula el error cuadrático medio de las estimaciones, Como un promedio de los errores cuadráticos medio ($\hat{Var}_k$) calculados para cada k-ésima estimación. Primero se calcula el $\hat{Var}_k$ para ara cada k-ésima estimación, por medio de la ecuación:

$$\hat{Var}_k(\hat{\delta}_{dk}^{Dir}) = \sum \sum_{s} \frac{\Delta_{dilk}}{\pi_{dilk}} \frac{\theta_{dik}^{vp}}{\pi_{dik}} \frac{\theta_{dlk}^{vp}}{\pi_{dlk}} \tag{4-5}$$

con $\Delta_{dilk} = \pi_{dilk} - \pi_{dik}\pi_{dlk}$.
Luego se ponderan de la siguiente forma para lograr obtener el error cuadrático medio de las estimaciones:

$$\hat{MSE}(\hat{\delta}_d^{Dir}) = \frac{1}{K} \sum_{k=1}^{K} \hat{Var}_k(\hat{\delta}_{dk}^{Dir}) + (1 + \frac{1}{K}) \frac{1}{(K-1)} \sum_{k=1}^{K} (\hat{\delta}_{dk}^{Dir} - \hat{\delta}_d^{Dir})^2 \tag{4-6}$$

Estas estimaciones, junto con las variables auxiliares, se incorporan en el modelo de Fay-Herriot espacial para mejorar la precisión. El modelo de Fay-Herriot espacial tiene en cuenta el factor espacial y la información externa asociada a la variable objetivo. Para los dominios que no fueron seleccionados, se estima la característica de interés utilizando modelos de estimación en áreas pequeñas, donde $\hat{\delta}_d^{SEBLUP} = x_d'\hat{\beta}$. En este enfoque, los dominios seleccionados se tratan como áreas pequeñas, y se utiliza el modelo de estimación en áreas pequeñas de Fay-Herriot espacial propuesto por Pratesi & Salvati (2008), el cual nos permite enlazar la información auxiliar y las estimaciones además del factor espacial y viene dado por:

$$\hat{\delta}_d^{SEBLUP} = x_d'\hat{\beta} + b_d'\{\hat{\sigma}_u^2[(I-\hat{\rho}W)(I-\hat{\rho}W')]^{-1}\}Z' \times \{diag(\psi_d) + Z\hat{\sigma}_u^2[(I-\hat{\rho}W)(I-\hat{\rho}W')]^{-1}Z'\}^{-1}(\hat{\delta}_d^{Dir} - x_d'\hat{\beta}) \tag{4-7}$$

Donde $\hat{\delta}_{dk}^{SEBLUP}$ es el estimador espacial para el área d, I es la matriz identidad, ρ es el coeficiente de autocorrelación espacial, permite medir la fuerza de las relaciones espaciales

entre los efectos aleatorios en las áreas vecinas. Mientras que W es la matriz de proximidades entre las áreas analizadas, $\hat{\beta}$ es el vector de coeficientes común en todas las áreas. b'_d es un vector con valor 1 en la d-ésima entrada y 0 en las demás entradas y $V = diag(\psi_d) + Z\hat{\sigma}_u^2[(I - \hat{\rho}W)(I - \hat{\rho}W')]^{-1}Z'$.

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}\hat{\delta}_d^{Dir} \tag{4-8}$$

Se calcula el error cuadrático medio de las estimaciones, bajo el modelo dado por Prasad & Rao (1990):

$$\hat{MSE}(\hat{\delta}_d^{SEBLUP}) \approx g_{1d}(\hat{\sigma}_u^2, \hat{\rho}) + g_{2d}(\hat{\sigma}_u^2, \hat{\rho}) + 2g_{3d}(\hat{\sigma}_u^2, \hat{\rho})$$

5 Estudio de simulación y aplicación a los datos de la prueba PISA 2018

5.1. Estudio de simulación

Se planteó este estudio con el fin de evaluar en diferentes escenarios el estimador propuesto (4-4) y compararlo con los distintos estimadores ya existentes como el Horvitz-Thompson, Compuesto y el EBLUP propuesto por Tellez (2020). Se compararán por medio del Sesgo Relativo $SB_d = \frac{\gamma_d - \hat{\gamma}_d}{\hat{\gamma}_d}$, el Error Estándar Relativo $EER_d = \frac{\sqrt{\hat{MSE}(\hat{\gamma}_d)}}{\gamma_d} \times 100$ y $\%\ Mejor =$ porcentaje de dominios en el que el estimador obtuvo menor EER_d.

Así mismo, se evaluarán los estimadores en diferentes escenarios donde cada uno de estos últimos está conformado por poblaciones con diferentes características relevantes para el análisis y cuya simulación se realizó combinando las metodologías de Tellez (2020) y Pratesi & Salvati (2008) de la siguiente manera:

1. Simular una población de N = 60.000 estudiantes con 75 ítems cada uno y D = 79 dominios y 2 variables auxiliares asociadas a la habilidad de cada estudiante mediante la metodología propuesta por Tellez (2020):

   a) El primer paso es definir la cantidad de estudiantes que habrá en cada uno de los 79 dominio. Para esto, se asigna cada estudiante a uno de los 79 dominios con una probabilidad generada mediante una distribución beta. luego de esto, se cuentan cuantos estudiantes fueron asignados a cada dominio. Esto con el fin de garantizar que los dominios no tengan la misma cantidad de estudiantes. Respetando que la suma de los tamaños de los dominios es N = 60.000.

   b) Para generar las habilidades de los estudiantes en cada dominio, se utiliza la distribución normal con promedio igual a la secuencia mencionada en el paso 3 y una varianza aleatoria (varía entre 0.3 y 0.7 por estudiante).

   c) Se crean las 2 variables auxiliares asociadas a la habilidad creada en el paso anterior, aumentando la varianza del componente aleatorio de un modelo lineal de la forma $X_{di} = \beta_{di}\theta_{di} + x_{di} + e_{di}$ donde θ_{di} es la habilidad del i-ésimo estudiante del dominio d, β_{di} es una constante generada de una distribución uniforme $U(5, 6)$, x_{di} se obtienen de una distribución normal multivariada con vector de media

$(150, 150)$ y matriz de varianza $\begin{bmatrix} 1 & 0{,}5 \\ 0{,}5 & 1 \end{bmatrix}$ donde e_{di} se obtiene de una distribución $N(0, \sigma_i^2)$, $\sigma_i^2 = 0,5$ si la correlación entre las variables auxiliares y la habilidad es alta ($> 80\,\%$), $\sigma_i^2 = 10$ si la correlación es media($60\,\%, 80\,\%$) y $\sigma_i^2 = 30$ si la correlación es baja($< 60\,\%$).

d) A cada estudiante se induce un 10 %, 20 % o 30 % de datos faltantes de manera aleatoria, según sea el escenario, usando la función *sample* para seleccionar aleatoriamente las respuestas que serán eliminadas.

2. Estimar diez valores plausibles para cada estudiante que repliquen la habilidad creada en el paso (1B), Usando la función *simdata* de la librería *mirt* mediante un modelo 3PL:

   a) Se estiman los parámetros del modelo 3PL utilizando el algoritmo EM. Esto se hace mediante la función *tam.mml.3pl* de la librería TAM.

   b) Se generan los diez valores plausibles utilizando la función *tam.pv* de la librería TAM.

3. Selección de la muestra:

   a) Se calcula el tamaño muestral por medio de la función *ss4p* de la librería *samplesize4surveys*, poniendo como parámetros *N=60000*, *p=0.5*, *conf=0.95*, *delta=0.04* y e*rror=cve*. Obteniendo como resultado un tamao muestral total de *n=619* el cual se distribuye proporcional al tamaño del dominio para así obtener el tamaño de muestreo en cada uno de los 79 dominios.

   b) Se realiza un diseño de muestro aleatorio simple, en cada dominio.

4. Se realiza la estimación directa ($\hat{\gamma}_d^{Dir}$), el estimador compuesto ($\hat{\gamma}_d^{Comp}$), para cada dominio seleccionado, junto con su desviación estándar.

5. Se crea la matriz estructural de los datos W aleatoriamente con una estructura Queen de la siguiente forma:

   a) Se toma un numero inicial de vecinos para cada dominio como una muestra aleatoria simple de valores enteros entre 1 y 5. Luego se generan un vector de repeticiones de 1 y 0 hasta tener un vector de tamaño 79 (dominios), la cantidad de veces que se repite el 1 es el número de vecinos seleccionados anteriormente y la cantidad de veces que se repite el cero es 79 menos la cantidad vecinos.

   b) Se genera una variable uniforme (0,1) para cada entrada del vector creado en el paso 1 y se organiza de menor a mayor con el fin de aleatorizar que dominios serán vecinos. Se va llenando la matriz C colocando cada vector generado como una fila hasta obtener una matriz cuadrada de 79×79.

c) De la matriz C, todas las entradas por debajo de la diagonal principal y las de la diagonal principal se convierten en 0 obteniendo la matriz A, de esta matriz resultante se obtiene la matriz B como la suma de $A' + A$.

d) Se obtiene la matrix W normalizando cada entrada de la matriz B de la forma $\frac{B_{ij}}{\sum_{i=1}^{79} B_i}$. Un ejemplo de la matriz resultante B es:

	d_1	d_2	d_3	d_4	d_5
d_1	0	0,5	0,5	0	0
d_2	0,33	0	0,33	0	0,33
d_3	0,33	0,33	0	0,33	0
d_4	0	0	0,5	0	0,5
d_5	0	0,5	0	0,5	0

e) Los pasos del (5a) hasta el (5d) se repiten hasta obtener una convergencia de $\rho = (0, \pm 0,25, \pm 0{,}5, \pm 0{,}75)$ en el modelo SEBLUP ajustado con las estimaciones del estimador directo y compuesto.

6. Calcular los sesgos relativos (SB_d) y los errores estándares relativos (EER_d) por dominio, para cada uno de los estimadores ($\hat{\gamma}_d^{SEBLUP}$), ($\hat{\gamma}_d^{EBLUP}$), ($\hat{\gamma}_d^{Dir}$), ($\hat{\gamma}_d^{Comp}$) y calcular:

a) $SBP = \frac{\sum_{d=1}^{D} SB_d}{D}$(Sesgo relativo promedio)

b) $EERP = \frac{\sum_{d=1}^{D} EER_d}{D}$(Error estándar relativo promedio)

c) $\%\ Mejor = \frac{\sum_{d=1}^{D} \lambda_i}{D}$ (λ_i una variable binomial que toma valor de 1 cuando $EER_d(i) < EER_d(j)$ para $i \neq j$)

7. Repetir los pasos del 3 al 5 para 500 muestras aleatorias y calcular:

a) $\overline{SBR} = \frac{\sum_{r=1}^{500} SBP_r}{500}$(Promedio de los sesgos relativos promedios)

b) $\overline{EERP} = \frac{\sum_{r=1}^{500} EERP_r}{500}$(Promedio del error estándar relativo promedio)

c) $\overline{\%\ Mejor} = \frac{\sum_{r=1}^{500} \%\ Mejor_r}{500}$(Porcentaje promedio de dominios donde el estimador tuvo el menor EER)

En este estudio se simuló la característica de interés (θ_{dj}, habilidad) de forma heterogénea, es decir, que la media y la varianza de la habilidad en los dominios sean diferentes, con el fin de evaluar el comportamiento de los estimadores. Se realizan las comparaciones con distintas correlaciones entre las variables auxiliares y la habilidad, diferentes niveles de correlación espacial entre los dominios, para así estudiar el efecto de cada uno de estos sobre los valores de $\overline{SBR}$, $\overline{EERP}$ y $\overline{\%\ Mejor}$.

5.1.1. Comparación de los estimadores

Se destaca la consistencia del estudio de simulación debido a que, en todos los escenarios, a medida que la correlación entre la habilidad y las variables auxiliares disminuye, el $\overline{EERP}$ aumenta y a medida que la correlación espacial ρ se acerca a cero el $\%\ Mejor$ del estimador SEBLUP disminuye.

La tabla 5-1 muestra que el estimador EBLUP espacial(SEBLUP) tiene un mejor desempeño en términos de $\overline{EERP}$ en todos los escenarios, en comparación con los estimadores Directo y Compuesto. Cuando la correlación entre la habilidad y las variables auxiliares (Cor) disminuye, se observa un aumento en todos los valores promedio del error cuadrático medio $\overline{EERP}$.

Cor	ρ	$\overline{EERP}\hat{\gamma}_d^{Dir}(\%)$	$\overline{EERP}\hat{\gamma}_d^{Comp}(\%)$	$\overline{EERP}\hat{\gamma}_d^{EBLUP}(\%)$	$\overline{EERP}\hat{\gamma}_d^{SEBLUP}(\%)$	$\overline{SBR}\hat{\gamma}_d^{EBLUP}(\%)$	$\overline{SBR}\hat{\gamma}_d^{SEBLUP}(\%)$	$\overline{\%\ Mejor}^{EBLUP}$	$\overline{\%\ Mejor}^{SEBLUP}$
A	-0.75	2.476	1.44	0.736	0.681	0.095	0.088	16 %	84 %
	-0.5	2.476	1.44	0.734	0.719	0.085	0.085	21 %	79 %
	-0.25	2.476	1.44	0.763	0.762	0.102	0.103	44 %	56 %
	0	2.476	1.44	0.74	0.763	0.058	0.038	54 %	46 %
	0.25	2.476	1.44	0.723	0.726	0.073	0.075	39 %	61 %
	0.5	2.476	1.44	0.723	0.721	0.078	0.081	25 %	75 %
	0.75	2.476	1.44	0.728	0.676	0.065	0.075	19 %	81 %
M	-0.75	2.492	2.513	1.186	1.164	-0.031	-0.046	38 %	62 %
	-0.5	2.492	2.513	1.213	1.2	-0.023	-0.031	40 %	60 %
	-0.25	2.492	2.513	1.224	1.221	-0.035	-0.038	39 %	61 %
	0	2.492	2.513	1.218	1.219	-0.043	-0.046	42 %	58 %
	0.25	2.492	2.513	1.241	1.23	-0.037	-0.036	33 %	67 %
	0.5	2.492	2.513	1.231	1.221	-0.03	-0.024	33 %	67 %
	0.75	2.492	2.513	1.206	1.154	-0.031	-0.025	33 %	68 %
B	-0.75	2.518	2.678	1.361	1.253	-0.057	-0.066	18 %	82 %
	-0.5	2.518	2.678	1.353	1.324	-0.046	-0.055	29 %	71 %
	-0.25	2.518	2.678	1.376	1.37	-0.014	-0.017	43 %	57 %
	0	2.518	2.678	1.379	1.379	-0.056	-0.056	53 %	47 %
	0.25	2.518	2.678	1.355	1.349	-0.039	-0.039	40 %	60 %
	0.5	2.518	2.678	1.354	1.327	-0.028	-0.029	28 %	73 %
	0.75	2.518	2.678	1.365	1.247	-0.024	-0.023	20 %	80 %

Tabla 5-1: Missing 10 %

En el caso del estimador Directo, esto se debe a que al estimar los valores plausibles que replican la habilidad, una disminución en la correlación de las variables auxiliares resulta en una mayor variabilidad de los valores plausibles.

En los demás estimadores, el aumento en el $\overline{EERP}$ se debe tanto a lo mencionado anteriormente como al hecho de que utilizan información auxiliar para generar las estimaciones.

Cuanto mayor sea la relación entre la variable a estimar y la información auxiliar, menor será el $\overline{EERP}$ de las estimaciones. Por el contrario, cuando existe una menor relación, el $\overline{EERP}$ tiende a ser mayor (Gutiérrez 2016, Särndal et al. 2003). Es importante destacar que el estimador Compuesto, en los niveles de correlación (Cor) Media y Baja, muestra un EERP promedio mayor que el estimador Directo.

En cuanto a las correlaciones espaciales entre los dominios ($\rho = \pm 0{,}75, \pm 0{,}5$), se observa que el estimador SEBLUP presenta un $\overline{EERP}$ menor que el estimador EBLUP, y el ($\overline{\%\ Mejor}^{SEBLUP}$) es significativamente mayor que el ($\overline{\%\ Mejor}^{EBLUP}$). Esto indica que la introducción del estimador SEBLUP cuando existe una correlación espacial alta o media ayuda a mejorar la eficiencia de las estimaciones en términos de $\overline{EERP}$. Sin embargo, cuando la correlación espacial es moderada ($\rho \pm 0{,}25$), no siempre el estimador SEBLUP presenta un $\overline{EERP}$ menor que el EBLUP. En el caso de una correlación espacial nula ($\rho = 0$), Pratesi & Salvati (2008) afirma que los estimadores EBLUP y SEBLUP son teóricamente iguales, aunque en la simulación pueden presentar diferencias leves. Esto puede deberse a un error en el proceso de estimación del error cuadrático medio por Monte Carlo, que no se ha tenido en cuenta (Pratesi & Salvati 2008).

En todos los niveles de correlación espacial, el estimador SEBLUP presenta un $\overline{\%\ Mejor}^{SEBLUP}$ mayor que cero (0 %). Esto significa que, sin importar el nivel de correlación espacial, en algunos dominios el estimador SEBLUP presenta el menor error cuadrático medio. Dado que en la práctica no se reporta el $\overline{EERP}$ sino el EER de cada dominio, incluir el efecto de área en las estimaciones ayuda a reducir el error cuadrático medio de algunas estimaciones. El número de estimaciones en las que el error cuadrático medio disminuye está determinado por la estructura de correlación W y el nivel de correlación espacial ρ.

Cor	ρ	$\overline{EERP}_{\hat{\gamma}_d^{Dir}}(\%)$	$\overline{EERP}_{\hat{\gamma}_d^{Comp}}(\%)$	$\overline{EERP}_{\hat{\gamma}_d^{EBLUP}}(\%)$	$\overline{EERP}_{\hat{\gamma}_d^{SEBLUP}}(\%)$	$\overline{SBR}_{\hat{\gamma}_d^{EBLUP}}(\%)$	$\overline{SBR}_{\hat{\gamma}_d^{SEBLUP}}(\%)$	$\%\ Mejor^{EBLUP}$	$\%\ Mejor^{SEBLUP}$
A	-0.75	2.473	1.457	0.753	0.694	0.099	0.104	15 %	85 %
	-0.5	2.473	1.457	0.749	0.726	0.087	0.089	19 %	81 %
	-0.25	2.473	1.457	0.747	0.741	0.097	0.104	42 %	58 %
	0	2.473	1.457	0.761	0.763	0.063	0.065	54 %	46 %
	0.25	2.473	1.457	0.746	0.744	0.075	0.081	35 %	65 %
	0.5	2.473	1.457	0.743	0.722	0.081	0.092	19 %	81 %
	0.75	2.473	1.457	0.751	0.695	0.075	0.08	18 %	83 %
M	-0.75	2.49	2.511	1.18	1.165	-0.033	-0.048	38 %	62 %
	-0.5	2.49	2.511	1.231	1.213	-0.009	-0.013	38 %	63 %
	-0.25	2.49	2.511	1.211	1.199	-0.029	-0.028	38 %	62 %
	0	2.49	2.511	1.222	1.207	-0.057	-0.057	39 %	61 %
	0.25	2.49	2.511	1.247	1.236	-0.036	-0.034	35 %	65 %
	0.5	2.49	2.511	1.257	1.243	-0.025	-0.022	33 %	67 %
	0.75	2.49	2.511	1.224	1.157	0.013	0.007	28 %	72 %
B	-0.75	2.512	2.673	1.357	1.273	-0.044	-0.058	23 %	78 %
	-0.5	2.512	2.673	1.375	1.349	-0.021	-0.03	28 %	72 %
	-0.25	2.512	2.673	1.364	1.358	-0.034	-0.033	44 %	56 %
	0	2.512	2.673	1.406	1.409	-0.048	-0.048	58 %	42 %
	0.25	2.512	2.673	1.366	1.36	-0.039	-0.035	41 %	59 %
	0.5	2.512	2.673	1.368	1.337	-0.038	-0.033	24 %	76 %
	0.75	2.512	2.673	1.377	1.264	-0.023	-0.012	18 %	82 %

Tabla 5-2: Missing 20 %

En la tabla 5-2, a pesar de que el porcentaje de datos faltantes aumenta del 10 % al 20 %, todavía se observan resultados similares a los de la tabla 5-1, Sin embargo, al comparar escenario por escenario entre ambas tablas, se evidencia que los valores $\overline{EERP}$ en la tabla 5-2 son mayores en comparación con la 5-1, pero la relación entre los estimadores se mantiene igual. El aumento de los $\overline{EERP}$se debe a que al incrementar el porcentaje de datos faltantes, aumenta la variabilidad entre los valores plausibles, lo cual afecta en menor medida al estimador directo pero sí impacta a los estimadores que utilizan información auxiliar.

En cuanto a los estimadores EBLUP y SEBLUP, presentan comportamientos similares en cuanto al promedio del sesgo relativo ($\overline{SBR}$) tanto en la tabla 5-2 como en la tabla 5-1. Aunque el $\overline{SBR}$ del estimador SEBLUP es ligeramente mayor que el del EBLUP en ambos escenarios, la magnitud de estos $\overline{SBR}$ es tan pequeña que se consideran despreciables.

Cor	ρ	$\overline{EERP}\hat{\gamma}_d^{Dir}(\%)$	$\overline{EERP}\hat{\gamma}_d^{Comp}(\%)$	$\overline{EERP}\hat{\gamma}_d^{EBLUP}(\%)$	$\overline{EERP}\hat{\gamma}_d^{SEBLUP}(\%)$	$\overline{SBR}\hat{\gamma}_d^{EBLUP}(\%)$	$\overline{SBR}\hat{\gamma}_d^{SEBLUP}(\%)$	$\%\ Mejor^{EBLUP}$	$\%\ Mejor^{SEBLUP}$
A	-0.75	2.473	1.482	0.764	0.711	0.1	0.101	16 %	84 %
	-0.5	2.473	1.482	0.751	0.731	0.083	0.087	19 %	81 %
	-0.25	2.473	1.482	0.757	0.756	0.094	0.102	44 %	56 %
	0	2.473	1.482	0.773	0.775	0.053	0.055	53 %	47 %
	0.25	2.473	1.482	0.759	0.756	0.07	0.077	35 %	65 %
	0.5	2.473	1.482	0.76	0.742	0.071	0.078	19 %	81 %
	0.75	2.473	1.482	0.769	0.71	0.075	0.077	18 %	83 %
M	-0.75	2.489	2.51	1.209	1.171	-0.005	-0.025	33 %	68 %
	-0.5	2.489	2.51	1.228	1.22	-0.027	-0.032	39 %	61 %
	-0.25	2.489	2.51	1.245	1.243	-0.05	-0.053	41 %	59 %
	0	2.489	2.51	1.237	1.224	-0.054	-0.054	41 %	59 %
	0.25	2.489	2.51	1.239	1.228	-0.017	-0.017	35 %	65 %
	0.5	2.489	2.51	1.246	1.228	-0.034	-0.037	33 %	67 %
	0.75	2.489	2.51	1.248	1.182	-0.025	-0.03	28 %	73 %
B	-0.75	2.509	2.671	1.393	1.286	-0.038	-0.052	16 %	84 %
	-0.5	2.509	2.671	1.396	1.365	-0.016	-0.02	27 %	73 %
	-0.25	2.509	2.671	1.392	1.388	-0.022	-0.022	44 %	56 %
	0	2.509	2.671	1.414	1.415	-0.05	-0.051	57 %	43 %
	0.25	2.509	2.671	1.382	1.377	-0.024	-0.024	39 %	61 %
	0.5	2.509	2.671	1.349	1.325	-0.04	-0.033	28 %	72 %
	0.75	2.509	2.671	1.401	1.283	-0.023	-0.016	18 %	82 %

Tabla 5-3: Missing 30 %

Dado que el porcentaje de datos faltantes es alto (30 %), se espera que la información auxiliar sea de calidad y altamente correlacionada con la variable de interés. En los casos donde el nivel de correlación lineal es bajo, los estimadores aún presentan sesgos despreciables, lo cual indica que la metodología de estimación mediante valores plausibles es efectiva y contribuye a la reducción del $\overline{EERP}$ y $\overline{SBR}$. En la tabla 5-3, ocurre algo similar a las tablas 5-1 y 5-2. Debido al aumento en el porcentaje de datos faltantes, los estimadores Compuesto, EBLUP y SEBLUP presentan un $\overline{EERP}$ mayor en comparación con las tablas 5-1 y 5-2.

Como se evidencia en el estudio de simulación, los estimadores mostraron caracteristicas favorables, especialmente los estimadores de áreas pequeñas (EBLUP y SEBLUP). Además, se mostró que el estimador SEBLUP es una buena opción para la estimación cuando la variable de interés presenta correlación espacial, ya sea alta, media o moderada.

Es importante destacar que todo el estudio de simulación se realizó en el software R.

5.2. Aplicación a los datos de la prueba PISA

En este capítulo, se llevará a cabo la estimación del promedio de habilidades de los estudiantes en las pruebas de lectura y matemáticas del programa PISA en el año 2018. Se compararán las estimaciones utilizando la metodología propuesta en este estudio, la metodología oficialmente empleada por la OCDE y la metodología propuesta por Tellez (2020), con el objetivo de analizar y comparar dichas estimaciones para los países que participaron en estas pruebas. Las estimaciones se evaluarán en términos de precisión. Además, se eliminará el 20 % de los países para realizar estimaciones y medir la capacidad predictiva del modelo propuesto, utilizando el sesgo relativo de la predicción tomando como valor de referencia el dato publicado por la OCDE.
Según la OCDE, el programa PISA evalúa los conocimientos y habilidades fundamentales para la participación plena en las sociedades modernas. Esto se realiza a través de tres materias principales: ciencia, lectura y matemáticas. En el año 2018, aproximadamente 600,000 estudiantes de 79 países y economías participantes presentaron las pruebas, lo que representa una muestra representativa de 32 millones de jóvenes de 15 años en los 79 países y economías participantes (OCDE, 2016). (OECD 2016).

Los datos de las pruebas PISA utilizados en esta investigación se obtuvieron de la OCDE, específicamente de su base de datos de 2018, disponible en el enlace: `https://www.oecd.org/pisa/data/2018database/`. Estos datos incluyen los valores plausibles para las distintas pruebas, particularmente aquellos relacionados con la prueba de lectura, que fue la asignatura principal evaluada en 2018, y los correspondientes a la prueba de matemáticas. Dado que la metodología de áreas pequeñas requiere información auxiliar para realizar las estimaciones, se tomaron en cuenta variables auxiliares relacionadas con la infraestructura, servicios básicos, inversión en educación, aspectos sociodemográficos y economía de los países. Las variables auxiliares consideradas son las siguientes:

- Desempleo total (% de participación total en la fuerza laboral - estimación nacional).
- Artículos en publicaciones científicas y técnicas.
- Gasto en investigación y desarrollo (% del PIB).
- Gasto público en educación, total (% del PIB).
- Índice de Gini.
- Personas que utilizan al menos servicio básico de agua potable (% de población).
- Acceso a la electricidad (% de población).

- Servidores de internet seguros (por cada millón de personas).
- Personas que usan Internet (% de la población).
- Población rural (% de la población total).
- Investigadores dedicados a investigación y desarrollo (por cada millón de personas).
- Tasa de incidencia de la pobreza, sobre la base de $\$1,90$ por día (2011 PPA) (% de la población).
- Densidad de población (personas por kilómetro).
- PIB (Producto interno bruto)

Se descartaron variables como "Artículos en publicaciones científicas"debido a su alta correlación del 98 % con la variable "PIB", lo cual generaría problemas de multicolinealidad en el modelo. Además, la variable "Gastos por Alumno, nivel secundario (% del PIB per cápita)"se descartó debido a que presentaba un 90 % de datos faltantes, lo cual afectaría la capacidad de estimación. Estos datos auxiliares pueden descargarse del sitio web del Banco Mundial en el siguiente enlace: `https://data.worldbank.org/`.

Es importante mencionar que la información auxiliar no estaba disponible para todos los dominios, por lo tanto, esta investigación contempló únicamente 77 de los 79 países y economías participantes en el estudio de simulación. Se excluyeron los datos correspondientes a China, Moscow y Tatarstan debido a la falta de disponibilidad de información auxiliar para estos dominios.

Diseño de muestreo

En el estudio PISA 2018, la población objetivo en cada país y economía participante fueron los estudiantes de 15 años en adelante, que abarcaban desde el séptimo grado en adelante, incluyendo aquellos que asistían a jornada completa, parcial, programas de formación profesional u otros programas educativos formales.

El diseño de muestreo utilizado en la mayoría de los países participantes, excepto Rusia, fue un diseño estratificado en dos etapas. En la primera etapa, las unidades primarias de muestreo (UPM) fueron las escuelas que tenían estudiantes de 15 años en adelante en séptimo grado y más. Estas escuelas se seleccionaron mediante un diseño sistemático con probabilidades proporcionales al tamaño (sistemático-πPT), donde el tamaño fue una estimación del número de estudiantes de 15 años elegibles para PISA. Antes de seleccionar las escuelas del marco muestral, se asignaron a grupos mutuamente excluyentes en función de características de la escuela llamadas estratos explícitos. Estos estratos se crearon para mejorar la precisión de los estimadores y se utilizaron variables de estratificación específicas para cada país o

economía, las cuales se pueden encontrar en la tabla 4.1 del informe de diseño de muestreo PISA 2018.

En la segunda etapa, las unidades de muestreo fueron los estudiantes de 15 años en adelante, de séptimo grado en adelante, dentro de las escuelas seleccionadas en la primera etapa. Para cada país participante, se definió un Tamaño Objetivo (TO) que dependía del tipo de evaluación. Para la Evaluación Basada en Computadora (EBC) y la Competencia Mundial (CM), el TO fue de 42 estudiantes, mientras que para la Evaluación Basada en Papel (EBP) o EBC sin CM, el TO fue de 35 estudiantes. A partir de la lista de estudiantes elegibles dentro de las escuelas seleccionadas, si la lista era mayor que el TO, se seleccionaba una muestra de 42 o 35 estudiantes mediante un muestreo aleatorio simple. En caso contrario, se realizaba un censo, es decir, se incluían todos los estudiantes de la lista en la muestra.

n el caso de Rusia, se empleó un diseño de tres etapas. En la primera etapa, se muestrearon áreas geográficas (UPM) mediante un muestreo πPT. En la segunda etapa, se seleccionaron escuelas (USM) dentro de las áreas geográficas mediante un diseño sistemático-πPT. En la tercera etapa, se seleccionaron estudiantes (UTM) dentro de las escuelas mediante un muestreo aleatorio simple. Para obtener más detalles sobre el diseño de muestra PISA 2018, se puede consultar el informe "Diseño de muestra PISA 2018"disponible en el siguiente enlace: `https://www.oecd.org/pisa/data/pisa2018technicalreport/PISA2018%20TecReport-Ch-04-Sample-Design.pdf`.

Los pesos de muestreo (factores de expansión) son publicados por la OCDE en su página web y se utilizan como insumo en las distintas estimaciones a realizar.

Valores plausibles

La metodología utilizada por PISA para imputar los valores plausibles se basa en la técnica propuesta por Rubin (1987) conocida como "imputación múltiple". Esta metodología se aplica a nivel agregado y no a nivel individual, ya que su objetivo es incluir la varianza estimada introducida por los valores plausibles.

La imputación de valores plausibles se utiliza para abordar el problema de que un estudiante solo responda a un subconjunto de los ítems de la prueba, y se necesita conocer cómo se comportaría el estudiante en el conjunto completo de ítems de la prueba Martínez Arias et al. (2006). Para predecir este comportamiento, se utilizan las respuestas a los ítems contestados y otras variables de "condicionamiento"que se obtienen de los cuestionarios de contexto. A partir de esta información, se genera una distribución a posteriori de los valores de cada estudiante, asumiendo generalmente una distribución normal. A partir de esta distribución, se generan aleatoriamente diez valores plausibles para cada estudiante, que son muestras de

su propia distribución.

El objetivo de generar múltiples valores plausibles es poder estimar la varianza derivada de la imputación y evitar sesgos que se producirían si se estimara la habilidad solo con el conjunto de ítems respondidos por el estudiante.

Para obtener más detalles sobre la metodología utilizada por PISA para imputar valores plausibles, se puede consultar el documento "La metodología de los estudios PISA"Martínez Arias et al. (2006) disponible en el siguiente enlace: `https://redined.educacion.gob.es/xmlui/bitstream/handle/11162/68713/00820073007016.pdf?sequence=1`.

Es importante tener en cuenta que debido a la privacidad de los datos publicados por PISA, no es posible realizar el ejercicio de estimación de los valores plausibles. Por lo tanto, se utilizarán los valores plausibles publicados por la OCDE en su página web como insumo para el análisis. Estos valores se pueden descargar de forma gratuita en el enlace mencionado: `https://www.oecd.org/pisa/data/pisa2018technicalreport/PISA2018%20TecReport-Ch-04-Sample-Design.pdf`.

Planteamiento del problema

El enfoque utilizado en esta investigación se divide en dos alternativas de estimación propuestas por SAE. La primera alternativa busca mejorar las estimaciones de los estimadores directos, sintéticos o compuestos utilizados en el estudio de PISA 2018 utilizando información auxiliar relevante. El segundo enfoque consiste en realizar estimaciones en dominios no observados, es decir, estimar los resultados de los países que no participaron en las pruebas PISA 2018, pero para los cuales se dispone de información auxiliar.

En la comparación del primer enfoque, se utiliza el indicador de precisión promedio ($\overline{EER}$) definido en la Sección 5.1 del estudio. Este indicador permite evaluar y comparar la calidad de las estimaciones obtenidas mediante el mejoramiento de los estimadores directos, sintéticos o compuestos.

Para el segundo enfoque, se lleva a cabo un proceso de prueba en el que se estiman algunos países (retirados aleatoriamente). Luego, se evalúan estas estimaciones mediante el cálculo del sesgo relativo, tomando como valor verdadero los resultados publicados por la OCDE. Posteriormente, se realizan estimaciones de países que no participaron en las pruebas PISA 2018, pero para los cuales se cuenta con información auxiliar.

En el contexto de PISA, se utiliza un modelo de dos parámetros logísticos (2PL) con valores plausibles para estimar la habilidad de los estudiantes. El modelo 2PL es un caso particular del modelo de tres parámetros logísticos (3PL) cuando $C_j = 0$ (pseudo azar), este modelo

tiene la siguiente forma:

$$P_i(\varepsilon_{ij} = 1|\theta_i, a_j, b_j) = \frac{e^{1,7a_j(\theta_i - b_j)}}{1 + e^{1,7a_j(\theta_i - b_j)}} \tag{5-1}$$

Donde a_j es la potencia discriminatoria del ítem y b_j es la dificultad del ítem. ε_{ij} es una variable dicotómica que toma valor de 1 cuando el individuo i responde correctamente el ítem j y 0 si responde de forma incorrecta con $i = 1, 2, \ldots, n_d$, n_d es el número de estudiantes seleccionados en el país.

Prueba de lectura

Se realizaron ajustes de modelos utilizando las variables auxiliares presentadas en la sección 5.2. Se ajustaron 14 modelos para los 33 países que tenían información disponible en las 13 variables propuestas. Además, se añadieron 8 modelos adicionales utilizando las variables que tenían un porcentaje de completitud igual o superior al 85 % Para así llegar a un total de 71 países con la información completa.

A continuación, se presenta la tabla del modelo final, la cual nos muestra la información auxiliar significativa, junto con los coeficientes correspondientes, utilizando un nivel de significancia del 10 %:

Variables	$\hat{\beta}$	p-valor
Intercepción	-0,0878	0,07298
% de población con acceso a la electricidad	13,13	0,00965
Tasa de incidencia de la pobreza, sobre la base de $1.90 por día (2011 PPA) (% de la población)	-8,585	0,0175
Investigadores dedicados a investigacion y desarrollo (por cada millon de personas)	0,0121	$1.07\text{x}10^{-5}$

Tabla 5-4: **Modelo 1**, 33 paises y 13 variables

En la tabla 5-4, al considerar todas las variables auxiliares y ajustar el modelo, se encontró que solo 3 de las 13 variables propuestas fueron significativas para estimar el rendimiento promedio en lectura. Estas variables incluyen el porcentaje de población con acceso a electricidad, que está asociada a la infraestructura; la tasa de incidencia de pobreza, que está relacionada con aspectos sociodemográficos; y el número de investigadores dedicados a la investigación y desarrollo, que está vinculado a la ciencia y tecnología.

En cuanto a las estimaciones directas del puntaje promedio en lectura en las pruebas PISA 2018, se observó una correlación espacial estimada de $\hat{\rho} = 0{,}18$ en los 33 países incluidos en el primer modelo. A continuación, se realizará una prueba de hipótesis de Moran para

determinar si esta correlación es significativa.

H_0 : Correlación espacial igual a cero
H_1 : Correlación espacial diferente de cero

Con un $P - value = 0{,}168$ a un nivel de significancia del 5 %, no existe suficiente evidencia estadística para rechazar la hipótesis nula. Por lo tanto, podemos concluir que la correlación espacial de las estimaciones directas del puntaje promedio de lectura en las pruebas PISA 2018 no es significativa.

Para el modelo 2, solo se tuvo en cuenta las variables que cumplieran la condición de completitud mayor o igual a 85 % de las cuales solo una de las que fueron significativas en el modelo 1 cumplió esta condición, el % de población con acceso a electricidad.

Variables	$\hat{\beta}$	p-valor
Intercepción	-31.18	0.08524
% de población que utilizan al menos servicio básico de agua potable	6.910	$9.36\text{x}10^{-4}$
Servidores de internet seguros (por cada millón de personas)	$6.526\text{x}10^{-4}$	$4.87\text{x}10^{-4}$
% de la población que usan internet	0,9001	0.0412
Densidad de población (personas por kilómetro)	$3.899\text{x}10^{-34}$	0.01709

Tabla 5-5: **Modelo 2**, 71 paises y variables con completitud $\geq 85\,\%$

En la tabla 5-5, se observa que al ajustar el modelo 2, la variable % de población con acceso a electricidad que resultó significativa en el modelo 1, no se encontró significativa en el modelo 2. En cambio, las variables % de población que utilizan al menos servicio básico de agua potable, Servidores de internet seguros (por cada millón de personas) y % de la población que usan internet, Densidad de población (personas por kilómetro) resultaron significativas en el modelo 2. Estas variables están relacionadas con aspectos de infraestructura y desarrollo.

En el modelo 2, que considera un total de 71 países, se obtiene una correlación espacial media de $\hat{\rho} = 0{,}399$ en la estimación directa del puntaje medio de lectura en las pruebas PISA 2018. A continuación, se llevará a cabo la prueba de hipótesis de Moran para determinar si esta autocorrelación espacial es estadísticamente significativa.

H_0 : Correlación espacial igual a cero
H_1 : Correlación espacial diferente de cero

Con un $P-value = 0{,}00022$ a un nivel de significancia del 5 %, se rechaza la hipótesis nula. Por lo cual concluimos que la correlación espacial de las estimaciones directas de la prueba de lecturas en los 71 países contemplados en el modelo 2 es significativa.

Resultados

Dado que los 33 países con los que se ajustó el modelo 1 están contenidos en el modelo 2, se reportaran las estimaciones obtenidas con el modelo 2 que presento mejor ajuste y mayor nivel de correlación espacial.

En la tabla 5-6, e presentan las estimaciones del rendimiento medio por país obtenidas a partir de la prueba de lectura, junto con los coeficientes de variación (Cve) publicados por PISA 2018. También se incluyen las estimaciones mediante la metodología propuesta por Tellez (2020) y las estimaciones con el estimador propuesto en esta investigación, junto con sus correspondientes $EER_d = \frac{\sqrt{\hat{MSE}(\hat{\gamma}_d^{Ep})}}{\hat{\gamma_d}^{Ep}} \times 100$. Además, se proporciona el nombre del estimador con el menor EER_d y la medida $Dif_{rel} = \frac{\hat{\sigma}_d^2 - \hat{MSE}(\hat{\gamma}_d^{Ep})}{\hat{\sigma}_d^2} \times 100$ del estimador con el menor EER_d. En esta medida, $\hat{\sigma}_d^2$ representa la varianza estimada publicada por la OCDE, y $\hat{MSE}(\hat{\gamma}_d^{Ep})$ es la varianza estimada mediante la metodología propuesta por Téllez (2020) y la metodología propuesta en esta investigación. Esta medida permite comparar si hubo una reducción en la variabilidad de las estimaciones por país. El estimador con menor EER_d se identifica en la columna **Mejor**.

Países	$\hat{\gamma}_d$	Cve(100 %)	$\hat{\gamma}_d^{EBLUP}$	EER_d^{EBLUP}	$\hat{\gamma}_d^{SEBLUP}$	EER_d^{SEBLUP}	Mejor	Dif_real
Albania	405	0.4741	405	0.4734	405	0.4733	SEBLUP	0.349
Alemania	498	0.6084	498	0.6063	498	0.6056	SEBLUP	1.0257
Arabia Saudita	399	0.7419	399	0.7384	399	0.7374	SEBLUP	1.0521
Australia	503	0.3241	503	0.3237	503	0.3237	SEBLUP	0.2488
Austria	484	0.5579	484	0.5563	484	0.5559	SEBLUP	0.7204
Bakú (Azerbaiyán)	389	0.6452	389	0.6434	389	0.6431	SEBLUP	0.5766
Bélgica	493	0.4706	493	0.4697	493	0.4696	SEBLUP	0.515
Bielorrusia	474	0.5148	474	0.5138	474	0.5134	SEBLUP	0.5802
Bosnia y Herzegovina	403	0.727	403	0.7244	403	0.7238	SEBLUP	0.8254
Brazil	413	0.5109	413	0.5099	413	0.5095	SEBLUP	0.4864
Brunei Darussalam	408	0.2206	408	0.2205	408	0.2205	SEBLUP	0.0742
Bulgaria	420	0.931	420	0.9246	421	0.9232	SEBLUP	1.3973
Canada	520	0.3462	520	0.3458	520	0.3457	SEBLUP	0.3018
Chile	452	0.5841	452	0.5824	452	0.5821	SEBLUP	0.614
Colombia	412	0.7888	412	0.7854	412	0.7847	SEBLUP	1.0387
Costa rica	426	0.8028	426	0.7988	426	0.7984	SEBLUP	1.0582
Dinamarca	501	0.3593	501	0.3588	501	0.3588	SEBLUP	0.238
Emiratos árabes unidos	432	0.5324	432	0.5311	432	0.5311	SEBLUP	0.4887
España	477	0.3312	477	0.3309	477	0.3308	SEBLUP	0.2621
Estados Unidos	505	0.7069	505	0.7036	505	0.7022	SEBLUP	1.3316
Estonia	523	0.3518	523	0.3514	523	0.3513	SEBLUP	0.3156
Federación Rusa	478	0.6444	478	0.6424	478	0.6409	SEBLUP	1.2131
Filipinas	340	0.9676	340	0.963	340	0.9623	SEBLUP	0.8673
Finlandia	520	0.4442	520	0.4434	520	0.4432	SEBLUP	0.5291
Francia	493	0.4706	493	0.4697	493	0.4694	SEBLUP	0.5565

Georgia	380	0.5684	380	0.5672	380	0.5668	SEBLUP	0.4414
Grecia	457	0.7921	457	0.7883	457	0.7874	SEBLUP	1.2558
Hong Kong	524	0.521	524	0.5197	524	0.5196	SEBLUP	0.6336
Hungría	476	0.4727	476	0.4718	476	0.4715	SEBLUP	0.4989
Indonesia	371	0.69	371	0.6887	371	0.6887	SEBLUP	0.4877
Irlanda	518	0.4324	518	0.4317	518	0.4315	SEBLUP	0.4682
Islandia	474	0.3671	474	0.3666	474	0.3665	SEBLUP	0.2654
Israel	470	0.7809	470	0.7768	469	0.7765	SEBLUP	1.3648
Italia	476	0.5126	476	0.5115	476	0.5111	SEBLUP	0.641
Japon	504	0.5298	504	0.5285	504	0.5284	SEBLUP	0.6554
Jordán	419	0.7017	419	0.6992	419	0.6988	SEBLUP	0.8065
Katar	407	0.1892	407	0.1891	407	0.1891	SEBLUP	0.0545
Kazajistán	387	0.3773	387	0.3769	387	0.3768	SEBLUP	0.199
Korea	514	0.572	514	0.5705	514	0.57	SEBLUP	0.7774
Letonia	479	0.3382	479	0.3379	479	0.3378	SEBLUP	0.263
Líbano	353	1.2238	354	1.2132	354	1.2109	SEBLUP	1.4802
Lituania	476	0.3193	476	0.3191	476	0.319	SEBLUP	0.2263
Luxemburgo	470	0.2404	470	0.2403	470	0.2403	SEBLUP	0.1207
Macao	525	0.2343	525	0.2343	525	0.2342	SEBLUP	0.0216
Macedonia del norte	393	0.2799	393	0.2797	393	0.2797	SEBLUP	0.1179
Malasia	415	0.6916	415	0.689	415	0.6884	SEBLUP	0.9577
Malta	448	0.3862	448	0.3857	448	0.3855	SEBLUP	0.2826
Marruecos	359	0.8719	359	0.8695	359	0.869	SEBLUP	0.6741
México	420	0.6548	420	0.6527	420	0.6524	SEBLUP	0.6486
Montenegro	421	0.2494	421	0.2493	421	0.2493	SEBLUP	0.1078
Noruega	499	0.4349	499	0.4341	499	0.4338	SEBLUP	0.4625
Nueva Zelanda	506	0.4032	506	0.4026	506	0.4025	SEBLUP	0.3876
Paises Bajos	485	0.5464	485	0.5448	485	0.5445	SEBLUP	0.563
Panamá	377	0.7825	377	0.7796	377	0.7789	SEBLUP	0.8685
Perú	401	0.7382	401	0.7362	401	0.7358	SEBLUP	0.8189
Polonia	512	0.5273	512	0.5262	512	0.5257	SEBLUP	0.7148
Portugal	492	0.4939	492	0.4929	492	0.4927	SEBLUP	0.5554
Reino Unido	504	0.5119	504	0.5107	504	0.5102	SEBLUP	0.6512
República Checa	490	0.5204	490	0.5191	490	0.5188	SEBLUP	0.6215
República Dominicana	342	0.8363	343	0.8321	343	0.8314	SEBLUP	0.7604
República Eslovaca	458	0.4869	458	0.4859	458	0.4856	SEBLUP	0.4855
Rumania	428	1.2009	429	1.1871	429	1.1835	SEBLUP	2.4506
Serbia	439	0.7449	439	0.7422	438	0.7419	SEBLUP	1.0378
Singapur	549	0.2896	549	0.2894	549	0.2894	SEBLUP	0.1708
Slovenia	495	0.2485	495	0.2484	495	0.2483	SEBLUP	0.146
Suecia	506	0.5968	506	0.595	506	0.5944	SEBLUP	0.8859
Suiza	484	0.6446	484	0.642	484	0.6414	SEBLUP	0.8864
Tailandia	393	0.8219	393	0.8181	393	0.8179	SEBLUP	0.8149
Turquía	466	0.4657	466	0.465	466	0.4649	SEBLUP	0.4554
Ucrania	466	0.7511	465	0.7487	465	0.7479	SEBLUP	1.1381
Uruguay	427	0.6464	427	0.6442	427	0.6439	SEBLUP	0.7145

Tabla 5-6: Resultados de las estimaciones del rendimiento medio de la prueba de lectura PISA 2018 con $\hat{\gamma}_d$, $\hat{\gamma}_d^{EBLUP}$ y $\hat{\gamma}_d^{SEBLUP}$ con sus medidas de calidad

En la tabla 5-6, se puede observar que los valores de EER_d para todas las estimaciones son menores que los coeficientes de variación (Cve) publicados por PISA 2018 para la prueba de lectura. Además, en los 71 países estimados, la metodología propuesta en esta investigación muestra un menor EER_d en comparación con la metodología propuesta en Tellez (2020).

El resultado obtenido fue mejor de lo esperado, ya que el estimador propuesto en la simulación demostró un mejor rendimiento en un menor número de dominios cuando la correlación era moderada-media. Se pudo comprobar que el estimador propuesto es superior al estimador Directo, Compuesto y, en algunos casos, incluso supera al EBLUP. Es importante destacar el buen desempeño del estimador SEBLUP en una gran cantidad de países estimados y cómo se logró reducir el error de estimación (Dif_{rel}) incluso después de aplicar la metodología propuesta en Tellez (2020). Con la metodología propuesta en esta investigación, se logró reducir aún más estos errores en algunos casos.

Habiendo comprobado la efectividad del estimador Fay-Herriot espacial (SEBLUP) en comparación con los demás estimadores, se procederá a abordar el segundo enfoque, que consiste en estudiar el error relativo de las estimaciones para los países que no participaron en PISA 2018. En primer lugar, se eliminarán aleatoriamente el 10 % de los países participantes y se estimarán, con el fin de comparar los resultados con los publicados por PISA 2018. Los países seleccionados para ser eliminados fueron Azerbaiyán, Estonia, Georgia, Panamá, Rumania, España y Uruguay. Después de realizar la estimación utilizando la información auxiliar de estos países, se presentarán los resultados de sus estimaciones junto con su sesgo relativo en la tabla 5-7.

Países	$\hat{\gamma}_d$	$\hat{\gamma}_d^{SEBLUP}$	SB_d
Baku, (Azerbaiyán)	389	416	-6,94
Estonia	523	482	7,84
Georgia	380	417	-9,74
Panamá	377	395	-4,77
Rumania	428	450	-5,14
España	477	462	3,14
Uruguay	427	448	-4,92

Tabla 5-7: Predicción del rendimiento medio en lectura para 7 países eliminados aleatoriamente

Como se puede observar en la tabla 5-7, los sesgos relativos obtenidos para las predicciones de los 7 países eliminados aleatoriamente fueron menores al 10 %, lo que indica que son predicciones acertadas. Por lo tanto, se procederá a realizar predicciones para algunos países que no participaron en PISA 2018, pero para los cuales se cuenta con información auxiliar. Las estimaciones se reportan con el estimador propuesto (SEBLUP) y el modelo 2.

La tabla 5-8 presenta las estimaciones del rendimiento medio en la prueba de lectura PISA 2018 para algunos países no participantes, para los cuales se cuenta con información auxiliar. Estas predicciones fueron realizadas utilizando el segundo enfoque descrito en la formulación del problema. Sin embargo, es importante tener precaución al analizar estas predicciones, ya que “es necesario realizar una discusión detallada sobre la conveniencia de reemplazar

la aplicación de pruebas por la estimación generada mediante modelos estadísticos"(Tellez 2020).

Países	$\hat{\gamma_d}^{SEBLUP}$
Bolivia	367
China	388
Ecuador	395
El Salvador	401
Honduras	376
India	327
Nigeria	238
Pakistán	325
Paraguay	434
Senegal	295

Tabla 5-8: Predicción del rendimiento medio en lectura para países no presentes en PISA 2018

Prueba de matemáticas

Siguiendo la misma metodología utilizada en la prueba de lectura, se ajustaron la misma cantidad de modelos con las mismas variables auxiliares como regresoras, con el objetivo de identificar las variables que contribuyen a explicar o predecir el rendimiento medio en matemáticas, con un nivel de significancia del 10 %. Los resultados obtenidos se muestran a continuación:

Variables	$\hat{\beta}$	p-valor
% de población con acceso a la electricidad	3,0466	$2{,}13\text{x}10^{-7}$
% de la población que usan Internet	1.4969	$8{,}3410^{-3}$
Tasa de incidencia de la pobreza, sobre la base de $1.90 por día (2011 PPA) (% de la población)	-6,9659	0,0340
Población rural (% de la población total)	0.8263	0.01867
Investigadores dedicados a investigacion y desarrollo (por cada millon de personas)	$8.64\text{x}10^{-3}$	$1.08\text{x}10^{-3}$

Tabla 5-9: **Modelo 1**, 33 paises y 13 variables

En la tabla 5-9, al ajustar diferentes modelos teniendo en cuenta todas las variables auxiliares como regresoras del rendimiento medio en la prueba de matemáticas PISA 2018, se encontró que cinco variables resultaron significativas. Dos de ellas están asociadas a infraestructura, como el porcentaje de población con acceso a la electricidad y el porcentaje de la población que utiliza internet. Dos variables están relacionadas con factores socio-demográficos, como la tasa de incidencia de la pobreza basada en $1.90 por día (2011 PPA) y el porcentaje de población rural. Además, una variable está relacionada con ciencia y tecnología, que es el

número de investigadores dedicados a la investigación y desarrollo. En este caso, el intercepto no resultó significativo.

Las estimaciones directas del puntaje medio de matemáticas en las pruebas PISA 2018 presentan una correlación espacial estimada de $\hat{\rho} = 0{,}499$ en los 33 países contemplados en el modelo 1. A continuación, se realizará una prueba de hipótesis de Moran para determinar si esta correlación es significativa.

H_0 : Correlación espacial igual a cero
H_1 : Correlación espacial diferente de cero

Con un $P - value = 0{,}02$ a un nivel de significancia del 5 %, se rechaza la hipótesis nula. Por lo cual concluimos que la correlación espacial de las estimaciones directas de la prueba de matemáticas en los 33 países contemplados en el modelo 1 es significativa.

Variables	$\hat{\beta}$	p-valor
Intercepción	-1,18x10^{3}	0,0134
Desempleo, total (% de participación total en la fuerza laboral- estimación nacional-)	-1,848	0,0809
% de población con acceso a la electricidad	15,67	2,071x10^{-3}
Servidores de internet seguros (por cada millón de personas)	7.354x10^{-4}	1.42x10^{-4}
% de la población que usan internet	0.9092	0.0467
Densidad de población (personas por kilómetro)	5.7x10^{-3}	8.54x10^{-4}

Tabla 5-10: **Modelo 2**, 71 paises y y variables con completitud $\geq 85\,\%$

En la tabla 5-10, se observa que en este modelo el intercepto sí resultó significativo, y se encontraron cinco variables significativas. Dos de ellas están asociadas al aspecto sociodemográfico, como el desempleo total (porcentaje de participación total en la fuerza laboral) y la densidad de población (personas por kilómetro). Tres variables están relacionadas con infraestructura, como el porcentaje de población con acceso a la electricidad, los servidores de internet seguros (por cada millón de personas) y el porcentaje de la población que utiliza internet.

En el modelo 2, que incluye un total de 71 países, se obtuvo una correlación espacial media estimada de $\hat{\rho} = 0{,}467$ en la estimación directa del puntaje medio de matemáticas en las pruebas PISA 2018. A continuación, se plantea una prueba de hipótesis de Moran para determinar si esta autocorrelación espacial es significativa.

H_0 : Autocorrelación espacial igual a cero
H_1 : Autocorrelación espacial diferente de cero

Con un $P - value = 1{,}037e^{-05}$ a un nivel de significancia del 5 %, se rechaza la hipótesis nula. Por lo cual concluimos que la ccorrelación espacial de las estimaciones directas de la prueba de matemáticas en los 71 países contemplados en el modelo 2 es significativa.

Resultados

Dado que los 33 países con los que se ajustó el modelo 1 están contenidos en el modelo 2, se reportaran las estimaciones obtenidas con el modelo 2 que presento mejor ajuste y mayor nivel de correlación espacial.

Países	$\hat{\gamma}_d$	Cve(100 %)	$\hat{\gamma}_d^{EBLUP}$	EER_d^{EBLUP}	$\hat{\gamma}_d^{SEBLUP}$	EER_d^{SEBLUP}	Mejor	Dif_real
Albania	437	0.5538	437	0.5526	437	0.5522	SEBLUP	0.586
Alemania	500	0.53	500	0.5286	500	0.5277	SEBLUP	0.8717
Arabia Saudita	373	0.8016	374	0.7975	374	0.7953	SEBLUP	1.2502
Australia	491	0.3951	491	0.3945	491	0.3944	SEBLUP	0.3691
Austria	499	0.5952	499	0.5934	499	0.5925	SEBLUP	0.9328
Bakú (Azerbaiyán)	420	0.6714	420	0.669	420	0.6683	SEBLUP	0.7944
Bélgica	508	0.4449	508	0.4441	508	0.4439	SEBLUP	0.522
Bielorrusia	472	0.5657	472	0.5642	472	0.5635	SEBLUP	0.7401
Bosnia y Herzegovina	406	0.7537	406	0.751	406	0.75	SEBLUP	0.9103
Brazil	384	0.5286	384	0.5276	384	0.5271	SEBLUP	0.4883
Brunei Darussalam	430	0.2698	430	0.2696	430	0.2696	SEBLUP	0.1266
Bulgaria	436	0.8761	436	0.8708	436	0.8695	SEBLUP	1.3399
Canada	512	0.4609	512	0.4601	512	0.4599	SEBLUP	0.5511
Chile	417	0.5803	417	0.5787	417	0.5784	SEBLUP	0.5131
Colombia	391	0.7647	391	0.7618	391	0.7611	SEBLUP	0.9037
Costa rica	402	0.8184	402	0.8143	402	0.8144	EBLUP	0.9126
Dinamarca	509	0.3418	509	0.3415	509	0.3414	SEBLUP	0.2335
Emiratos árabes unidos	435	0.492	435	0.491	435	0.4909	SEBLUP	0.4453
España	481	0.3035	481	0.3033	481	0.3032	SEBLUP	0.2211
Estados Unidos	478	0.6778	478	0.6748	478	0.673	SEBLUP	1.2557
Estonia	523	0.3327	523	0.3324	523	0.3322	SEBLUP	0.2987
Federación Rusa	488	0.6066	488	0.6048	488	0.6028	SEBLUP	1.323
Filipinas	353	0.983	353	0.9804	353	0.979	SEBLUP	0.7737
Finlandia	507	0.3886	507	0.388	507	0.3878	SEBLUP	0.4133
Francia	495	0.4687	495	0.4678	495	0.4674	SEBLUP	0.6061
Georgia	398	0.6533	398	0.6514	398	0.6509	SEBLUP	0.6775
Grecia	451	0.6851	451	0.6832	451	0.6825	SEBLUP	0.9405
Hong Kong	551	0.5445	551	0.543	551	0.5429	SEBLUP	0.7376
Hungría	481	0.4823	481	0.4814	481	0.481	SEBLUP	0.5708
Indonesia	379	0.8232	379	0.8205	379	0.8202	SEBLUP	0.7404
Irlanda	500	0.44	500	0.4392	500	0.4388	SEBLUP	0.5065
Islandia	495	0.3939	495	0.3933	495	0.3932	SEBLUP	0.3466
Israel	463	0.7559	463	0.7523	463	0.7516	SEBLUP	1.3033
Italia	487	0.5708	487	0.5694	487	0.5686	SEBLUP	0.9058
Japon	527	0.4687	527	0.4678	527	0.4676	SEBLUP	0.5766
Jordán	400	0.8275	400	0.824	400	0.8244	EBLUP	0.9078
Katar	414	0.2971	414	0.2969	414	0.2969	SEBLUP	0.1431
Kazajistán	423	0.4515	423	0.4508	423	0.4505	SEBLUP	0.362
Korea	526	0.5932	526	0.5915	526	0.5908	SEBLUP	0.9156
Letonia	496	0.3952	496	0.3947	496	0.3944	SEBLUP	0.4134
Líbano	393	1.0305	394	1.0223	394	1.0202	SEBLUP	1.5835

Lituania	481	0.4054	481	0.4048	481	0.4046	SEBLUP	0.395
Luxemburgo	483	0.2277	483	0.2276	483	0.2276	SEBLUP	0.1213
Macao	558	0.2742	558	0.2741	558	0.2741	SEBLUP	0.0361
Macedonia del norte	394	0.3959	394	0.3956	394	0.3954	SEBLUP	0.2169
Malasia	440	0.6545	440	0.6522	440	0.6508	SEBLUP	1.1532
Malta	472	0.4025	472	0.402	472	0.4018	SEBLUP	0.3508
Marruecos	368	0.9049	368	0.9003	368	0.8991	SEBLUP	1.0382
México	409	0.6088	409	0.6072	409	0.607	SEBLUP	0.5595
Montenegro	430	0.2884	430	0.2882	430	0.2881	SEBLUP	0.1659
Noruega	501	0.4431	501	0.4423	501	0.4419	SEBLUP	0.5184
Nueva Zelanda	494	0.3462	494	0.3458	494	0.3457	SEBLUP	0.2844
Paises Bajos	519	0.5067	519	0.5054	519	0.5051	SEBLUP	0.5873
Panamá	353	0.7705	353	0.7689	353	0.7683	SEBLUP	0.5388
Perú	400	0.6525	400	0.6516	399	0.6515	SEBLUP	0.5494
Polonia	516	0.5039	516	0.5028	516	0.5024	SEBLUP	0.7174
Portugal	492	0.5447	492	0.5434	492	0.543	SEBLUP	0.6874
Reino Unido	502	0.51	502	0.5088	502	0.5083	SEBLUP	0.7078
República Checa	499	0.493	499	0.4919	499	0.4915	SEBLUP	0.6078
República Dominicana	325	0.8062	326	0.8025	326	0.8015	SEBLUP	0.6667
República Eslovaca	486	0.5267	486	0.5255	486	0.525	SEBLUP	0.6811
Rumania	430	1.1395	431	1.1278	431	1.1238	SEBLUP	2.4165
Serbia	448	0.7054	448	0.7028	448	0.7019	SEBLUP	1.1139
Singapur	569	0.2812	569	0.281	569	0.281	SEBLUP	0.1785
Slovenia	509	0.2672	509	0.267	509	0.267	SEBLUP	0.1888
Suecia	502	0.5279	502	0.5266	502	0.5261	SEBLUP	0.7313
Suiza	515	0.565	515	0.5633	515	0.5627	SEBLUP	0.8187
Tailandia	419	0.8234	419	0.8199	419	0.8192	SEBLUP	1.0018
Turquía	454	0.4978	454	0.4969	454	0.4966	SEBLUP	0.5331
Ucrania	453	0.8057	453	0.8021	453	0.8007	SEBLUP	1.3267
Uruguay	418	0.6292	418	0.6272	418	0.627	SEBLUP	0.6862

Tabla 5-11: Resultados de las estimaciones del rendimiento medio de la prueba de matématicas PISA 2018 con $\hat{\gamma}_d$, $\hat{\gamma}_d{}^{EBLUP}$ y $\hat{\gamma}_d{}^{SEBLUP}$ con sus medidas de calidad

En la tabla 5-11, se puede observar que en 69 de los 71 países estimados, la metodología propuesta proporciona un menor Error Estándar Relativo (EER_d) en comparación con la metodología propuesta en Tellez (2020) y también en comparación con los Coeficientes de Variación (Cve) publicados por PISA 2018 para la prueba de matemáticas. Además, se muestra la disminución porcentual del error de estimación (Dif_{rel}), que es menor con la metodología propuesta en comparación con la metodología oficial de la OCDE y en algunos casos menor a la propuesta por Tellez (2020).

Es destacable que, al igual que en la prueba de lectura, los resultados superan las expectativas, ya que con una correlación media estimada de $\hat{\rho} = 0{,}467$, se obtienen mejores resultados incluso en comparación con simulaciones realizadas con una correlación alta ($\rho = 0{,}75$) y un 30 % de datos faltantes.
A continuación, se aplicó el segundo enfoque de la metodología propuesta para evaluar la capacidad de predicción del rendimiento medio en matemáticas de los modelos ajustados. Este enfoque se evaluó utilizando el sesgo relativo tomando los resultados publicados por

la OCDE como los valores reales. Los países eliminados en este análisis fueron Azerbaiyán, Estonia, Georgia, Panamá, Rumania, España y Uruguay. Se realizó la estimación utilizando la información auxiliar de estos países y los resultados correspondientes se presentan junto con su sesgo relativo en la tabla 5-12.

Países	$\hat{\gamma}_d$	$\hat{\gamma}_d^{SEBLUP}$	SB_d
Baku (Azerbaiyán)	420	451	-7,38
Estonia	523	486	7,07
Georgia	398	412	-3,52
Panamá	353	381	-7,93
Rumania	430	449	-4,42
España	481	440	8,52
Uruguay	418	442	-5,74

Tabla 5-12: Predicción del rendimiento medio en lectura para 7 países eliminados aleatoriamente

La tabla 5-12 presenta las predicciones del rendimiento medio de matemáticas para los 7 países eliminados aleatoriamente, donde se puede identificar que los sesgos relativos son menores al 8 % lo cual indica que son buenas predicciones. Se continua prediciendo el rendimiento medio de matemáticas de algunos países que no presentaron las pruebas PISA 2018 de los cuales se cuenta con información auxiliar. Se reportan los resultados de la misma manera explicada para lo correspondiente la prueba de lectura.

Países	$\hat{\gamma}_d^{SEBLUP}$
Afganistan	327
Bolivia	339
China	428
Cabo Verde	313
Cuba	433
Argelia	398
Ecuador	414
Gabón	287
India	337
Nicaragua	270
Puerto Rico	431
Paraguay	426
El Salvador	380
Sudáfrica	230

Tabla 5-13: Predicción del rendimiento medio de matématicas para países no presentes en PISA 2018

En la tabla 5-13 se muestran las estimaciones del rendimiento medio en la prueba de matemáticas de PISA 2018 utilizando el segundo enfoque de la metodología propuesta. Es importante resaltar que se debe llevar a cabo una discusión detallada sobre la conveniencia de reemplazar la aplicación de pruebas por la estimación generada por modelos estadísticos, como se menciona en Tellez (2020).

6 Conclusiones y recomendaciones

Este estudio reveló mediante simulaciones que el estimador propuesto (SEBLUP) presenta sesgos mínimos, inferiores al 1 %, lo cual sugiere un posible insesgamiento en las estimaciones. Además, se observó que en la mayoría de casos el estimador SEBLUP supera al estimador EBLUP en términos de EERP (error estándar relativo promedio). A medida que la correlación espacial aumenta, se incrementa la proporción de casos en los que el estimador SEBLUP es más eficiente que el estimador EBLUP en términos de EERP. Asimismo, a medida que la correlación lineal disminuye, aumenta la diferencia de EERP entre el estimador SEBLUP y el EBLUP. En todos los casos, el estimador propuesto exhibe un EERP inferior al 2 %, lo cual, según Gutiérrez (2016), se considera bajo.

En la aplicación práctica de las pruebas de lectura y matemáticas PISA 2018, se lograron mejoras en las estimaciones en términos de precisión en comparación con los resultados publicados por la OCDE en PISA 2018. Además, se realizaron predicciones para algunos países que no participaron en PISA 2018. En la mayoría de los países estimados, el estimador SEBLUP demostró ser el más preciso. Por lo tanto, la metodología propuesta en esta investigación se presenta como una nueva alternativa de estimación en pruebas estandarizadas, al igual que la metodología propuesta por Tellez (2020), ya que ambas muestran un gran potencial predictivo que ayuda a mejorar la precisión de las estimaciones. Estas metodologías también pueden ser aplicadas en otras pruebas estandarizadas, como TIMMS, PRILS e ICFES.

Un aspecto fundamental de este trabajo consistió en identificar los factores que influyen en el rendimiento de los estudiantes. los cuales fueron las variables auxiliares que mostraron una significancia estadística en cada una de las pruebas realizadas, tanto en lectura como en matemáticas. Estos hallazgos resultan relevantes para tenerlos en cuenta al momento de formular políticas públicas relacionadas con la educación, ya que desempeñan un papel crucial en el desempeño de los estudiantes.

Bibliografía

Attorresi, H. F., Lozzia, G. S., Abal, F. J. P., Galibert, M. S. y Aguerri, M. E. (2009), 'Teoría de respuesta al ítem. conceptos básicos y aplicaciones para la medición de constructos psicológicos', *Revista Argentina de Clínica Psicológica* **18**(2), 179–188.

Birnbaum, A. (1958), 'On the estimation of mental ability', *Series Rep* **15**, 7755–7723.

Chandra, H. y Salvati, N. (2018), 'Small area estimation of proportions under a spatial dependent aggregated level random effects model', *Communications in Statistics-Theory and Methods* **47**(5), 1234–1255.

Dempster, A. P., Laird, N. M. y Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the em algorithm', *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1), 1–22.

Deville, J.-C. y Särndal, C.-E. (1992), 'Calibration estimators in survey sampling', *Journal of the American statistical Association* **87**(418), 376–382.

Drew, D., Singh, M. y Choudhry, G. (1982), 'Evaluation of small area estimation techniques for the canadian labour force survey', *Survey Methodology* **8**(1), 17–47.

Embretson, S. E. y Reise, S. P. (2013), *Item response theory*, Psychology Press.

Fay, R. y Herriot, R. (1979), 'Estimates of income for small places: an application of james-stein procedures to census data', *Journal of the American Statistical Association* **74**(366a), 269–277.

Gutiérrez, H. A. (2009), *Estrategias de muestreo diseño de encuestas y estimación de parámetros.*, Universidad Santo Tomas, Bogota (Colombia).

Gutiérrez, H. A. (2016), *Estrategias de muestreo: diseño de encuestas y estimación de parámetros*, Ediciones de la U.

Harville, D. A. y Jeske, D. R. (1992), 'Mean squared error of estimation or prediction under a general linear model', *Journal of the American Statistical Association* **87**(419), 724–731.

Horvitz, D. G. y 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 (2022), 'Informe nacional de resultados saber 3°, 5°, 7° y 9°'.

Longford, N. T. (2005), 'On selection and composition in small area and mapping problems', *Statistical methods in medical research* **14**(1), 3–16.

Martínez Arias, M. R. et al. (2006), 'La metodología de los estudios pisa', *Revista de educación* .

Martínez, C. (2012), *Estadística y muestreo-13ra Edición*, Ecoe ediciones.

Matas Terrón, A. et al. (2010), 'Introducción al análisis de la teoría de respuesta al ítem'.

Mislevy, R. J. (1991), 'Randomization-based inference about latent variables from complex samples', *Psychometrika* **56**(2), 177–196.

Mislevy, R. J., Beaton, A. E., Kaplan, B. y Sheehan, K. M. (1992), 'Estimating population characteristics from sparse matrix samples of item responses', *Journal of Educational Measurement* **29**(2), 133–161.

Molina, I. (2019), *Desagregación de datos en encuestas de hogares: metodologías de estimación en áreas pequeñas*, División de Estadísticas de la Comisión Económica para América Latina y el Caribe (CEPAL).

OECD (2016), 'Pisa 2018. draft analytical frameworks, may 2016'.

Petrucci, A. y Salvati, N. (2006), 'Small area estimation for spatial correlation in watershed erosion assessment', *Journal of agricultural, biological, and environmental statistics* **11**(2), 169.

Pfeffermann, D. (2002), 'Small area estimation-new developments and directions', *International Statistical Review* **70**(1), 125–143.

Prasad, N. N. y Rao, J. N. (1990), 'The estimation of the mean squared error of small-area estimators', *Journal of the American statistical association* **85**(409), 163–171.

Pratesi, M. y Salvati, N. (2005), *Small area estimation: the EBLUP estimator with autoregressive random area effects*, Università di Pisa, Dipartimento di statistica e matematica applicata all ....

Pratesi, M. y Salvati, N. (2008), 'Small area estimation: the eblup estimator based on spatially correlated random area effects', *Statistical methods and applications* **17**(1), 113–141.

Rao, J. (2003), *Small Area Estimation*, Wiley, New York.

Rao, J. N. y Molina, I. (2015), *Small area estimation*, John Wiley & Sons.

Rubin, D. B. (1987), *Multiple imputation for nonresponse in surveys*, John Wiley & Sons.

Salvati, N. (2004*a*), *La correlazione spaziale nella stima per piccole aree: metodi proposti e casi di studio*, Università degli Studi di Firenze.

Salvati, N. (2004*b*), 'Small area estimation by spatial models: the spatial empirical best linear unbiased prediction (spatial eblup)', *Working Paper 2004/03. Firenze: Universitá degli Studi di Firenze* pp. 59–50134.

Särndal, C.-E., Swensson, B. y Wretman, J. (2003), *Model assisted survey sampling*, Springer Science & Business Media.

Tellez, C. F. (2020), Estimación de áreas pequeñas utilizando imputación múltiple en modelos logísticos de tres parámetros, PhD thesis.

Tellez, C. F., Arias, I. R., Gómez, S. G. y Oyola, L. T. (2021), 'Estimation of educational establishments performance in saber 5◦ tests in colombia. an approach from small area estimation', *Boletın de Estadıstica e Investigación Operativa BEIO* p. 169.

Triviño, A. F. P., Piñerez, C. F. T. y Oyola, L. T. (2020), 'Estimación de los resultados en matemáticas y ciencias de las pruebas timss 2015: un nuevo enfoque desde la metodología de áreas pequeñas', *Comunicaciones en Estadística* **13**(2), 62–77.

Von Davier, M., Gonzalez, E. y Mislevy, R. (2009), 'What are plausible values and why are they useful', *IERI monograph series* **2**(1), 9–36.

Wackerly, D., Mendenhall, W. y Scheaffer, R. L. (2014), *Mathematical statistics with applications*, Cengage Learning.

Wawrowski, Ł. (2016), 'The spatial fay-herriot model in poverty estimation', *Folia Oeconomica Stetinensia* **16**(2), 191–202.

Zimmerman, D. L. y Cressie, N. (1992), 'Mean squared prediction error in the spatial linear model with estimated covariance parameters', *Annals of the institute of statistical mathematics* **44**(1), 27–43.