

Aplicación del Algoritmo de Expectación-Maximización con Bootstrap para la Imputación de Indicadores en Modelos de Regresión Beta Multinivel Bayesianos con enlace Logit: un análisis de panel para explicar la relación del Gasto público en educación, el Factor productivo nacional y el Factor demográfico sobre la Pobreza en educación del 2015 al 2023.

Michael Stibenson Montoya Casas

Universidad Santo Tomás

Facultad de Estadística

Bogotá D.C.

2024

**Aplicación del Algoritmo de Expectación-Maximización con Bootstrap para la
Imputación de Indicadores en Modelos de Regresión Beta Multinivel
Bayesianos con enlace Logit: un análisis de panel para explicar la relación del
Gasto público en educación, el Factor productivo nacional y el Factor
demográfico sobre la Pobreza en educación del 2015 al 2023.**

Michael Stibenson Montoya Casas

Proyecto de grado para optar por el título Estadístico.

**Director de Proyecto de Grado
Docente. Mario José Pacheco López
Magister en estadística**

**Universidad Santo Tomás
División de Ciencias Económicas y Administrativas
Facultad de Estadística
Bogotá D.C.
2024**

Contenido

INTRODUCCIÓN	5
OBJETIVOS	6
OBJETIVO GENERAL:	6
OBJETIVOS ESPECÍFICOS:	6
JUSTIFICACIÓN	7
MARCO REFERENCIAL	8
ANTECEDENTES Y ESTUDIOS	8
MARCO TEÓRICO	10
CONTEXTO DE LA EDUCACIÓN EN COLOMBIA	10
IDENTIFICACIÓN DE INDICADORES DE POBREZA EDUCATIVA.	12
<i>Analfabetismo</i>	12
<i>Inasistencia Escolar</i>	14
<i>Bajo Logro Educativo</i>	16
<i>Rezago Escolar</i>	18
METODOLOGÍA	19
TIPO DE INVESTIGACIÓN	19
IMPUTACIÓN DEL PANEL DE DATOS	21
<i>Algoritmo de imputación</i>	21
<i>Medidas de severidad de Rubín</i>	23
APLICACIÓN DEL MODELO	25
RESULTADOS	28
ANALFABETISMO	28
INASISTENCIA ESCOLAR	29
BAJO LOGRO EDUCATIVO	29
REZAGO ESCOLAR	30
ANÁLISIS INTEGRAL DE LOS RESULTADOS.	31
CONCLUSIONES	32
REFERENCIAS	33
ANEXOS	37

Tablas

<i>Tabla 1 Variables y covariables utilizadas para el estudio</i>	20
<i>Tabla 2 Medidas de severidad Rubín para las variables de pobreza en educación imputadas.</i>	25
<i>Tabla 3 Covariables significativas al 95% y 90% de confianza del modelo combinado de Analfabetismo</i>	28
<i>Tabla 4 Departamentos significativos al 95% y 90% de confianza del modelo combinado de Analfabetismo</i>	29
<i>Tabla 5 Covariables significativas al 95% y 90% de confianza del modelo combinado de Inasistencia Escolar.</i>	29
<i>Tabla 6 Covariables significativas al 95% y 90% de confianza del modelo combinado de Bajo Logro Educativo</i>	29
<i>Tabla 7 Departamentos significativos al 95% y 90% de confianza del modelo combinado de Bajo Logro Educativo</i>	30
<i>Tabla 8 Covariables significativas al 95% y 90% de confianza del modelo combinado de Rezago Escolar</i>	30
<i>Tabla 9 Departamentos significativos al 95% y 90% de confianza del modelo combinado de Rezago escolar</i>	31

Tabla de ilustraciones

<i>Ilustración 1 Comportamiento del gasto público en educación de Colombia 2015 – 2022, precios constantes con base en 2015, cifras en billones de pesos COP, total nacional.</i>	11
<i>Ilustración 2 Indicador de Analfabetismo en Colombia por Departamento años 2018 – 2020 – 2022.</i>	13
<i>Ilustración 3 Indicador de Inasistencia Escolar en Colombia por Departamento años 2018 – 2020 – 2022.</i>	15
<i>Ilustración 4 Indicador de Bajo Logro Educativo en Colombia por Departamento años 2018 – 2020 – 2022.</i>	17
<i>Ilustración 5 Indicador de Rezago Escolar en Colombia por Departamento años 2018 – 2020 – 2022.</i>	19
<i>Ilustración 6 Comparativo de la densidad que tienen las variables imputadas contra los casos completos.</i>	23

Tabla de anexos

Anexo 1	37
Anexo 2	40
Anexo 3	43
Anexo 4	46

Introducción

La búsqueda de conocimiento se puede considerar como un factor intrínseco de la raza humana, diversos factores (i) sociales, (ii) económicos, (iii) políticos e incluso (iv) ambientales de las sociedades se ven íntimamente relacionados con el nivel académico que poseen, Amartya Sen (1998) expuso que la integración de la educación y la productividad económica generó cambios en los cuatro factores mencionados anteriormente, dando como resultado transformaciones significativas en la estructura social y condiciones de vida de todo el mundo.

Para conocer el impacto y la relación que tienen los factores mencionados en la calidad de vida en los diversos países del mundo, la Organización de las Naciones Unidas (2015) planteó 17 objetivos de desarrollo sostenible, siendo su principal objetivo acabar con la pobreza, asegurar la prosperidad de todas las personas y proteger el medio ambiente, los objetivos tienen una agenda de seguimiento y evaluación hasta el 2030.

Por otra parte, “la inversión pública desempeña un rol clave en el crecimiento económico y en el logro de los Objetivos de la Agenda 2030, resulta importante contar con información acerca de en qué funciones y sectores se ha priorizado la inversión” (Podestá, 2020, pág. 48). La educación como punto clave del desarrollo es uno de los principales rubros de inversión pública, de este modo, evaluar la relación e impacto que efectivamente tiene la política pública sobre el nivel educativo sirve como determinante para la distribución de recursos nacionales en futuros planes de gobierno.

El Departamento Administrativo Nacional de Estadística (DANE) es la entidad a cargo de realizar la medición del nivel de pobreza y avance educativo en Colombia, el Índice de Pobreza Multidimensional (IPM) contiene 15 indicadores, de los cuales 4 miden el nivel educativo de la población (Analfabetismo, Inasistencia escolar, Rezago escolar y Bajo logro educativo). Esta medición de pobreza y educación junto con los objetivos de desarrollo mundial motivaron las siguientes preguntas de investigación, ¿Cuál es el efecto que la inversión pública junto al factor económico y demográfico en Colombia tuvieron sobre la pobreza educativa del 2015 – 2023? y ¿cuál es el panorama frente a la agenda de desarrollo mundial 2015 - 2030?

Objetivos

Objetivo General:

Modelar los indicadores de pobreza educativa en Colombia, frente a la agenda de desarrollo mundial a nivel departamental, teniendo en cuenta el gasto público junto a los factores económicos y demográficos del país del 2015 al 2023.

Objetivos Específicos:

- Identificar en los repositorios oficiales de la información nacional, las variables que dan a conocer el comportamiento histórico de los indicadores de pobreza en educación, gasto público en educación, factores demográficos (densidad poblacional & migración) y el factor de producción económica (PIB per cápita) a nivel departamental.

- Aplicar algoritmos de datos faltantes para las variables de pobreza en educación desde 2015 a 2018 a nivel departamental, para mejorar la robustez de su modelamiento y capturar de mejor forma su comportamiento histórico utilizando el algoritmo Expectación-Maximización con Bootstrap.

- Estimar modelos de Regresión Beta Multinivel Bayesianos con Enlace Logit, mediante el método de Monte Carlo con Cadenas de Markov, para determinar la relación que tienen los indicadores de pobreza en educación con el gasto público en educación, el factor demográfico y la producción económica a nivel departamental del 2015 a 2023.

- Realizar el análisis y comparativo de los resultados obtenidos de pobreza en educación frente a la agenda de desarrollo mundial 2030.

Justificación

En su informe y comunicado de prensa el Banco Mundial, BM, (2022a) dio a conocer el preocupante aumento de pobreza en aprendizajes, que tuvieron los países de ingreso medio y bajo a partir de la coyuntura mundial ocurrida en el año 2020 por la pandemia, el panorama se detalla a nivel mundial como una catástrofe para el aprendizaje y avance social de la población, el BM estima que para 2022 el 70% de los niños de 10 años no puede comprender una lectura o texto básico y Colombia no es la excepción, en su boletín técnico el Departamento Administrativo Nacional de Estadística, DANE, (2024c) dio a conocer que el porcentaje de personas en condición de pobreza multidimensional fue del 12.1% para 2023, disminuyó 0.8% con respecto al año 2022, pero aún es bastante elevado.

En su informe el Banco de la República de Colombia (2023) dio a conocer la diferencia de acceso a la educación que tiene la población dependiendo su grupo social, esta diferencia se presenta desde la edad colonial y aunque ya se ha opacado en gran medida, aún se presentan diferencias marcadas del nivel educativo dependiendo el estrato o grupo social. El gobierno colombiano en el año 2011 anunció un nuevo Plan Nacional de Desarrollo¹, el principal enfoque de este fue la reducción de la pobreza y la desigualdad, para el cual, el Departamento Nacional de Planeación, DNP, ideó un Índice de pobreza con el método Alkire-Foster² (Angulo, Díaz y pardo, 2013, pág. 36). Por su parte, el DANE lleva a cabo la medición del índice de Pobreza Multidimensional (IPM) basado en la Encuesta Nacional de Calidad de Vida (ECV), este índice se utiliza para evaluar o plantear políticas públicas, para investigaciones de todo tipo referentes a pobreza y para conocimiento del comportamiento social que tiene el fenómeno de la pobreza.

En línea con la agenda de desarrollo mundial, es importante saber si las acciones que está tomando el gobierno en materia de gasto público, además de las condiciones demográficas y económicas a nivel nacional, tienen un real efecto al momento de disminuir la pobreza y mejorar el nivel educativo de la población, como

¹ Plan Nacional de Desarrollo (PND), Juan Manuel Santos 2010 – 2014 con las iniciativas de “instrumento de innovación”, y “locomotoras del crecimiento” (DNP, 2011, págs. 88 - 203).

² Cumple con las propiedades axiomáticas planteadas por Amartya Sen (1976-1979) y es un indicador de brecha, severidad e incidencia (Angulo, Díaz y pardo, 2011, pág. 6-11)

mencionó Amartya Sen (1998) un avance educativo trae consigo una mejora en las condiciones de calidad de vida que posee toda la población.

Marco Referencial

Antecedentes y estudios

A nivel teórico el estudio de la pobreza, la educación y el gasto público tienen un amplio acervo de investigación histórica, en el caso de la teoría clásica del desarrollo y crecimiento económico Mankiw, Romer y Weil (1992) mencionan que la inversión en el capital humano trae consigo un mayor crecimiento económico a largo plazo, Barro (1991), Keller (2006) y Romer (1990) por su lado mencionan que los países con un mayor nivel educativo experimentan mejores tasas de desarrollo a nivel económico y social, a nivel de bienestar social Gary Becker (1964) en su investigación evidenció que la relación existente entre el incremento del ingreso y la mejora del bienestar está relacionada con la acumulación de conocimientos, estudios y habilidades.

A nivel técnico y más específicamente a nivel estadístico se han desarrollado diversas metodologías de medición en este campo, partiendo de; toma de los datos que “abarcan desde información recopilada en encuestas de hogares hasta píxeles captados por imágenes satelitales” (Banco Mundial, 2023) y la clasificación de los no pobres en un indicador de distintas dimensiones Alkire-Foster (2011).

Desde el año 2018 Colombia cuenta con la medición del IPM a nivel departamental, pero en la presente investigación se busca tener un pasado más amplio, Según Box, Jenkins, Ljung, Reinsel (2015) y Hamilton (1994) una serie temporal más extensa permite estimar relaciones y parámetros con mayor precisión, además Hyndman & Athanasopoulos (2021) hacen hincapié en la importancia de una serie temporal más extensa para mejorar el ajuste de su modelamiento y pronóstico.

Para obtener los beneficios de ajuste con una serie de tiempo más extensa Rubin (1987) y Enders (2010) utilizaron covariables para mejorar el ajuste de las imputaciones y los modelos resultantes, por otro lado, Baltagi (2013) dio a conocer una visión detallada del uso de covariables para mejorar el ajuste de modelos estadísticos de datos panel, los puntos positivos que tienen las variables de control incluyen la disminución de sesgo por estimación y control de factores no observados

en el modelo, Wooldridge (2015) también resaltó el uso de covariables en modelos de datos panel para capturar efectos espaciales y temporales, pues mejora la capacidad para explicar el fenómeno de estudio.

Realizando una revisión de corte más empírica Del Callejo, Canal, Vernazza, Urruticoechea & Álvarez (2022) con una base de datos panel que contenía un 24% de datos faltantes generados al azar, utilizaron el software Rstudio (librería Amelia), con la actualización del algoritmo EMB para aplicar un algoritmo Monte Carlo con Cadenas de Markov MCMC y así generar su imputación, el estudio demostró que incluir restricciones con información a priori del comportamiento que deben tener los datos puede mejorar la distribución de la imputación.

Rodriguez, Freire & Pais (2017) aplicaron un modelo de Efectos Fijos en los países pertenecientes a la OCDE, tomaron variables económicas (como importaciones, exportaciones, consumo, PIB per-cápita), variables sociales (como el índice de desarrollo humano) y el gasto público en educación para los años 1980 – 2011, el estudio arrojó una relación positiva entre la inversión pública en educación y el desarrollo y bienestar social.

Para el caso de Colombia, Galvis (2015) utilizó la medición del gasto territorial mediante el sistema general de participaciones (SGP) para medir la eficiencia de la inversión pública en el sistema educativo de Colombia, el análisis envolvente utilizado demostró que las regiones con menos ciudades y más apartadas del centro tienen una relación menos optima entre gasto público y el nivel educativo del país, conclusión también demostrada por Chamorro (2013).

Los estudios realizados en Colombia sobre la relación del gasto público y la pobreza en educación con desagregación a nivel departamental son escasos, al no existir un histórico extenso se limita bastante la aplicación de modelos con series de tiempo. Utilizando algoritmos de imputación para expandir una base de datos panel con covariables relacionadas al tema educativo, económico, demográfico y política fiscal de Colombia, se puede realizar un estudio con evidencia empírica que dé cuenta de la relación que tienen dichas covariables sobre los indicadores de pobreza en educación. Adicionalmente, un modelo bayesiano utilizando el método de Monte Carlo Con Cadenas de Markov (MCMC) para simular muestras de la distribución posterior, es innovador para este tipo de análisis en Colombia, pues es muy útil en casos de

poca disponibilidad histórica en la data, tal como lo resaltaron Gelman, Carlin, Stern, Dunson, Vehtari & Rubin en (2013). “Los métodos MCMC son atractivos porque proporcionan una forma sencilla e intuitiva de simular valores de una distribución desconocida y utilizar esos valores simulados para realizar análisis posteriores. Esto les permite ser aplicables en una amplia variedad de dominios” (Speagle. 2019. Pag.2).

Marco Teórico

Contexto de la educación en Colombia

Según Galvis (2015) y Chamorro (2013) la división geográfica y política de Colombia dificulta tener control real sobre el nivel de avance educativo en los departamentos más alejados del centro, el gasto público no está correctamente enfocado y factores coyunturales como la corrupción toman su parte de los recursos destinados a la educación.

Al contrastar los datos del gasto público en educación (Sistema General de Participaciones³) a precios constantes⁴ contra el plan de gobierno de turno se encontró que; la iniciativa “Colombia, la más educada” correspondiente al Plan Nacional de Desarrollo 2014 – 2018 (DNP. 2015. págs. 75 - 107), en cabeza del presidente Juan Manuel Santos, tuvo un crecimiento constante, pasando de 17.3 a 19.2 billones de pesos COP (Ilustración 1), para el siguiente plan de gobierno la iniciativa “Pacto por la equidad, Educación de calidad para un futuro con oportunidades para todos” correspondiente al Plan Nacional de Desarrollo 2018 – 2022 (DNP. 2019. págs. 257 - 510), en cabeza del presidente Iván Duque Márquez presentó un crecimiento acelerado en los dos primeros años de mandato para después presentar un comportamiento decreciente, pasó de 19.2 a 20.3 billones de pesos COP (Ilustración 1), para el año 2023 el plan de gobierno corresponde a la iniciativa “Educación de calidad para reducir la desigualdad” determinada en el Plan Nacional de Desarrollo 2022 – 2026 (DNP. 2023. págs. 113 - 119), en cabeza del

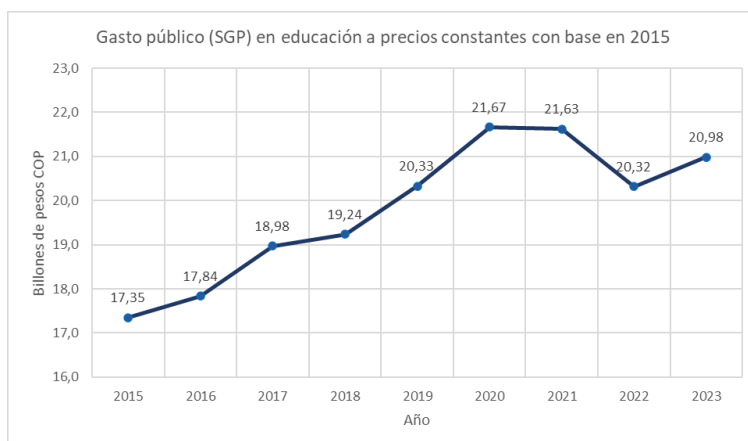
³ El Sistema General de Participaciones:

“Es el conjunto de recursos que la Nación transfiere por mandato de los artículos 356 y 357 de la Constitución Política (modificados por los Actos Legislativos 01 de 2001 y 04 de 2007), a las entidades territoriales (departamentos, distritos y municipios) para la financiación tanto de los servicios a su cargo, como para cumplir con las competencias asignadas por la Ley” (DNP, 2024)

⁴ Los precios constantes eliminan el factor de inflación del valor real con un año base.

actual presidente Gustavo Francisco Petro Urrego presentó una subida de 20.3 a 21 billones de pesos COP (Ilustración 1) del SGP perteneciente al rubro de educación.

Ilustración 1 Comportamiento del gasto público en educación de Colombia 2015 – 2022, precios constantes con base en 2015, cifras en billones de pesos COP, total nacional.



Fuente: Elaboración propia basado en el DNP (2015), DNP (2019), DNP (2023), las series fueron deflactadas con los datos del IPC provenientes del “Anexo del SGP” publicado por el SICODIS⁵ (2024).

Por su parte, Ferrero y Parrado (2019) evidenciaron que históricamente Colombia no ha tenido un gasto público diferencial a nivel educativo si se compara con los otros países de la región o con los otros países pertenecientes a la OCDE, de este modo se evidencia una mayor necesidad de control y enfoque sobre los recursos destinados a la educación, tal como resalta el Banco Mundial (2024) en su análisis del panorama actual de estudio mundial para países en vías de desarrollo “realizar inversiones inteligentes y eficaces en la educación de las personas resulta imprescindible para desarrollar el capital humano con el que se pondrá fin a la pobreza extrema”.

Colombia aún tiene que recorrer un largo camino para mejorar a nivel de cobertura educativa, el DANE en su boletín técnico de IPM para el año 2022 dio a conocer que “en el total nacional, la dimensión que tuvo mayor contribución a la pobreza multidimensional fue educación 35,0%” (2023. Pág. 12), el boletín técnico también resaltó que la dimensión de educación se vio más afectada en centros

⁵ SICODIS: Es el Sistema de información y consulta de distribuciones de recursos territoriales.

poblados y rural disperso⁶, en este punto es importante tener en cuenta el contraste de pobreza en educación con el decreto 1411 del 2022 [Ministerio de Educación] por el cual se establecen la universalidad, la equidad, la complementariedad, la intersectorialidad, la inclusión y diversidad como los principios que orientan el servicio de educación inicial (29 de julio 2022).

La CEPAL (2023) realizó recomendaciones en términos de avance en el nivel educativo de los países de América Latina y El Caribe, se centró en universalizar el acceso a la educación primaria y fortalecer la conclusión de los niveles educativos de bachillerato, cabe destacar que un mejor desarrollo en su nivel de aprendizaje secundario fortalecerá el intento de acceso a niveles de educación superior, frente a esto la CEPAL indica que:

“Una de las medidas clave para seguir avanzando en cobertura y calidad consiste en promover la inversión en la educación desde la primera infancia. Las bases del aprendizaje se sientan en las etapas tempranas de la niñez; es en ese período cuando actúan los principales impulsores de la desigualdad” (CEPAL, 2022, pág. 228).

Identificación de indicadores de pobreza educativa.

El Departamento Nacional de Planeación, DNP, en el año 2017 demostró que un buen nivel de significancia a nivel departamental de la ECV permitiría tomar mejores decisiones de política e inversión pública, de este modo, el DANE en 2018 por primera vez publicó el IPM⁷ y sus 15 indicadores con desagregación y significancia a nivel departamental.

Analfabetismo

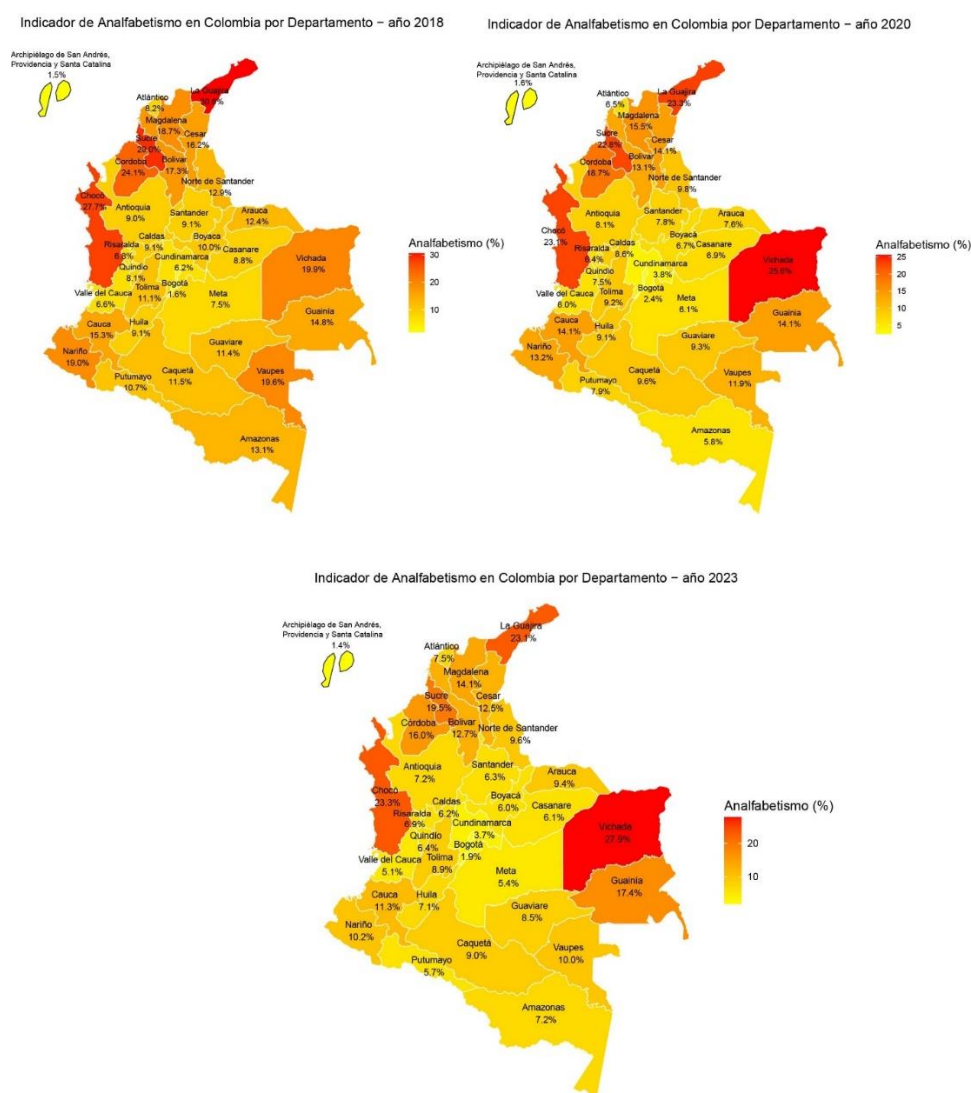
La base de datos panel contiene información del año 2018 al año 2023 a nivel departamental para el indicador de analfabetismo. El indicador de Analfabetismo muestra en porcentaje la proporción de hogares privados, según el proceso de construcción del indicador DNP (2011) “aquellos hogares donde menos del 100% de las personas de 15 años y más saben leer y escribir [...]”. Cuando en el hogar no se

⁶ La mayor contribución al IPM para centros poblados y rural disperso en el año 2022 fue de la dimensión de educación 36,6%, en el área de cabeceras la dimensión de educación fue de 33,5%. (DANE, 2023)

⁷ El IPM contiene 15 indicadores asignados en las siguientes 5 dimensiones: (i) Educación, (ii) Condiciones de la niñez y Juventud, (iii) Trabajo, (iv) Salud, (v) Condiciones de la Vivienda.

reportan personas de 15 o más años, el hogar se considera en privación” (Pág. 15), en este contexto de medición el analfabetismo es abarcado a nivel de hogar⁸, si el individuo o alguna de las personas que viven con él y están asociados a él no sabe leer o escribir teniendo 15 años o más, se considera una persona privada del acceso a la educación, al ser un indicador que abarca la pobreza en varios aspectos también se considera privado de este derecho el hogar que solo contiene personas menores de 15 años.

Ilustración 2 Indicador de Analfabetismo en Colombia por Departamento años 2018 – 2020 – 2022.



Fuente: Elaboración propia basado en el DANE (2024d) “Anexo Departamental de Pobreza Multidimensional 2023”

⁸ Hogar:

“una persona o grupo de personas que ocupan la totalidad o parte de una vivienda y que se han asociado para compartir la dormida y/o la comida. Pueden ser familiares o no entre sí. Los empleados del servicio doméstico y sus familiares forman parte del hogar siempre y cuando duerman en la misma vivienda donde trabajan.” (DANE, 2024)

El indicador de Analfabetismo pertenece a la dimensión de Condiciones Educativas, gráficamente es posible identificar los departamentos con mayor porcentaje de privación en la periferia del país, los departamentos de Vichada, el Chocó y la Guajira no han podido disminuir del 20% el indicador de pobreza, los departamentos de Amazonas, Córdoba, la Guajira, Nariño, Sucre, Putumayo y Vaupés presentaron un salto positivo de al menos 5% del 2018 al 2023, los departamentos de Guainía, Risaralda y Vichada tuvieron un incremento en el nivel de analfabetismo para el periodo de muestra, los departamentos de San Andrés, Providencia & Santa Catalina, Casanare, Valle del Cauca, Risaralda, Quindío, Meta y Cundinamarca son los departamentos con menos tasa de privación del 2018 al 2023, cabe resaltar e incluir a Bogotá en este grupo (Ilustración 2).

Para el año 2018 Colombia presentó un indicador de analfabetismo en 10%, para el año 2020 a nivel nacional se tuvo un indicador de 8,4% y para el año 2023 el indicador de analfabetismo fue de 7,5%, DANE (2024c), este indicador no presentó alteraciones en su comportamiento decreciente por la coyuntura mundial del año 2020, si bien las condiciones de educación fueron distintas por el aislamiento, no se presentó un efecto significativo para ese año y los años siguientes.

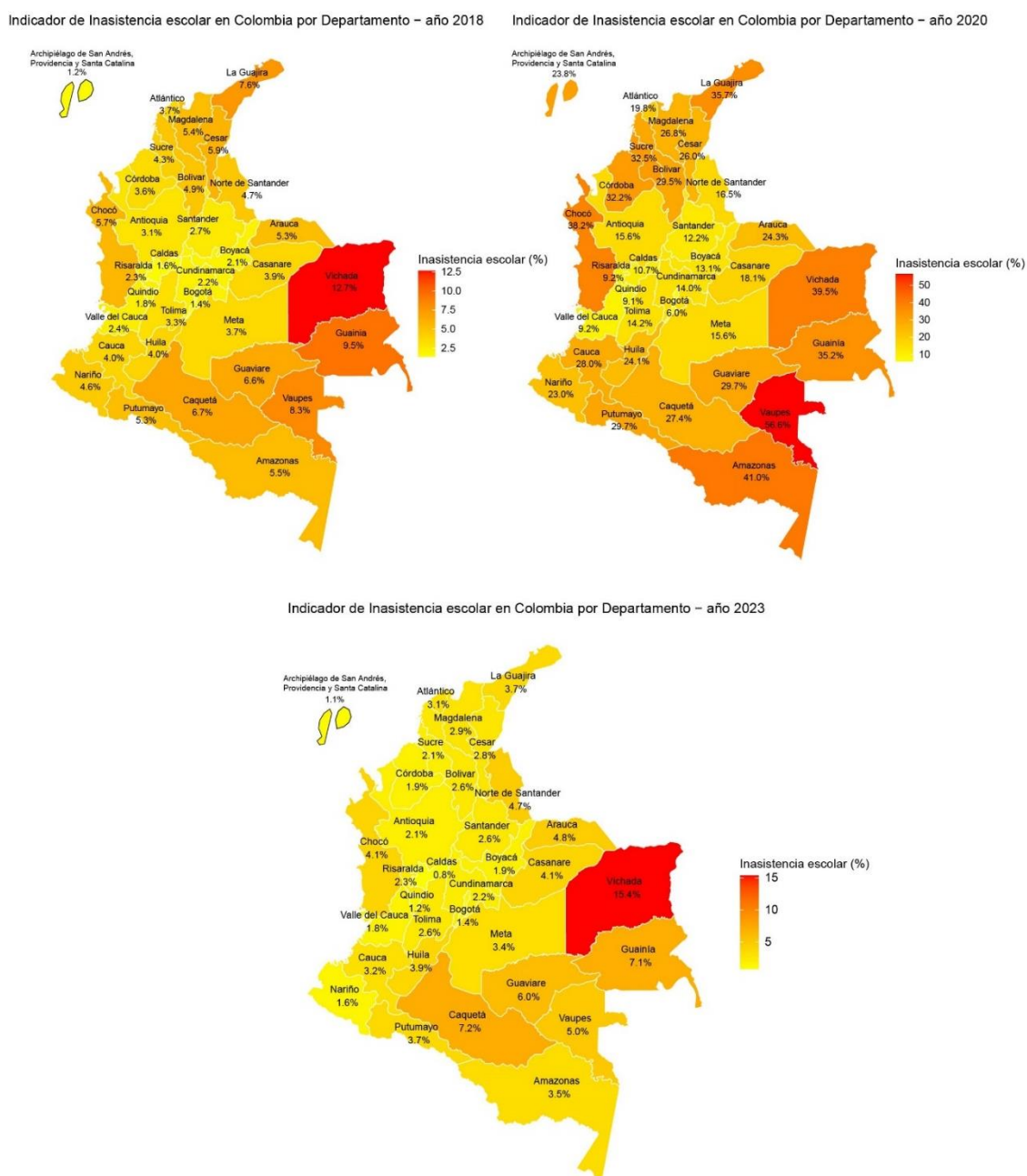
Inasistencia Escolar

La base de datos panel contiene la información del año 2018 al año 2023 a nivel departamental para el indicador de Inasistencia Escolar. El indicador de Inasistencia Escolar muestra en porcentaje la proporción de hogares privados, según el proceso de construcción del indicador DNP (2011) “un hogar está privado si menos del 100% de los niños entre 6 y 16 años asiste al colegio. Si en el hogar no hay niños en edad escolar, el hogar no se considera en privación de asistencia escolar” (Pág. 15). En este contexto de medición la inasistencia escolar es abarcada a nivel hogar, si el individuo o alguna de las personas que viven con él y están asociados con él tiene entre 6 y 16 años y no asiste a una institución educativa, se considera una persona privada del acceso a la educación, el indicador excluye a los hogares que no tienen entre sus integrantes personas con el rango de edad entre 6 y 16 años.

El indicador de inasistencia escolar pertenece a la dimensión de Condiciones de la niñez y juventud. Para el año 2018 Colombia presentó el indicador en 3,1%, para el año 2020 el indicador fue de 16,4% y para el año 2023 se estabilizó

nuevamente con 2,3%, DANE (2024c), la asistencia escolar se vio fuertemente afectada por el contexto de pandemia y aislamiento a nivel mundial del año 2020, la presencialidad que los jóvenes tenían en las instituciones educativas tuvo que cambiar por otras estrategias pedagógicas implementadas por el ministerio de educación, tales como plataformas virtuales, guías de contenidos pedagógicos, radio, llamadas telefónicas, entre otras, DANE (2021).

Ilustración 3 Indicador de Inasistencia Escolar en Colombia por Departamento años 2018 – 2020 – 2022.



Fuente: Elaboración propia basado en el DANE (2024d) “Anexo Departamental de Pobreza Multidimensional 2023”

Gráficamente se evidencia el salto de escala en la inasistencia escolar para el año 2020, los departamentos más afectados en este año fueron Vaupés, Amazonas, Vichada, Chocó y la Guajira, de este modo se intuye que los departamentos que pertenecen a la periferia del país y están más alejados de la capital fueron los más afectados, pues tenían menos recursos y métodos para garantizar el goce del derecho a la educación.

Mientras que, los departamentos de Caquetá, Guainía, Guaviare y Vichada son los que presentan más inasistencia escolar del 2018 – 2023, los departamentos de San Andrés, Providencia & Santa Catalina, Caldas y Quindío son los departamentos que mejor comportamiento presentan en este indicador, cabe resaltar e incluir a Bogotá en este grupo. Los departamentos de Caquetá y Vichada presentaron un crecimiento en su indicador de Inasistencia Escolar del 2018 al 2023 (Ilustración 3).

Bajo Logro Educativo

La base de datos panel contiene la información del año 2018 al año 2023 a nivel departamental para el indicador de Bajo Logro Educativo. El indicador de Bajo Logro Educativo muestra en porcentaje la proporción de hogares privados, según el proceso de construcción del indicador DNP (2011) “se considera que un hogar está privado cuando el promedio del logro educativo de las personas de 15 años y más que lo componen es menor a 9 años escolares” (Pág. 15), en este contexto de medición, el bajo logro educativo es abarcado a nivel hogar, si el individuo o alguna de las personas que viven con él y se asocian con el teniendo más de 15 años cumplidos no ha completado por lo menos 9 años de estudio en una institución educativa, se considera una persona privada del acceso a la educación.

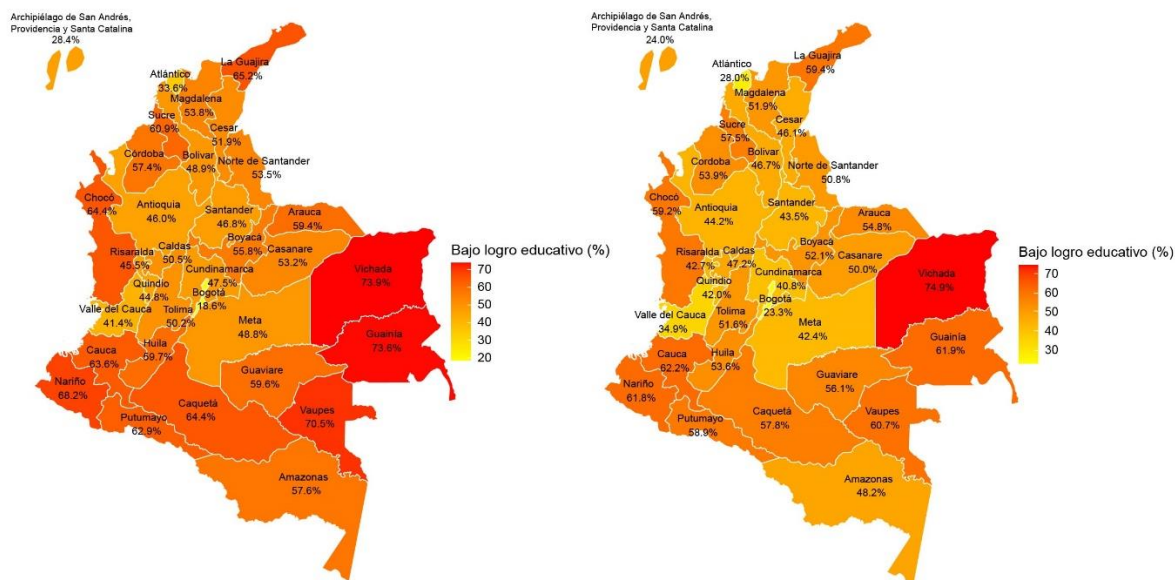
El indicador de bajo logro educativo pertenece a la dimensión de Condiciones educativas. Los departamentos de Caquetá, Cauca, Córdoba, Chocó, Huila, la Guajira, Nariño, Sucre, Arauca, Guainía, Guaviare, Vaupés y Vichada no han podido disminuir del 50% este indicador de pobreza, dejando en evidencia la gran oportunidad de mejora del nivel educativo en áreas periféricas del país. Los departamentos de Cundinamarca, Putumayo, Amazonas y Vaupés disminuyeron en 10% o más su indicador de Bajo logro Educativo del 2018 al 2023, el departamento del Vichada y la capital nacional Bogotá fueron los únicos que tuvieron un aumento de su bajo logro educativo del 2018 al 2023, se deben resaltar los departamentos de

Atlántico, San Andrés, Providencia & Santa Catalina, Valle del cauca y la capital Bogotá como los lugares con menos tasa de privación del 2018 al 2023 (Ilustración 4).

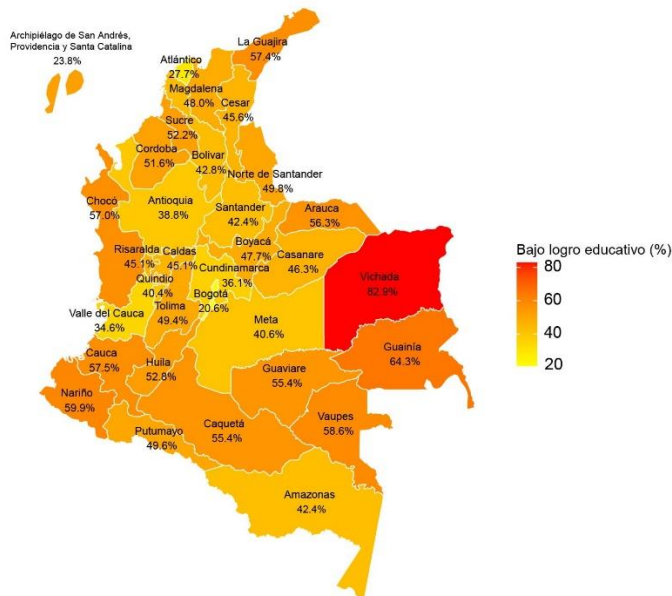
Ilustración 4 Indicador de Bajo Logro Educativo en Colombia por Departamento años 2018 – 2020 – 2022.

Indicador de Bajo logro educativo en Colombia por Departamento – año 2018

Indicador de Bajo logro educativo en Colombia por Departamento – año 2020



Indicador de Bajo logro educativo en Colombia por Departamento – año 2023



Fuente: Elaboración propia basado en el DANE (2024d) “Anexo Departamental de Pobreza Multidimensional 2023”

Para el año 2018 Colombia presentó un indicador de Bajo Logro Educativo del 44,5%, para el año 2020 el indicador fue de 44,2% y para el año 2023 fue de 39,4%, DANE (2024c), este indicador no presentó alteraciones en su comportamiento

decreciente por el contexto de pandemia y aislamiento del año 2020, la coyuntura mundial no tuvo un efecto grave para ese año y los años siguientes.

Rezago Escolar

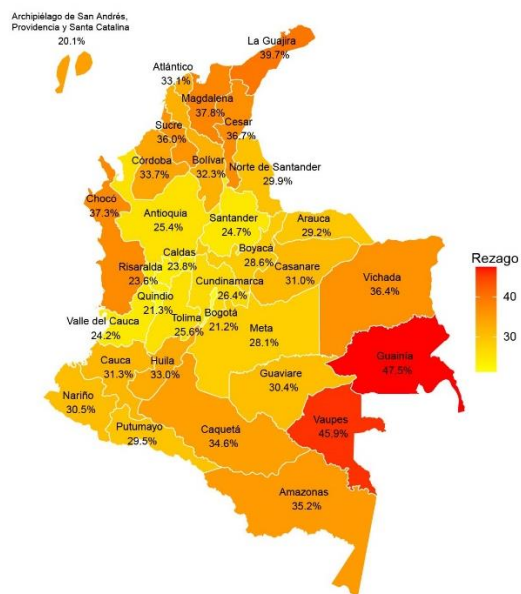
La base de datos panel contiene la información del año 2018 al año 2022 a nivel departamental para el indicador de Rezago Escolar. El indicador de Rezago Escolar muestra en porcentaje la proporción de hogares privados, según el proceso de construcción del indicador DNP (2011) “un hogar tiene privación [...] si alguno de los niños entre 7 y 17 años tiene rezago escolar. [...] En los casos en que no hay niños entre 7 y 17 años [...], se considera que no enfrenta rezago escolar” (Pág. 15). Teniendo en cuenta la información del indicador Rezago Escolar que contiene la (Tabla 1), el DNP en (2011) anunció que “el rezago escolar se define como la diferencia entre el número de años normativos y años aprobados por un niño entre 7 y 17 años de edad” (Pág. 15). En este contexto de medición el rezago escolar es abarcado a nivel hogar, si el individuo o alguna de las personas que viven con él y están asociadas con él tiene entre 7 - 17 años y la diferencia de años aprobados con años normativos de educación es negativa, se considera una persona privada del acceso a la educación.

El indicador de Rezago escolar pertenece a la dimensión de Condiciones de la Niñez y Juventud, gráficamente se identifica que los departamentos con mayor porcentaje de privación de la educación por este indicador están ubicados en la periferia del país, Atlántico, Cesar, Guainía, la Guajira, Magdalena, Chocó, Sucre y Vaupés no han podido disminuir del 30% el indicador de pobreza desde el año 2018, se resaltan los departamentos de Vaupés y Vichada que disminuyeron más del 10% el indicador de Rezago Escolar del 2018 al 2023, los departamentos de Guaviare y San Andrés, Providencia & Santa Catalina fueron los únicos que aumentaron su nivel de Rezago Escolar, del 2018 al 2023 (Ilustración 5).

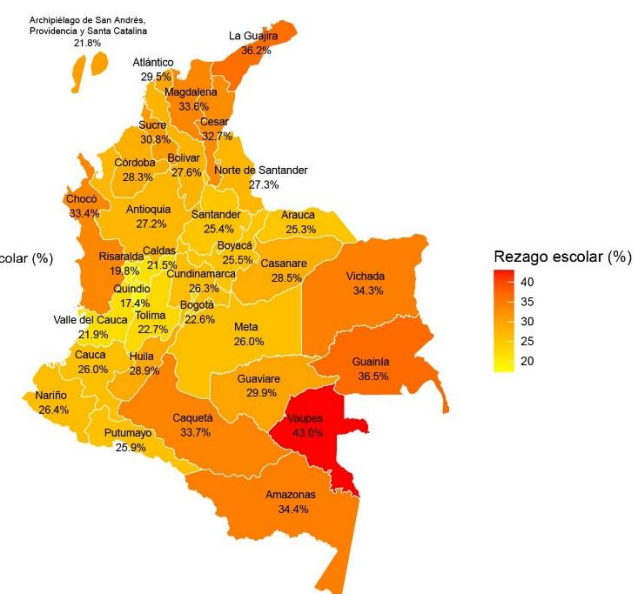
Para el año 2018 Colombia presentó el indicador de Rezago Escolar con un 27,4% de privación, para el año 2020 el indicador fue de 25,9% y para el 2023 disminuyó a 24,3%, DANE (2024c), de este modo se concluye que el indicador de Rezago Escolar no presentó alteraciones en su comportamiento decreciente por la coyuntura mundial del año 2020, tampoco presentó un efecto en la tendencia del indicador para los años siguientes.

Ilustración 5 Indicador de Rezago Escolar en Colombia por Departamento años 2018 – 2020 – 2022.

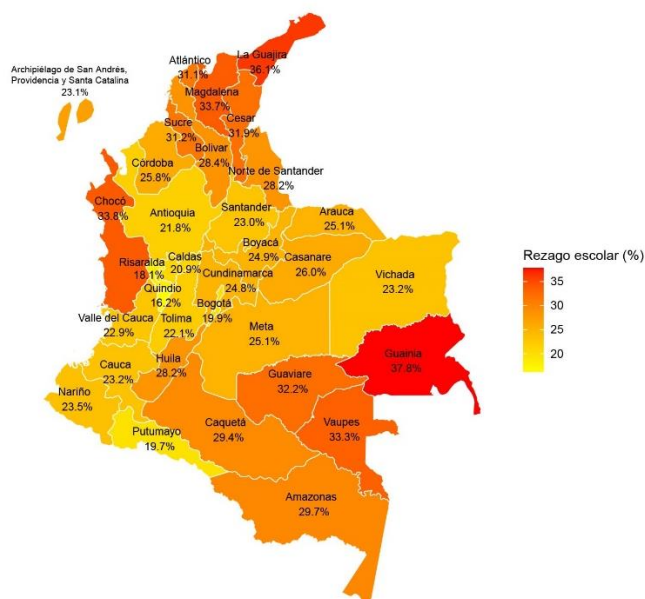
Indicador de Rezago escolar en Colombia por Departamento – año 2018



Indicador de Rezago escolar en Colombia por Departamento – año 2020



Indicador de Rezago escolar en Colombia por Departamento – año 2023



Fuente: Elaboración propia basado en el DANE (2024d) “Anexo Departamental de Pobreza Multidimensional 2023”

Metodología

Tipo de Investigación

La presente investigación es de corte cuantitativa – relacional, el periodo de estudio inicialmente es de 2018 – 2023 a nivel departamental para Colombia, teniendo en cuenta que el curso metodológico de la investigación inicia en 2015 con el inicio

de la vigencia que tuvieron los Objetivos de Desarrollo propuestos por la ONU, se aplicó un método de imputación utilizando el algoritmo EMB (expectation-maximization con bootstrap) el cual se detallará más adelante, el algoritmo se aplicó utilizando covariables asociadas al Gasto Público en Educación, la densidad poblacional del departamento, la migración interna del país y migración internacional por departamento, también se tomó en cuenta el Producto Interno Bruto Per-cápita, estas se encuentran explicadas en la tabla 1.

Tabla 1 Variables y covariables utilizadas para el estudio

Indicadores de pobreza en educación	Analfabetismo	Indica si las personas de 15 años y más del hogar son analfabetas o no
	Bajo logro educativo	Indica el número de años de escolaridad alcanzados por cada miembro del hogar. El hogar tiene un bajo logro educativo si alguno de sus miembros de 15 años o más tiene menos de 8 años de educación
	Inasistencia escolar	Indica si las personas entre 6 y 16 años asisten o no a una institución educativa
	Rezago escolar	Indica el rezago escolar. Esta se crea a partir de la variable "Años de Educación" de las personas del hogar en edades entre 7 y 17 años: • Tiene 7 años y no tiene al menos 1 año de educación • Tiene 8 años y no tiene al menos 2 años de educación • Tiene 9 años y no tiene al menos 3 años de educación • Tiene 17 años y no tiene al menos 11 años de educación
Covariables de gasto público	SGP_Educacion	Sistema General de Participaciones en educación a precios constantes con base en 2015 (Gasto público en educación) per-cápita
	SGP_Alimentacion_escolar	Sistema General de Participaciones en alimentación escolar a precios constantes con base en 2015 (Gasto público en alimentación escolar) per-cápita
Covariables demográficas	Migracion_interna	Densidad poblacional por departamento de la migración interna del país
	Migracion_internacional	Densidad poblacional por departamento de la migración internacional
	Densidad_Poblacional	Densidad poblacional del departamento
Covariable asociada al desarrollo económico	PIB	Producto Interno Bruto departamental a precios constantes con base en 2015 per-cápita

Nota: Las definiciones correspondientes a indicadores de pobreza provienen del DANE (2014)⁹, las series de tiempo extraídas provienen de repositorios del DANE y el DNP¹⁰.

⁹ Documento oficial de Metodología Índice De Pobreza Multidimensional (IPM).

¹⁰ - Las series de tiempo correspondientes a los cuatro indicadores provienen del Anexo Departamental de Pobreza Multidimensional 2022 generado por el DANE.

- Las series de tiempo correspondientes al sistema general de participaciones (Gasto público) provienen del Sistema de Información y Consulta de Distribuciones de Recursos Territoriales "SICODIS", y se deflactaron los precios con los datos del IPC provenientes del Anexo del IPC publicado por el DANE (2024a), los datos poblacionales se tomaron de las Proyecciones de Población reportadas por el DANE a nivel departamental del 2005-2019 y 2020-2050.

El método de imputación aplicado permite obtener más información del pasado para estimar de mejor modo la tendencia de las variables de respuesta (Analfabetismo, Inasistencia escolar, Bajo logro educativo, Rezago escolar), luego de este proceso la base de datos captura el comportamiento de los datos del año 2015 al año 2023.

Para el presente estudio se desarrollaron Modelos de Regresión Beta Multinivel Bayesianos con Enlace Logit, estos fueron estimados mediante el método de Monte Carlo con Cadenas de Markov (MCMC) a través del software cuantitativo R. Los indicadores de pobreza fueron contruidos con una muestra significativa de la población nacional a nivel Departamental, por lo tanto, este es el universo de estudio.

Imputación del panel de datos

Algoritmo de imputación

La imputación se aplicó utilizando el paquete “Amelia” del software estadístico Rstudio, este paquete originalmente se planteó con el algoritmo de datos faltantes EM desarrollado por (Dempster, Laird y Rubin, 1977), el algoritmo del paquete fue mejorado por Honaker, King & Blackwell (2011), dando como resultado una nueva versión llamada “AmeliaII” con el algoritmo EMB, actualmente se aplica con el nombre del paquete original.

El algoritmo utiliza el conocido algoritmo EM (expectation-maximization) en múltiples muestras bootstrap de los datos originales incompletos para extraer valores de los parámetros de los datos completos. Luego, el algoritmo obtiene valores imputados de cada conjunto de parámetros obtenidos por Bootstrap,

- El PIB fue extraído del Anexo PIB por Departamento 2022, la serie de tiempo se deflactó en base los datos de IPC publicados por el DANE (2024b) y los datos poblacionales corresponden a las Proyecciones de Población reportadas por el DANE a nivel departamental del 2005-2019 y 2020-2050.

- Las covariables demográficas se tomaron de (i) las Proyecciones de Población reportadas por el DANE a nivel departamental, los anexos correspondientes del 2005 - 2019 y 2020 - 2050, (ii) las proyecciones de los componentes del cambio demográfico 2020 – 2070, anexo Migración - estimación por sexo nacional y departamental con base en el CNPV 2018 publicadas por el DANE, (iii) las retroproyecciones de los componentes del cambio demográfico 1985 – 2019, anexo Migración - estimación por sexo nacional y departamental con base en el CNPV 2018 publicadas por el DANE. Las variables fueron descompuestas en densidad poblacional teniendo en cuenta el área por departamentos reportada por la Sociedad Geográfica de Colombia (2024) y la Alcaldía Mayor de Bogotá (2010).

reemplazando los valores faltantes con estos valores generados. (Honaker, King & Blackwell, 2011, pág. 2)

Honaker, King y Blackwell (2011) agregaron el componente Bootstrap al algoritmo EM¹¹ planteado por (Dempster, Laird y Rubin, 1977), dando como resultado la siguiente ecuación de verosimilitud con expectativas iteradas para la aplicación del Bootstrap:

$$p(D_o|\theta) = \int p(D|\theta) dDm \quad \text{Ecuación 1}$$

Donde $\theta = (\mu, \Sigma)$, hace referencia a los datos completos, D_o hace referencia a los datos observados, Dm hace referencia a los datos faltantes y $p(D|\theta)$ hacer referencia a la probabilidad de obtener los datos (faltantes y observados) dado θ . De este modo se tiene que la probabilidad que tienen los datos observados dado θ es igual a integrar sobre todas las realizaciones posibles de Dm . De este modo se puede concluir la siguiente ecuación posterior del algoritmo utilizando la ley de Bayes:

$$p(\theta|D_o) \propto p(D_o|\theta) * p(\theta) \quad \text{Ecuación 2}$$

Donde $p(\theta)$ es la probabilidad a priori de los parámetros y $p(D_o|\theta)$ es la ecuación de verosimilitud con la aplicación del Bootstrap.

Para el presente caso de investigación se siguió el ejemplo propuesto por Del Callejo, Canal, Vernazza, Urruticoechea & Álvarez (2022) de incluir restricciones en la imputación para mejorar su resultado, la imputación se delimitó con la variable aleatoria Departamentos (que indica el departamento de Colombia), adicionalmente se indicó la variable Año como una serie temporal y se delimitó la variable de respuesta con valores entre 0 y 1¹² para generar mayor la asertividad en el proceso de imputación.

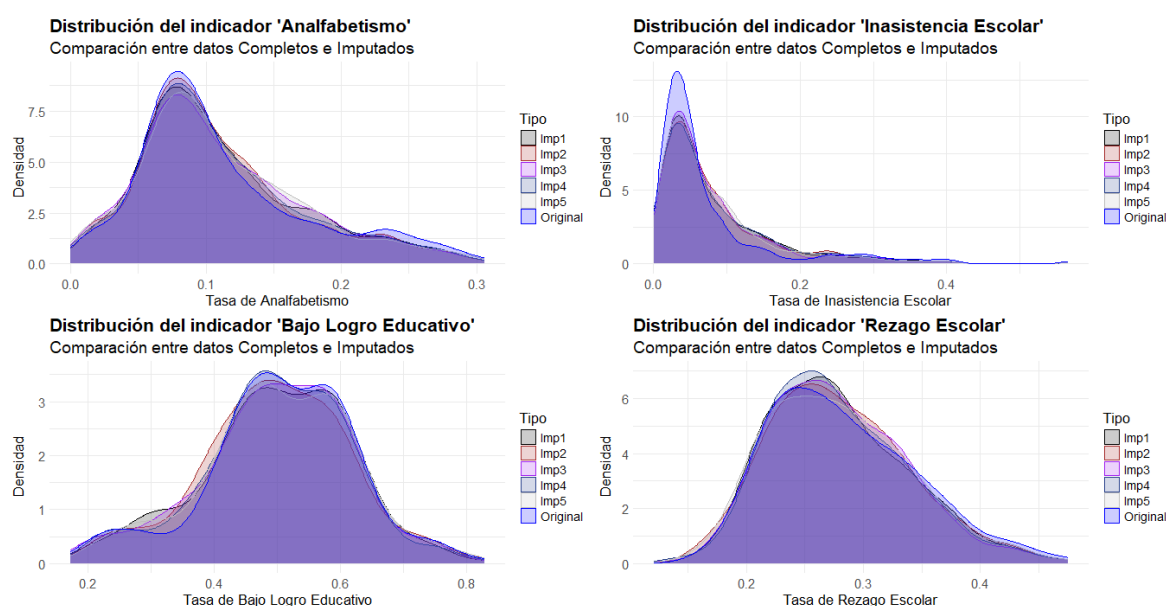
Teniendo en cuenta a los autores Honaker, King & Blackwell (2011) “a menos que la tasa de valores faltantes sea muy alta, $m = 5$ ”. (pág. 2), la salida del paquete en Rstudio son 5 bases de datos panel ya imputadas. En la Ilustración 6 se observa

¹¹ adicionalmente se debe tener en cuenta que “El algoritmo EM [...], converge hacia las estimaciones máximo-verosímiles de los parámetros” (García, Albaladejo & Fernández, 2006, pág. 254).

¹² Las variables de respuesta son los 4 indicadores de pobreza en educación (Analfabetismo, Inasistencia escolar, Bajo logro educativo Y Rezago escolar), los indicadores originalmente toman valores entre 0 y 100, para obtener el comportamiento de una distribución beta se dividieron entre 100, por eso la variable respuesta toma valores entre 0 y 1.

el comparativo de la densidad que tienen los casos completos (datos aún sin imputar) contra las 5 imputaciones generadas por el algoritmo EMB para cada variable respuesta, gráficamente la densidad es similar entre las variables y las imputaciones, Analfabetismo e Inasistencia Escolar tienen asimetría a la izquierda, Bajo logro Educativo tiende a tener un comportamiento bimodal.

Ilustración 6 Comparativo de la densidad que tienen las variables imputadas contra los casos completos.



Fuente: Elaboración propia basado en el DANE, el DNP y el proceso de imputación realizado.

Medidas de severidad de Rubín

Para evaluar la calidad y el impacto que las imputaciones realizadas tienen sobre las bases de datos Little & Rubín (2020) propusieron tres medidas de calidad para identificar cuán severamente el proceso realizado puede afectar los análisis estadísticos posteriores sobre las bases de datos.

Por ello, para todas las bases imputadas se ajustaron las variables de respuesta a una distribución beta y se calcularon las medidas de severidad colapsando la media de las estimaciones y la varianza de las imputaciones con las fórmulas propuestas por Rubín (2020) para este tipo de aplicaciones:

$$\bar{\theta} = \frac{1}{m} \sum_{k=1}^m \hat{\theta}^{(k)} \quad \text{Ecuación 3}$$

Donde $\bar{\theta}$ es la combinación de los estimadores obtenidos en la imputación, m es el número de imputaciones y $\hat{\theta}^{(k)}$ es el estimador del parámetro θ en la k -ésima imputación.

$$V(\hat{\theta}) = \bar{\Sigma}^{13} + \left(1 + \frac{1}{m}\right) B^{14} \text{ Ecuación 4}$$

Donde $V(\hat{\theta})$ es la combinación de las varianzas obtenidas después de cada imputación, $\bar{\Sigma}$ es la varianza que se obtuvo de cada imputación, B es la varianza entre las imputaciones y m es el número de imputaciones.

A continuación, se presenta la formulación propuesta por Little & Rubin (2020) para evaluar las tres medidas de calidad sobre la imputación:

$$\lambda = (1 + m^{-1})q^{-1}tr(B[\hat{V}(\bar{\theta})]^{-1}) \text{ Ecuación 5}$$

Donde λ está relacionada con la información que se retiene después de la imputación (Little & Rubin, 2020), indica la proporción de varianza debido a los datos faltantes, m es el número de imputaciones, q es el número de bases imputadas, B es la matriz de varianza que existe entre las imputaciones, tr indica la traza de la matriz y $V(\hat{\theta})$ es la combinación de las varianzas obtenidas después de cada imputación.

$$r = (1 + m^{-1})q^{-1}tr(B\bar{\Sigma}^{-1}) \text{ Ecuación 6}$$

Donde r identifica la variabilidad que se tiene dentro de las imputaciones y la varianza total (Little & Rubin, 2020), indica el incremento relativo de la varianza y $\bar{\Sigma}$ es la varianza que se obtuvo de cada imputación para los estimadores.

$$\gamma = (1 + r)^{-1}[r + 2(\mathcal{V}^{15} + 3)^{-1}] \text{ Ecuación 7}$$

Donde γ ajusta la varianza estimada de los parámetros obtenidos de las bases imputadas para mostrar la incertidumbre adicional introducida por el proceso de imputación (Little & Rubin, 2020), indica el porcentaje de información faltante sobre θ .

¹³ $\bar{\Sigma} = \frac{1}{m} \sum_{k=1}^m \hat{\Sigma}^{(k)}$. (Little & Rubin, 2020)

¹⁴ $B = \frac{1}{m-1} \sum_{k=1}^m (\hat{\theta}^{(k)} - \bar{\theta})(\hat{\theta}^{(k)} - \bar{\theta})^T$. (Little & Rubin, 2020)

¹⁵ Grados de libertad $\mathcal{V} = \frac{n(n-1)}{(n+2)}(1 - \lambda)$, donde n es el tamaño de la muestra o cantidad de datos. (Little & Rubin, 2020),

Tabla 2 Medidas de severidad Rubín para las variables de pobreza en educación imputadas.

Indicador	λ	r	γ
Analfabetismo	9,76%	11,32%	10,84%
Inasistencia Escolar	10,21%	12,62%	11,87%
Bajo Logro Educativo	10,66%	12,03%	11,42%
Rezago Escolar	9,14%	10,91%	10,51%

Nota: Todas las medidas de severidad evaluadas para las 4 variables de respuesta son menores al 30. Elaboración propia basado en el DANE, el DNP y el proceso de imputación realizado.

Los resultados obtenidos luego del proceso de imputación y evaluación de las medidas de severidad se detallan en la Tabla 2, no superan el 15% en cuanto a incremento de la varianza por los datos faltantes y el proceso de imputación, los resultados son robustos e indican que la inferencia estadística que se realice sobre los datos no tendrá un gran sesgo debido al proceso realizado y la falta de datos.

Aplicación del modelo

Se consideraron varios factores al momento de identificar el modelo de análisis, (i) se tiene una base de datos con estructura panel y covariables relacionadas al contexto de educación, por lo cual se optó por aplicar un modelo multinivel o jerárquico que tome en cuenta el factor aleatorio “Departamento” y las covariables descritas, (ii) las variables de respuesta tienen una distribución beta, (iii) aunque se realizó un proceso de imputación robusto para capturar de mejor forma la tendencia de los datos, el histórico de la base de datos no es muy extenso, (iv) el propósito del modelo es encontrar la relación existente entre las covariables y la variable de respuesta.

Los modelos Beta dan flexibilidad en cuanto a la distribución de la variable respuesta, (Ferrari & Cribari-Neto, 2004) para lo cual se puede ajustar la distribución que presentan los indicadores de pobreza en educación, Nelder y McCullagh (1989) explican que utilizar un enlace logit en la variable de respuesta de los modelos, mejora la versatilidad y la interpretabilidad que tienen los resultados, es fundamental al momento de modelar proporciones. Adaptando el enlace logit a la distribución beta, se transforma la media de la distribución que pertenece a valores entre 0 y 1 en escalas del predictor lineal, se obtendrán entonces valores que abarcan de $-\infty$ a $+\infty$, la ecuación planteada por Berkson para enlaces logit (1944) se ve de la siguiente forma:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) \quad \text{Ecuación 8}$$

Donde p es la probabilidad de éxito. Teniendo en cuenta que un modelo lineal se presenta en su forma general como en la ecuación 9 (Nelder y McCullagh, 1989), la adaptación del logit a un modelo con distribución beta en su variable de respuesta se presenta en la ecuación 10 (Ferrari & Cribari-Neto, 2004).

$$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p \quad \text{Ecuación 9}$$

Donde la variable respuesta η es el predictor lineal, $\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_p$ son los coeficientes del modelo (tomando β_0 como intercepto) y $X_1, X_2, X_3, \dots, X_p$ como variables independientes o predictoras.

$$\mu_k = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \beta_3 X_{3k} + \dots + \beta_p X_{pk})}} \quad \text{Ecuación 10}$$

Donde μ_k corresponde a una variable con valores entre 0 y 1, es la media de la variable dependiente o variable respuesta para la k -ésima observación, e corresponde a la función exponencial, los β son los coeficientes resultantes del modelo, los X son las variables predictoras en la k -ésima observación. Teniendo en cuenta nuevamente la estructura de datos panel, se debe agregar una variable aleatoria que dé cuenta del comportamiento y relación de las variables de una forma jerárquica, agregando este nuevo factor la forma general del modelo queda planteada en la siguiente forma por Grün, Kosmidis y Zeileis (2012), ecuación 11.

$$g(\mu_{ij}) = \eta_{ij} = X_{ij}\beta + Z_{ij}b_i \quad \text{Ecuación 11}$$

Donde $g(.)$ Es la función de enlace (para este caso corresponde a la ecuación 8), μ_{ij} corresponde a la media de la respuesta con distribución beta para la observación j en el grupo i , de este modo, η_{ij} corresponde al predictor lineal o variable de respuesta, X_{ij} corresponde a una matriz que contiene los efectos fijos para la observación j en el grupo i , β corresponde a los coeficientes de los efectos fijos, Z_{ij} corresponde a una matriz que contiene los efectos aleatorios para la observación j en el grupo i , b_i corresponde a los efectos aleatorios del grupo i , asumiendo que se distribuya $N(0, D)$ donde D es una matriz de covarianza de los efectos aleatorios. Agregando el factor aleatorio en la ecuación planteada por Ferrari & Cribari-Neto (2004) se tiene la ecuación 12.

$$\mu_{ki} = \frac{1}{1 + e^{-(\beta_0 + u_{0j} + \beta_1 X_{1ki} + \beta_2 X_{2ki} + \beta_3 X_{3ki} + \dots + \beta_p X_{pki})}} \quad \text{Ecuación 12}$$

Donde μ_{ki} corresponde a una variable con valores entre 0 y 1, es la media de la variable dependiente o variable respuesta para la k-ésima observación en el i-ésimo grupo, u_{0j} corresponde al factor aleatorio que varía por grupo i, e corresponde a la función exponencial, los β son los coeficientes resultantes del modelo, los X son las variables predictoras en la k-ésima observación y en el i-ésimo grupo.

Para mejorar la flexibilidad del modelo y aplicarlo efectivamente a datos reales y complejos como los planteados en este estudio Gelman, Carlin, Stern, Dunson, Vehtari, Rubin (2013) mencionaron que la aplicación del método de Monte Carlo con Cadenas de Markov¹⁶ (MCMC) facilita la aproximación de distribuciones complejas y también facilita el análisis si se poseen modelos con una gran cantidad de parámetros o variables.

Teniendo en cuenta la ecuación base de la inferencia bayesiana (ecuación 2), se utilizó el paquete “brms” del software estadístico Rstudio para realizar la aplicación de modelos de regresión beta multinivel con enlace logit mediante el método MCMC, se realizó el ejercicio para las 5 bases imputadas por cada variable de respuesta (Analfabetismo, Inasistencia escolar, Rezago escolar, Bajo logro educativo), la variable “Departamento” se tomó como el factor aleatorio del modelo. Para igualar la escala de las covariables y facilitar el cálculo técnico del modelo, se cambiaron a escala logarítmica, teniendo en cuenta la distribución beta de la variable respuesta se utilizaron los parámetros de “link = logit” que realiza la aplicación del enlace logit de la ecuación 8 y “link_phi = log” que es la medida de precisión que garantiza que los valores estimados sean siempre positivos Ferrari & Cribari-Neto (2004), Cribari-Neto & Zeileis (2010).

Utilizando la ecuación 3 y la ecuación 4 planteadas por Little & Rubin (2020), se colapsaron los resultados obtenidos de los modelos aplicados a las 5 imputaciones para cada variable de respuesta, de este modo se unifican los resultados de varianza, coeficientes y significatividad por cada variable de respuesta.

¹⁶ Una Cadena de Markov es un proceso estocástico que indica un nuevo punto inicial. (Bertsekas & Tsitsiklis, 2002), (Ross, 2000), (Peña, 2002)

Resultados

La combinación de los modelos imputados se encuentra a nivel de variable respuesta para la interpretación de las estimaciones de los Modelos de Regresión Beta Multinivel Bayesianos con Enlace Logit se debe recurrir a la ecuación logística inversa, Berkson (1944) y McCullagh & Nelder (1989) indicaron que debido al enlace logit aplicado se debe recurrir a este recurso.

$$y = \frac{e^{est.log}}{1 + e^{est.log}} \quad \text{Ecuación 13}$$

En donde est.log hace referencia a la estimación obtenida en el modelo con enlace logit, adicionalmente, la evaluación y análisis de los resultados se realiza al 95% y 90% de confianza.

Analfabetismo

Tabla 3 Covariables significativas al 95% y 90% de confianza del modelo combinado de Analfabetismo

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
SGP_Alimentacion_escolar	0,35913	0,13626	0,58200	0,17209	0,54617
PIB	-0,82026	-1,17174	-0,46878	-1,11523	-0,52529

Nota: El modelo muestra significatividad en las variables de Gasto público en alimentación escolar y PIB per-cápita al 95% y al 90% de confianza. Elaboración propia basado en el DANE y el DNP.

Utilizando la función logística para transformar las estimaciones a una escala de probabilidad, cuando el Gasto público en alimentación escolar per-cápita (considerando la escala logarítmica de la variable predictora) aumenta en 1%, la probabilidad de analfabetismo aumenta en aproximadamente 58.88%, esto implica que, un pequeño incremento del gasto en este rubro está asociado a un gran incremento en el indicador de analfabetismo, la relación obtenida indica que los mayores programas de alimentación escolar se implementan en departamentos con mayor tasa de analfabetismo, llevando la misma línea de análisis, un aumento del 1% en el PIB per-cápita produce un decrecimiento de 30.57% en la probabilidad del indicador de analfabetismo, este resultado va de la mano con las teorías del crecimiento planteadas por Mankiw, Romer & Weil (1992), Barro (1991), Keller (2006) & Romer (1990) y Gary Becker (1964), pues se espera que el aumento del bienestar económico mejore la situación de pobreza educativa.

Tabla 4 Departamentos significativos al 95% y 90% de confianza del modelo combinado de Analfabetismo

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
Amazonas	-0,26601	-0,56383	0,03182	-0,51595	-0,01606
La Guajira	0,36464	0,05442	0,67486	0,10429	0,62498
Putumayo	-0,57169	-1,05879	-0,08458	-0,98047	-0,16290
San Andrés	-0,52114	-1,00462	-0,03766	-0,92689	-0,11539
Sucre	0,29076	0,01549	0,56603	0,05974	0,52177

Nota: El modelo muestra significatividad al 95% de confianza en los departamentos de la Guajira, Putumayo, Sucre, San Andrés, Providencia & Santa catalina, al 90% de confianza Amazonas también es significativo. Elaboración propia basado en el DANE y el DNP.

Los departamentos de Putumayo y San Andrés, Providencia & Santa Catalina presentan una menor probabilidad de analfabetismo en comparación al total nacional al 95% de confianza, el departamento del Amazonas también presenta este comportamiento con un 90% de confianza, por otro lado, los departamentos de la Guajira y Sucre tienen mayor probabilidad de analfabetismo en comparación al total nacional con un 95% de confianza.

Inasistencia Escolar

Tabla 5 Covariables significativas al 95% y 90% de confianza del modelo combinado de Inasistencia Escolar.

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
SGP_Alimentacion_escolar	0,21818	-0,02391	0,46028	0,01501	0,42136
PIB	-0,37244	-0,69814	-0,04675	-0,64578	-0,09911

Nota: El modelo muestra significatividad en la variable de PIB per-cápita al 95% y de Gasto público en alimentación escolar al 90% de confianza. Elaboración propia basado en el DANE y el DNP.

El aumento en 1% del PIB per-cápita produce un decrecimiento de 40.79% en la probabilidad del indicador de Inasistencia escolar con el 95% de confianza, por el contrario, el aumento en 1% del Gasto público en alimentación escolar aumenta en 55.43% la probabilidad del indicador de Inasistencia escolar con un 90% de confianza.

Bajo Logro Educativo

Tabla 6 Covariables significativas al 95% y 90% de confianza del modelo combinado de Bajo Logro Educativo

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
SGP_Educacion	0,61594	0,19657	1,03532	0,26399	0,96790
PIB	-0,43393	-0,67740	-0,19045	-0,63825	-0,22960

Nota: El modelo muestra significatividad en las variables de Gasto público en educación y PIB per-cápita al 95% y 90% de confianza. Elaboración propia basado en el DANE y el DNP.

Utilizando la función logística para transformar las estimaciones a una escala de probabilidad, cuando el Gasto público en educación per-cápita aumenta en 1%, la probabilidad de Bajo logro educativo aumenta en aproximadamente 64.93% con el 95% de confianza, la relación obtenida indica que los mayores programas de gasto público en educación se implementan en departamentos con mayor indicador de Bajo logro educativo, por el contrario, un aumento del 1% en el PIB per-cápita produce un decrecimiento de 39.82% en la probabilidad del indicador de Bajo Logro educativo con el 95% de confianza, lo cual es bastante intuitivo y acertado con la teoría.

Tabla 7 Departamentos significativos al 95% y 90% de confianza del modelo combinado de Bajo Logro Educativo

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
Atlántico	-0,29751	-0,50896	-0,08606	-0,47497	-0,12006
Bogotá	-0,23123	-0,47096	0,00851	-0,43242	-0,03003
Nariño	0,17047	-0,02747	0,36840	0,00436	0,33658
San Andrés	-0,33770	-0,58799	-0,08741	-0,54775	-0,12765
Vichada	0,43500	0,10358	0,76641	0,15686	0,71313

Nota: El modelo muestra significatividad al 95% de confianza en los departamentos de Atlántico, Vichada y San Andrés, Providencia & Santa Catalina, al 90% de confianza Bogotá y Nariño también son significativos. Elaboración propia basado en el DANE y el DNP.

Los departamentos de Atlántico y San Andrés, Providencia & Santa Catalina presentan una menor probabilidad de bajo logro educativo en comparación al total nacional al 95% de confianza, Bogotá también presenta este comportamiento con un 90% de confianza, por otro lado, Vichada al 95% de confianza presenta mayor probabilidad de bajo logro educativo en comparación al total nacional, el departamento de Nariño también presenta este comportamiento al 90% de confianza.

Rezago Escolar

Tabla 8 Covariables significativas al 95% y 90% de confianza del modelo combinado de Rezago Escolar

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
SGP_Alimentacion_escolar	0,13149	-0,00327	0,26626	0,01839	0,24459
PIB	-0,16726	-0,32691	-0,00760	-0,30124	-0,03327
Densidad_poblacional	-0,02560	-0,05262	0,00142	-0,04828	-0,00293

Nota: El modelo muestra significatividad en la variable PIB per-cápita al 95%, además, en las variables Gasto público en alimentación escolar y Densidad poblacional se tiene significatividad 90% de confianza. Elaboración propia basado en el DANE y el DNP.

El aumento en 1% del PIB per-cápita produce un decrecimiento de 45.83% en la probabilidad del indicador de rezago escolar con el 95% de confianza, el incremento de 1% en la Densidad poblacional por departamento decrece la probabilidad de Rezago escolar en 49,36% con el 90% de confianza, por el contrario, el aumento de 1% en el Gasto público en alimentación escolar trae asociado un incremento del 53.28% en la probabilidad de rezago escolar del departamento con el 90% de confianza.

Tabla 9 Departamentos significativos al 95% y 90% de confianza del modelo combinado de Rezago escolar

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
Atlántico	0,15360	0,00635	0,30085	0,03002	0,27718
La Guajira	0,12116	-0,01887	0,26119	0,00364	0,23868
Quindío	-0,19083	-0,34989	-0,03178	-0,32431	-0,05735
Tolima	-0,12112	-0,25847	0,01622	-0,23639	-0,00586

Nota: El modelo muestra significatividad al 95% de confianza en los departamentos de Atlántico y Quindío, al 90% de confianza los departamentos de la Guajira y Tolima también son significativos. Elaboración propia basado en el DANE y el DNP.

Por su parte, el departamento del Atlántico presenta una mayor probabilidad de Rezago escolar en comparación al total nacional con el 95% de confianza, el departamento de la Guajira también presenta este comportamiento con un 90% de confianza, por otro lado, Quindío al 95% de confianza presenta menor probabilidad de Rezago escolar en comparación al total nacional, Tolima también presenta este comportamiento al 90% de confianza.

Análisis integral de los resultados.

En primera medida se resalta la falta de relación entre la densidad poblacional de los departamentos con su nivel de pobreza en educación, si bien el rezago escolar tiene relación con esta variable predictora, al 90 % de confianza, no tiene un valor significativo en las otras dimensiones de pobreza en educación.

El Gasto público en educación y el Gasto público en alimentación escolar son significativos al 95% de confianza en los modelos de Analfabetismo y Bajo logro

educativo, para los otros dos indicadores son significativos al 90% de confianza, la relación indica que un aumento en el gasto público per cápita en el rubro de educación está asociado a un incremento elevado en los indicadores de pobreza en educación, si bien, el análisis es contra intuitivo, la revisión empírica de los datos históricos da a conocer que en los departamentos con contextos más críticos de pobreza en educación, se tienen los mayores proyectos e iniciativas de gasto público en educación.

El PIB per-cápita tiene una relación significativa al 95% de confianza con los cuatro indicadores de pobreza en educación, teóricamente, la relación es complementaria entre la producción nacional y el nivel educativo del país, de este modo, la evidencia estadística da un completo apoyo a la teoría clásica y demuestra que Colombia ha mejorado su panorama de pobreza en educación debido a unas mejores condiciones económicas y de calidad de vida del país.

A nivel departamental se concluye que, en su mayoría, los departamentos que presentan niveles más elevados de privación al derecho a la educación en comparación al total nacional, están alejados de la capital, lo cual demuestra la falta de igualdad en el acceso que tiene el país.

En relación con los objetivos propuestos por la OCDE en su agenda de mejora para el 2030, Colombia ha mantenido un comportamiento decreciente en sus indicadores de pobreza (excepto para el año 2020 en Inasistencia escolar), el gasto público enfocado en los departamentos con mas bajos niveles de educación tiene efecto a mediano y largo plazo, en línea con Galvis (2015) y Chamorro (2013), se tiene menos control del gasto público en los departamentos periféricos del país lo cual se debe tener en cuenta si se busca presentar un mejor nivel educativo real en el año 2030.

Conclusiones

El año 2020 fue un punto atípico en el comportamiento decreciente de la pobreza en educación, demostró la diferencia de recursos y métodos de acceso a la educación que tienen los departamentos del país.

El uso de la imputación por el algoritmo EMB permitió capturar de mejor forma las tendencias e información relevante del contexto educativo en Colombia para poder

explicar con modelos de regresión beta multinivel bayesianos y enlace logit, que a nivel departamental en Colombia (i) no existe una relación significativa entre la densidad poblacional y la pobreza en educación a nivel departamental, (ii) la relación entre el Gasto público y la pobreza en educación es significativa y relevante, pues el gasto público es mayor en áreas con menos acceso a la educación, (iii) la producción nacional tiene una relación positiva e importante con los indicadores de pobreza en educación, lo cual va de la mano con los estudios empíricos y teóricos de desarrollo económico.

La teoría del desarrollo económico que relaciona positivamente el nivel educativo y la producción nacional se ve reflejada claramente en el contexto colombiano, pero también pone en evidencia la desigualdad de acceso a la educación a nivel departamental. Aunque el Gasto público es mayor en áreas con peores indicadores, aún se debe mejorar el enfoque y control de los recursos.

En relación con los objetivos de desarrollo planteados por la OCDE, Colombia no presenta un comportamiento acelerado, pero si constante de disminución en la pobreza educativa, por lo cual se espera seguir con la tendencia y presentar un mejor nivel educativo para el año 2030.

Referencias

- Alkire, S., & Foster, J. (2011). *Counting and Multidimensional Poverty Measurement*. Oxford University Press.
- Angulo, R. C., Díaz, B. Y., & Pardo, R. P. (2013). *A Counting Multidimensional Poverty Index in Public Policy Context: the case of Colombia*. Oxford Department of International Development.
- Angulo, R. C., Díaz, Y. C., & Pardo, R. P. (2011). *Índice de Pobreza Multidimensional para Colombia (IPM-Colombia) 1997-2010*. ARCHIVOS DE ECONOMÍA.
- Baltagi, B. H. (2013). *Econometric analysis of panel data*. John Wiley & Sons.
- Banco de la República. (2023). Persistencia de la desigualdad en la educación de Colombia: un análisis histórico. *BANREP*.
- Banco Mundial. (2022). *The State of Global Learning Poverty: 2022 Update*. worldbank.org.
- Banco Mundial. (2022a). *El 70 % de los niños de 10 años se encuentran en situación de pobreza de aprendizajes y no pueden leer y comprender un texto simple*. bancomundial.org. COMUNICADO DE PRENSA N.º 2022/074/EDU.
- Banco Mundial. (2024). Educación: Panorama general. *bancomundial.org*, <https://www.bancomundial.org/es/topic/education/overview>.

- BancoMundial. (17 de Oct de 2023). *bancomundial.org*. Obtenido de Entendiendo a la pobreza: https://www.bancomundial.org/es/topic/poverty/overview?cid=ECR_GA_worldbank_ES_EX_TP_Search&s_kwid=AL!18468!3!665400387540!b!!g!!como%20se%20mide%20la%20pobreza&gad_source=1#2
- Barro, R. J. (1977). *Economic Growth in a cross section of Countries*”, *Quarterly Journal of Economics*. Vol. CVI.
- Becker, G. (1964). Human Capital. *Chicago IL: University of Chicago Press*. .
- Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American Statistical Association*, págs. 357 - 365.
- Bertsekas, D., & Tsitsiklis, J. (2002). *Introducyion to Probability*. USA: Athena Scientific.
- Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control*. Wiley.
- Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 395-411.
- Castillo-Mateo, J., & Guajardo, A. C. (2022). *Métodos Monte Carlo basados en cadenas de Markov*. Zaragoza: Universidad de Zaragoza.
- Chamorro, D. B. (2013). Eficiencia del Gasto Público en educación en Colombia. Un Análisis a Nivel Departamental para los años 2005 y 2010. *Santiago de Cali: Universidad del Valle*. Obtenido de Santiago de Cali: Universidad del Valle.
- Cribari-Neto, F., & Ferrari, S. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, 799-815.
- Cribari-Neto, F., & Zeileis, A. (2010). Beta Regression in R. *Journal of statistical Software*.
- DANE . (2024). Glosario - Gran encuesta integrada de hogares. *Departamento Administrativo Nacional de Estadística*.
- DANE. (2014). METODOLOGÍA ÍNDICE DE POBREZA MULTIDIMENSIONAL (IPM). Departamento Administrativo Nacional de Estadística.
- DANE. (2021). *Anexo técnico No. 2. Construcción del indicador inasistencia escolar del Índice de Pobreza Multidimensional 2020*. Bogotá. Colombia: Departamento Administrativo Nacional de Estadística.
- DANE. (2022). *Gastos del gobierno general por finalidad (GGF). Anexo estadístico ggf*. Obtenido de <https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-anuales/gastos-del-gobierno-general-por-finalidad>
- DANE. (2023). *Boletín Técnico Pobreza multidimensional en Colombia año 2022*. Departamento Administrativo Nacional de Estadística.
- DANE. (2023a). *dane.gov.co*. Obtenido de Cambio demográfico de la Retroproyección de la población 1985-2019: <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/estimaciones-del-cambio-demografico>

- DANE. (2023b). *dane.gov.co*. Obtenido de Proyecciones de los componentes del cambio demográfico a nivel nacional 2020-2070 y departamental 2020-2050 actualización post COVID-19: <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/estimaciones-del-cambio-demografico>
- DANE. (2023c). Serie departamental de población por área, para el periodo 2005-2019. Proyecciones y retroproyecciones de población departamental para el periodo 1985-2019 y 2020-2050 con base en el CNPV 2018. *Departamento Administrativo Nacional de Estadística*.
- DANE. (2023d). Serie departamental de población por área, para el periodo 2020-2050. *Departamento Administrativo Nacional de Estadística. Proyecciones y retroproyecciones de población departamental para el periodo 1985-2019 y 2020-2050 con base en el CNPV 2018*.
- DANE. (2024a). Índice de Precios al Consumidor - IPC. Variaciones porcentuales - mayo 2024. Departamento Administrativo Nacional de Estadística.
- DANE. (2024b). PIB por departamento. *Departamento Administrativo Nacional de Estadística. Anexos estadísticos PIB departamental 2005-2023*.
- DANE. (2024c). Pobreza multidimensional. Anexo nacional. Departamento Administrativo Nacional de Estadística.
- DANE. (2024d). Pobreza multidimensional. Anexo departamental. *Departamento Administrativo Nacional de Estadística*.
- Del Callejo Canal, D., Canal-Martínez, M., Vernazza, E., Urruticoechea, A., & Álvarez-Vaz, R. (2022). *Consideraciones a la imputación múltiple. Un caso de estudio con datos panel*. Cuadernos Del CIMBAGE, 1(24), 33-47. .
- Dempster, A., Laird, N. M., & Rubin, D. (1977). *Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society B, 39, 1–38.
- DNP. (2015). *Plan Nacional de Desarrollo 2014 - 2018 (Tomo 1)*. . Bogotá DC: Departamento Nacional de Planeación.
- DNP. (2017). Panorámica regional. Pobreza monetaria y multidimensional departamental, necesidad de políticas públicas diferenciadas . *Departamento Nacional de Planeación*.
- DNP. (2019). *Plan de Desarrollo Nacional 2018 - 2022 (Tomo 1)*. . Bogotá DC: Departamento Nacional de Planeación.
- DNP. (2023). *Plan de Desarrollo Nacional 2022 - 2026* . Bogotá DC: Departamento Nacional de Planeación.
- DNP. (2024). Distribuciones de los recursos del Sistema General de Participaciones (SGP). Detalle histórico por cantidad y concepto SGP. *SICODIS*, Series gasto público y gasto público en educación 2015 -2022.
- Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford Press.
- Forero, D., & Parrado, C. V. (2019). *Gasto en el Sector Educativo en Colombia*. FEDESARRROLLO.
- Galvis Aponte, L. A. (2015). LA EFICIENCIA DEL GASTO PÚBLICO EN EDUCACIÓN EN COLOMBIA. . *Cartagena: Universidad Tecnológica de Bolívar*.

- Galvis, L. A. (2015). La eficiencia del gasto público en educación en Colombia. *Economía & Región*.yr: 2015;no: 9;iss: 2.
- García, J. G., Albaladejo, J. P., & Fernández, J. A. (2006). Métodos de inferencia estadística con datos faltantes. Estudio de simulación sobre los efectos de las estimaciones. *ESTADÍSTICA ESPAÑOLA*, págs. 241 - 270.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC Press. .
- Grün, B., Kosmidis, I., & Zeileis, A. (2012). *Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned*. Journal of Statistical Software. Vol. 48. #11.
- Hamilton, J. D. (1994). Time Series Analysis. *Princeton University Press*.
- Honaker, J., King, G., & Blackwell, M. (2011). *Amelia II: A Program for Missing Data*. Journal of Statistical Software, 45(7), 1–47.
- Hyndman, R., & Athanasopoulos, G. (2021). Forecasting: principles and practice. *3rd edition, OTexts: Melbourne, Australia*.
- Keller, K. (2006). Investment in Primary, Secondary, and Higher Education and the Effects on Economic Growth. *Contemporary Economic Policy* 24 (1), 18 – 34.
- Little, R. J., & Rubin, D. B. (2020). *Statistical Analysis with Missing Data*. John Wiley & Sons, Hoboken, 3rd edition.
- Little, R. J., & Rubin, D. B. (2020). *Statistical Analysis with Missing Data*. John Wiley & Sons, Hoboken, NJ, 3rd edition.
- Mankiw, N., Romer, D., & Weil, D. (1992). "A Contribution to the empirics of Economic Growth. *Quarterly Journal of Economics*, (107), 3, pp.407- 437.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. Chicago: Second Edition. Chapman and Hall.
- Ministerio de Educación. (2022). *Decreto 1411 de 2022 Ministerio de Educación*. MINEDUC. Función pública.
- ONU. (2024). *Objetivos de desarrollo sostenible*. Obtenido de un.org: <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>
- Peña, P. (2002). *Cadena de Markov para modelar la permanencia de pacientes en una unidad de cuidados intensivos cardiológicos*. Chile: Universidad Técnica Federico Santa María.
- Podestá, A. (2020). *Gasto público para impulsar el desarrollo económico e inclusivo y lograr los Objetivos de Desarrollo Sostenible*. Santiago: serie Macroeconomía del Desarrollo, N° 214, Comisión Económica para América Latina y el Caribe (CEPAL).
- Robert, C., & Casella, G. (2009). *Introducing Monte Carlo Methods with R*. 1st Edition. Springer.
- Rodríguez, M., Freire, M., & Pais, C. (2017). *El efecto del gasto público sanitario y educativo en la determinación del bienestar de los países de la OCDE: un modelo con datos de panel*. Cuadernos de Economía, 41, 104-118.
- Romer, P. (1990). Endogenous Technological Change. *Journal of Political Economy*.

- Ross, S. (2000). *Introduction to Probability Models*. USA: Academic Press.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.
- SDP. (2010). *Densidades Urbanas: El caso de Bogotá*. BOGOTÁ D.C., COLOMBIA 2010: ALCALDÍA MAYOR DE BOGOTÁ. Boletín No. 22.
- Sen, A. (1998). Las teorías del desarrollo a principios del siglo XXI. *Cuadernos de Economía, Banco Interamericano de Desarrollo*.
- SOGEOCOL. (Sociedad Geográfica de Colombia). *Departamentos de Colombia*. Obtenido de [sogeocol.edu.co: https://sogeocol.edu.co/Home_B/esgeo/departamentos/](https://sogeocol.edu.co/Home_B/esgeo/departamentos/)
- Speagle, J. S. (2019). *A Conceptual Introduction to Markov Chain Monte Carlo Methods*. Harvard.
- Trucco, D. (2023). *Mejorar la educación es crucial para un desarrollo social y económico inclusivo y sostenible*. CEPAL.
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Nelson Education.

Anexos

Anexo 1

Analfabetismo resultados modelo

- Coeficientes e intervalos de confianza para el modelo combinado de Analfabetismo

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
Intercept	13,26694	1,07053	25,46335	3,03139	23,50249
SGP_Educacion	-0,39919	-1,04339	0,24502	-0,93982	0,14145
SGP_Alimentacion_escolar	0,35913	0,13626	0,58200	0,17209	0,54617
PIB	-0,82026	-1,17174	-0,46878	-1,11523	-0,52529
Densidad_poblacional	0,03922	-0,02668	0,10512	-0,01608	0,09452

Nota: El modelo muestra significatividad en las variables de Gasto público en alimentación escolar y PIB per-cápita al 95% y al 90% de confianza. Elaboración propia basado en el DANE y el DNP.

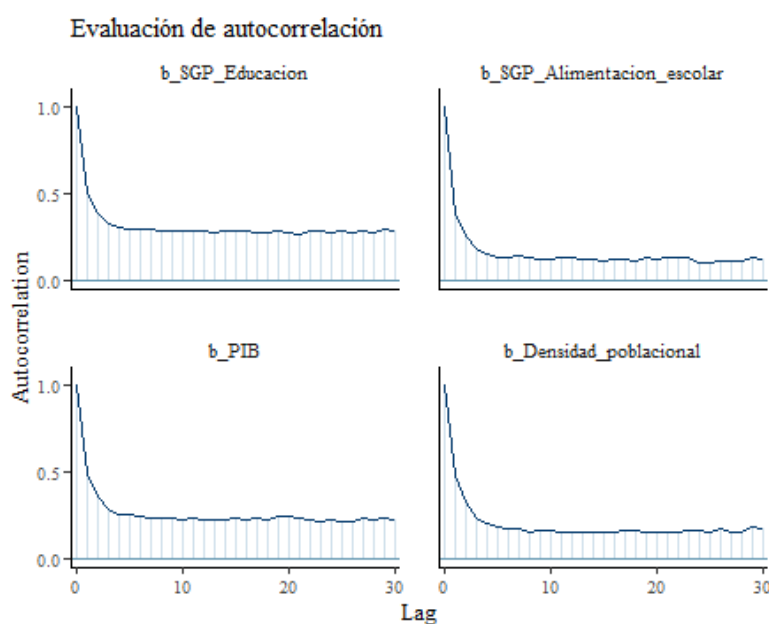
- Coeficientes e intervalos de confianza para el modelo combinado de Analfabetismo evaluado por departamento

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
Amazonas	-0.26601	-0.56383	0.03182	-0.51595	-0.01606
Antioquia	0.18989	-0.12703	0.50682	-0.07608	0.45587
Arauca	0.01461	-0.27563	0.30485	-0.22897	0.25819
Atlántico	0.03407	-0.25762	0.32575	-0.21072	0.27886
Bogotá	-0.22825	-0.61885	0.16235	-0.55605	0.09955
Bolívar	0.25365	-0.05174	0.55904	-0.00264	0.50994
Boyacá	0.14520	-0.34164	0.63203	-0.26337	0.55376
Caldas	0.17057	-0.12582	0.46695	-0.07817	0.41930
Caquetá	-0.14723	-0.45990	0.16543	-0.40963	0.11516
Casanare	-0.03877	-0.50129	0.42374	-0.42693	0.34938
Cauca	0.03618	-0.23806	0.31043	-0.19397	0.26634
Cesar	0.19602	-0.07430	0.46633	-0.03084	0.42287
Chocó	0.14898	-0.09729	0.39524	-0.05769	0.35565
Córdoba	0.17429	-0.09643	0.44500	-0.05291	0.40148
Cundinamarca	-0.22134	-0.68997	0.24729	-0.61463	0.17194
Guainía	-0.04812	-0.42809	0.33185	-0.36700	0.27076
Guaviare	-0.14456	-0.44617	0.15705	-0.39768	0.10856
Huila	0.02901	-0.23064	0.28866	-0.18889	0.24692
La Guajira	0.36464	0.05442	0.67486	0.10429	0.62498
Magdalena	0.03654	-0.26565	0.33873	-0.21707	0.29014
Meta	0.11277	-0.29503	0.52057	-0.22947	0.45501
Nariño	-0.03225	-0.27315	0.20866	-0.23442	0.16993
Norte de Santander	-0.22222	-0.48866	0.04423	-0.44583	0.00139
Putumayo	-0.57169	-1.05879	-0.08458	-0.98047	-0.16290
Quindío	0.05595	-0.29448	0.40639	-0.23814	0.35005
Risaralda	-0.02748	-0.43849	0.38353	-0.37241	0.31745
San Andrés	-0.52114	-1.00462	-0.03766	-0.92689	-0.11539
Santander	0.17252	-0.36921	0.71425	-0.28212	0.62715
Sucre	0.29076	0.01549	0.56603	0.05974	0.52177
Tolima	0.02239	-0.26334	0.30812	-0.21740	0.26218
Valle del Cauca	0.10128	-0.24478	0.44733	-0.18914	0.39170
Vaupés	-0.14074	-0.47065	0.18917	-0.41761	0.13613
Vichada	0.02571	-0.33916	0.39059	-0.28050	0.33193

Nota: El modelo muestra significatividad al 95% de confianza en los departamentos de La guajira, Putumayo, Sucre, San Andrés, Providencia & Santa catalina, al 90% de confianza Amazonas también es significativo. Elaboración propia basado en el DANE y el DNP.

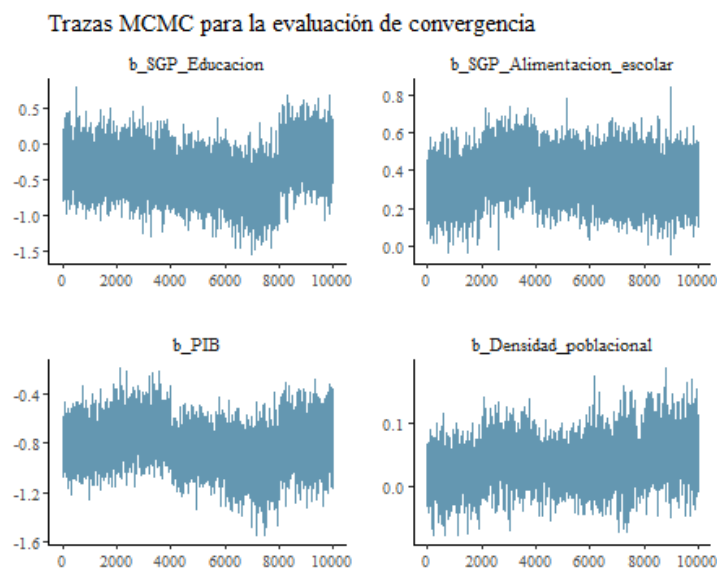
Analfabetismo pruebas del modelo

- Gráfico de autocorrelación para las covariables del modelo de Analfabetismo.



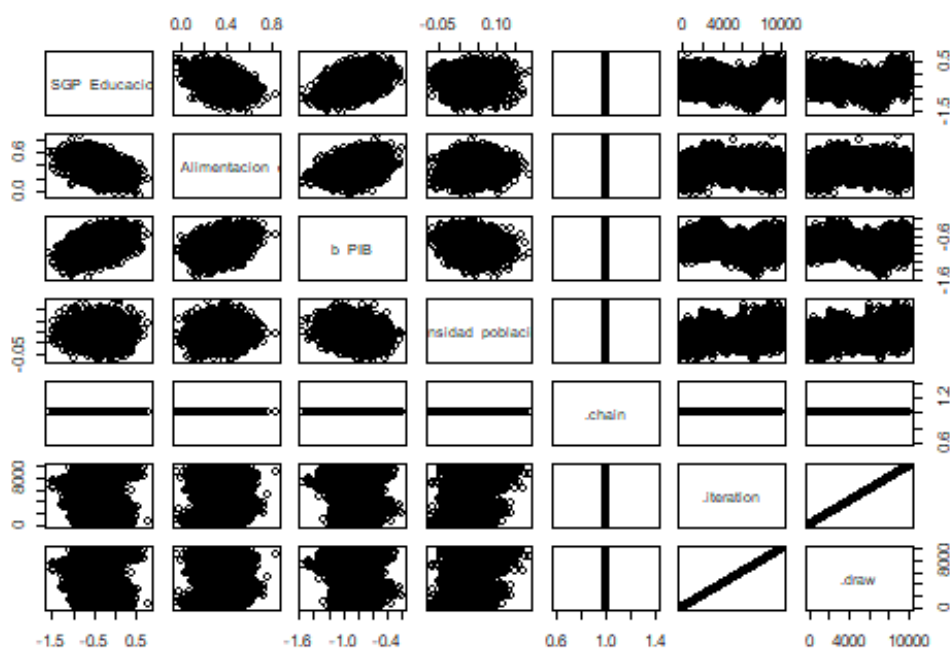
Nota: El modelo muestra convergencia hacia a distribución posterior de los parámetros del modelo. Elaboración propia basado en el DANE y el DNP.

- Gráfico de evaluación de convergencia para las covariables del modelo de Analfabetismo.



Nota: Las covariables del modelo no evidencian una tendencia marcada, cada 2000 iteraciones representan los resultados para un modelo en una imputación. Elaboración propia basado en el DANE y el DNP.

- Prueba de Correlación y convergencia. Gráfico de pares para evaluar la relación entre las covariables del modelo de Analfabetismo



Nota: Las covariables del modelo no evidencian tendencias marcadas entre sus respectivas relaciones, por lo tanto, son independientes entre sí, el comportamiento de los valores a través de las iteraciones y las dos cadenas de Markov tampoco muestra tendencias marcadas, por lo cual, se logró una correcta convergencia. Elaboración propia basado en el DANE y el DNP.

Anexo 2

Inasistencia Escolar resultados modelo

- Coeficientes e intervalos de confianza para el modelo combinado de Inasistencia Escolar

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
Intercept	2,80669	-7,84368	13,45707	-6,13139	11,74477
SGP_Educacion	-0,07141	-0,65630	0,51347	-0,56226	0,41944
SGP_Alimentacion_escolar	0,21818	-0,02391	0,46028	0,01501	0,42136
PIB	-0,37244	-0,69814	-0,04675	-0,64578	-0,09911
Densidad_poblacional	-0,00908	-0,06969	0,05152	-0,05994	0,04178

Nota: El modelo muestra significatividad en la variable de PIB per-cápita al 95% y de Gasto público en alimentación escolar al 90% de confianza. Elaboración propia basado en el DANE y el DNP.

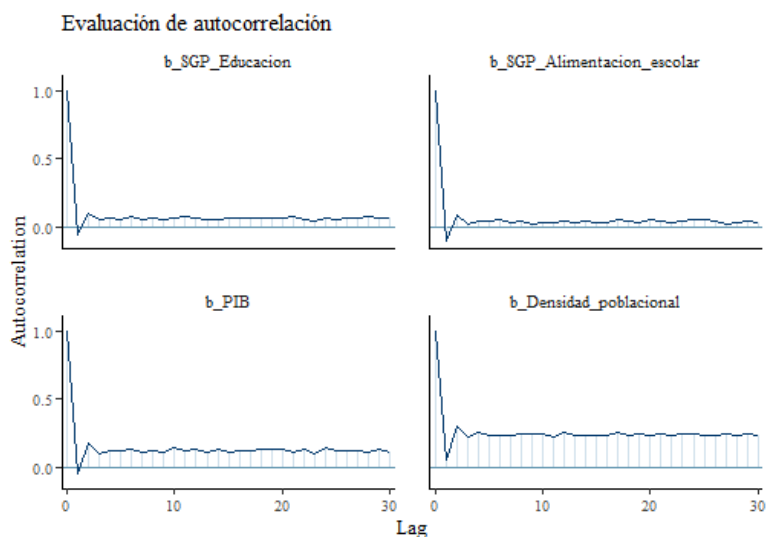
- Coeficientes e intervalos de confianza para el modelo combinado de Inasistencia Escolar evaluado por departamento

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
Amazonas	0,00154	-0,13585	0,13892	-0,11376	0,11684
Antioquia	-0,00160	-0,14318	0,13999	-0,12042	0,11723
Arauca	0,00843	-0,13888	0,15574	-0,11519	0,13206
Atlántico	0,01795	-0,12944	0,16534	-0,10574	0,14164
Bogotá	0,00421	-0,13792	0,14635	-0,11507	0,12350
Bolívar	0,01498	-0,12850	0,15846	-0,10543	0,13539
Boyacá	0,00884	-0,13903	0,15671	-0,11526	0,13294
Caldas	-0,01031	-0,15215	0,13153	-0,12934	0,10873
Caquetá	0,00077	-0,13576	0,13729	-0,11381	0,11534
Casanare	0,01912	-0,13680	0,17503	-0,11173	0,14997
Cauca	-0,00110	-0,13663	0,13444	-0,11484	0,11265
Cesar	-0,00124	-0,14528	0,14280	-0,12212	0,11964
Chocó	-0,01721	-0,16184	0,12742	-0,13859	0,10416
Córdoba	-0,02131	-0,16538	0,12275	-0,14222	0,09959
Cundinamarca	0,00056	-0,14693	0,14805	-0,12322	0,12433
Guainía	0,00661	-0,13431	0,14753	-0,11166	0,12487
Guaviare	-0,00555	-0,14422	0,13311	-0,12193	0,11082
Huila	0,01209	-0,12784	0,15202	-0,10535	0,12952
La Guajira	-0,00148	-0,14829	0,14533	-0,12469	0,12173
Magdalena	-0,00334	-0,14284	0,13615	-0,12041	0,11372
Meta	0,01769	-0,12987	0,16524	-0,10615	0,14152
Nariño	-0,01811	-0,16657	0,13036	-0,14270	0,10649
Norte de Santander	-0,00486	-0,14673	0,13701	-0,12392	0,11420
Putumayo	-0,00664	-0,15057	0,13730	-0,12743	0,11416
Quindío	-0,01744	-0,16333	0,12845	-0,13987	0,10499
Risaralda	-0,01071	-0,15205	0,13064	-0,12932	0,10791
San Andrés	-0,01817	-0,17025	0,13391	-0,14580	0,10946
Santander	0,01441	-0,13562	0,16444	-0,11150	0,14032
Sucre	-0,01519	-0,15747	0,12710	-0,13460	0,10422
Tolima	-0,01199	-0,15912	0,13515	-0,13546	0,11149
Valle del Cauca	0,00167	-0,14180	0,14513	-0,11873	0,12207
Vaupés	-0,00050	-0,14078	0,13978	-0,11822	0,11722
Vichada	0,02737	-0,12615	0,18090	-0,10147	0,15622

Nota: El modelo no muestra significatividad al 95% o al 90% en ningún departamento. Elaboración propia basado en el DANE y el DNP.

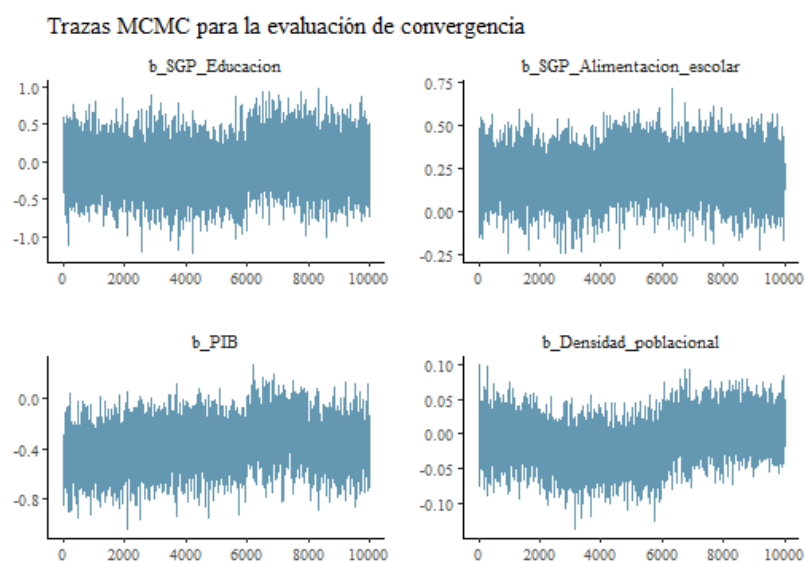
Inasistencia Escolar pruebas del modelo

- Gráfico de autocorrelación para las covariables del modelo de Inasistencia Escolar.



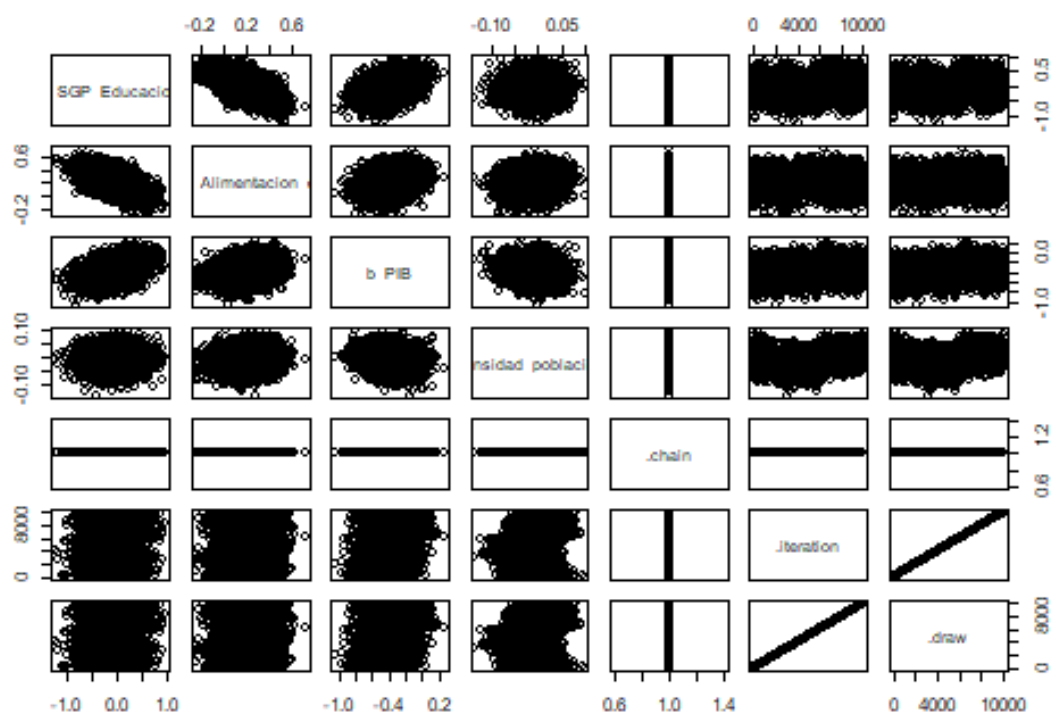
Nota: El modelo muestra convergencia hacia a distribución posterior de los parámetros del modelo. Elaboración propia basado en el DANE y el DNP.

- Gráfico de evaluación de convergencia para las covariables del modelo de Inasistencia Escolar.



Nota: Las covariables del modelo no evidencian una tendencia marcada, cada 2000 iteraciones representan los resultados para un modelo en una imputación. Elaboración propia basado en el DANE y el DNP.

- Prueba de Correlación y convergencia. Gráfico de pares para evaluar la relación entre las covariables del modelo de Inasistencia Escolar



Nota: Las covariables del modelo no evidencian tendencias marcadas entre sus respectivas relaciones, por lo tanto, son independientes entre sí, el comportamiento de los valores a través de las iteraciones y las dos cadenas de Markov tampoco muestra tendencias marcadas, por lo cual, se logró una correcta convergencia. Elaboración propia basado en el DANE y el DNP.

Anexo 3

Bajo Logro Educativo

- Coeficientes e intervalos de confianza para el modelo combinado de Bajo Logro Educativo

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
Intercept	-1,44656	-9,51354	6,62042	-8,21658	5,32346
SGP_Educacion	0,61594	0,19657	1,03532	0,26399	0,96790
SGP_Alimentacion_escolar	0,05031	-0,11021	0,21084	-0,08440	0,18503
PIB	-0,43393	-0,67740	-0,19045	-0,63825	-0,22960
Densidad_poblacional	0,00817	-0,03285	0,04919	-0,02626	0,04259

Nota: El modelo muestra significatividad en las variables de Gasto público en educación y PIB per-cápita al 95% y 90% de confianza. Elaboración propia basado en el DANE y el DNP.

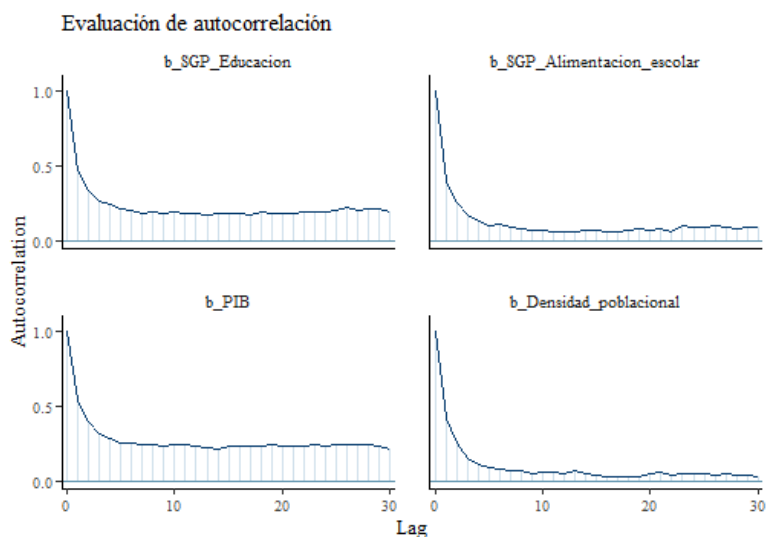
- Coeficientes e intervalos de confianza para el modelo combinado de Bajo Logro Educativo evaluado por departamento

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
Amazonas	-0,22685	-0,42181	-0,03190	-0,39046	-0,06324
Antioquia	0,09722	-0,09129	0,28573	-0,06099	0,25542
Arauca	0,15713	-0,07880	0,39305	-0,04087	0,35512
Atlántico	-0,29751	-0,50896	-0,08606	-0,47497	-0,12006
Bogotá	-0,23123	-0,47096	0,00851	-0,43242	-0,03003
Bolívar	-0,10273	-0,31382	0,10837	-0,27988	0,07443
Boyacá	-0,08409	-0,33153	0,16336	-0,29175	0,12357
Caldas	0,07181	-0,12592	0,26954	-0,09413	0,23775
Caquetá	0,01284	-0,19304	0,21873	-0,15994	0,18563
Casanare	0,12975	-0,10442	0,36392	-0,06678	0,32627
Cauca	0,13263	-0,04541	0,31067	-0,01679	0,28205
Cesar	-0,00694	-0,21952	0,20565	-0,18535	0,17147
Chocó	-0,08714	-0,29800	0,12372	-0,26410	0,08982
Córdoba	-0,08116	-0,26229	0,09998	-0,23317	0,07086
Cundinamarca	0,05127	-0,18593	0,28848	-0,14780	0,25034
Guainía	-0,02614	-0,25447	0,20219	-0,21776	0,16548
Guaviare	-0,04687	-0,24494	0,15120	-0,21310	0,11935
Huila	0,10398	-0,09051	0,29847	-0,05924	0,26720
La Guajira	0,06373	-0,14503	0,27249	-0,11147	0,23892
Magdalena	-0,11090	-0,30936	0,08756	-0,27745	0,05565
Meta	0,12824	-0,08538	0,34186	-0,05103	0,30752
Nariño	0,17047	-0,02747	0,36840	0,00436	0,33658
Norte de Santander	-0,04264	-0,27846	0,19317	-0,24055	0,15526
Putumayo	0,09362	-0,19169	0,37893	-0,14582	0,33306
Quindío	-0,06063	-0,28937	0,16811	-0,25260	0,13134
Risaralda	0,03830	-0,16135	0,23796	-0,12925	0,20586
San Andrés	-0,33770	-0,58799	-0,08741	-0,54775	-0,12765
Santander	0,10566	-0,09985	0,31118	-0,06681	0,27813
Sucre	-0,03981	-0,26409	0,18447	-0,22803	0,14841
Tolima	0,02731	-0,16186	0,21649	-0,13145	0,18607
Valle del Cauca	0,04286	-0,18945	0,27517	-0,15210	0,23782
Vaupés	-0,09908	-0,35710	0,15893	-0,31561	0,11745
Vichada	0,43500	0,10358	0,76641	0,15686	0,71313

Nota: El modelo muestra significatividad al 95% de confianza en los departamentos de Atlántico, Vichada y San Andrés, Providencia & Santa catalina, al 90% de confianza Bogotá y Nariño también son significativos. Elaboración propia basado en el DANE y el DNP.

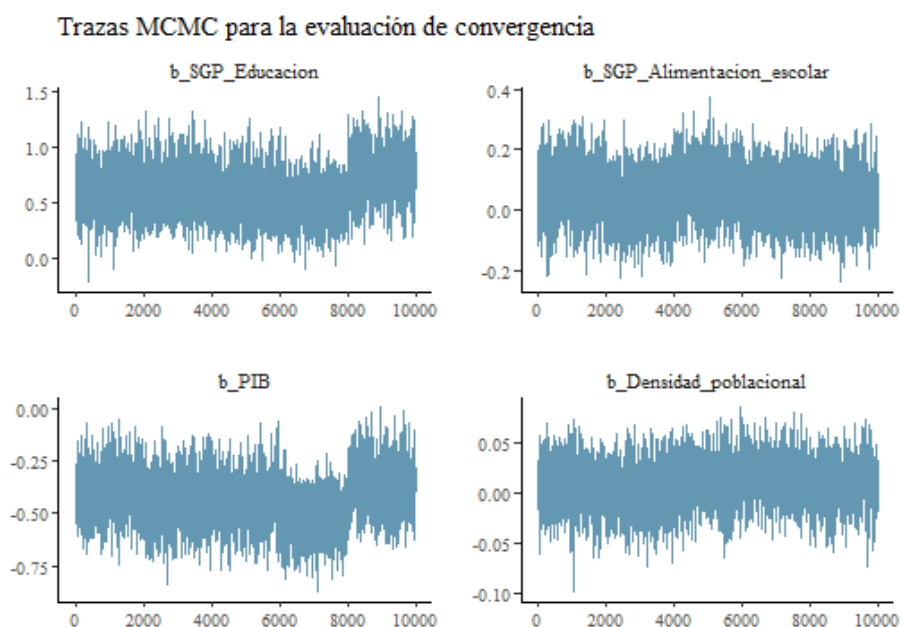
Bajo Logro Educativo pruebas del modelo

- Gráfico de autocorrelación para las covariables del modelo de Bajo Logro Educativo.



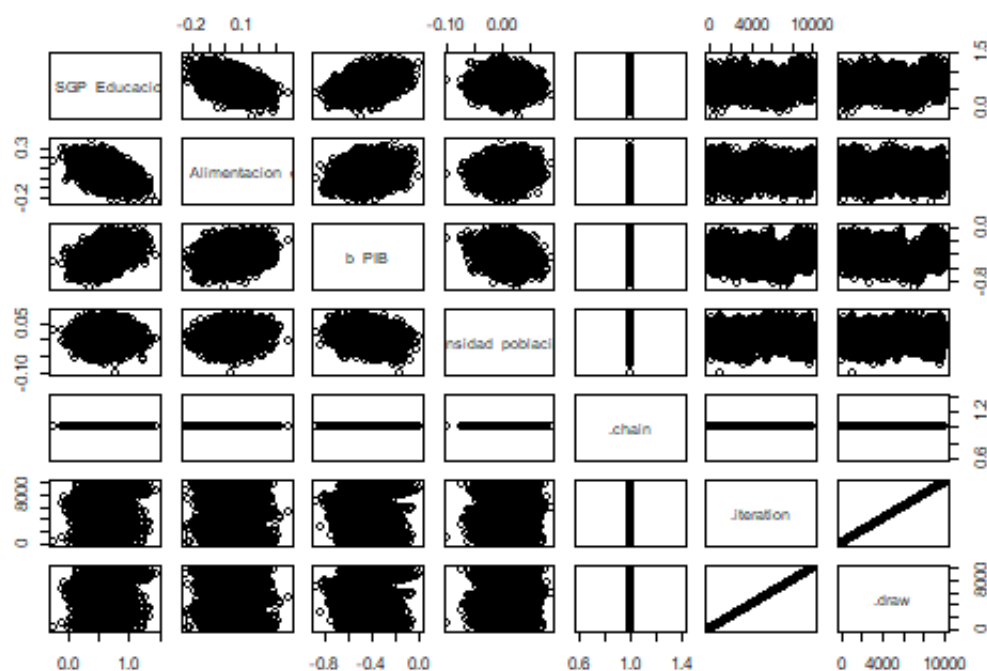
Nota: El modelo muestra convergencia hacia a distribución posterior de los parámetros del modelo. Elaboración propia basado en el DANE y el DNP.

- Gráfico de evaluación de convergencia para las covariables del modelo de Bajo Logro Educativo.



Nota: Las covariables del modelo no evidencian una tendencia marcada, cada 2000 iteraciones representan los resultados para un modelo en una imputación. Elaboración propia basado en el DANE y el DNP.

- Prueba de Correlación y convergencia. Gráfico de pares para evaluar la relación entre las covariables del modelo de Bajo Logro Educativo.



Nota: Las covariables del modelo no evidencian tendencias marcadas entre sus respectivas relaciones, por lo tanto, son independientes entre sí, el comportamiento de los valores a través de las iteraciones y las dos cadenas de Markov tampoco muestra tendencias marcadas, por lo cual, se logró una correcta convergencia. Elaboración propia basado en el DANE y el DNP.

Anexo 4

Rezago Escolar

- Coeficientes e intervalos de confianza para el modelo combinado de Rezago Escolar

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
Intercept	-0,50919	-5,34009	4,32171	-4,56341	3,54503
SGP_Educacion	0,09690	-0,19528	0,38908	-0,14831	0,34210
SGP_Alimentacion_escolar	0,13149	-0,00327	0,26626	0,01839	0,24459
PIB	-0,16726	-0,32691	-0,00760	-0,30124	-0,03327
Densidad_poblacional	-0,02560	-0,05262	0,00142	-0,04828	-0,00293

Nota: El modelo muestra significatividad en la variable PIB per-cápita al 95%, además, en las variables Gasto público en alimentación escolar y Densidad poblacional se tiene significatividad 90% de confianza. Elaboración propia basado en el DANE y el DNP.

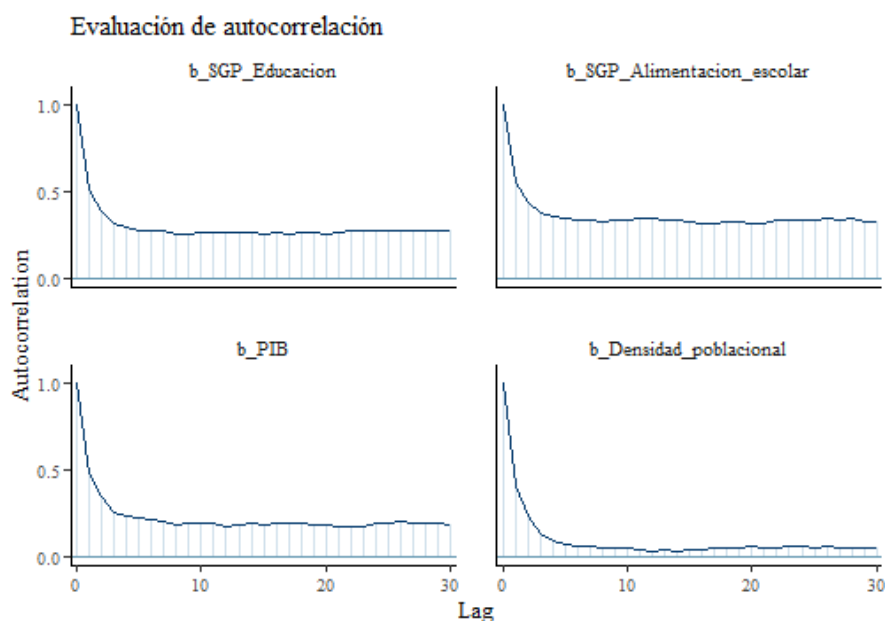
- Coeficientes e intervalos de confianza para el modelo combinado de Rezago Escolar evaluado por departamento

Variable	Estimación	Q 2.5	Q 97.5	Q 5	Q 95
Amazonas	0,12556	-0,00899	0,26010	0,01264	0,23847
Antioquia	0,01432	-0,12615	0,15479	-0,10356	0,13221
Arauca	-0,02533	-0,18512	0,13446	-0,15943	0,10877
Atlántico	0,15360	0,00635	0,30085	0,03002	0,27718
Bogotá	0,09436	-0,07249	0,26122	-0,04567	0,23439
Bolívar	0,06761	-0,08667	0,22189	-0,06186	0,19709
Boyacá	0,00412	-0,15642	0,16466	-0,13061	0,13885
Caldas	-0,05151	-0,21979	0,11677	-0,19274	0,08972
Caquetá	0,04304	-0,10577	0,19185	-0,08185	0,16793
Casanare	0,09945	-0,07441	0,27331	-0,04646	0,24536
Cauca	-0,07477	-0,22890	0,07936	-0,20412	0,05458
Cesar	0,08587	-0,04126	0,21300	-0,02082	0,19256
Chocó	0,01954	-0,11307	0,15214	-0,09175	0,13082
Córdoba	-0,05301	-0,18476	0,07874	-0,16357	0,05756
Cundinamarca	0,05418	-0,07599	0,18434	-0,05506	0,16342
Guainía	0,09056	-0,06489	0,24600	-0,03990	0,22101
Guaviare	-0,10679	-0,25877	0,04520	-0,23433	0,02076
Huila	0,02741	-0,09882	0,15364	-0,07853	0,13334
La Guajira	0,12116	-0,01887	0,26119	0,00364	0,23868
Magdalena	0,09307	-0,05146	0,23760	-0,02823	0,21436
Meta	0,03757	-0,11848	0,19363	-0,09339	0,16854
Nariño	-0,11226	-0,24647	0,02195	-0,22489	0,00037
Norte de Santander	-0,00640	-0,14441	0,13161	-0,12222	0,10943
Putumayo	-0,11519	-0,29689	0,06652	-0,26768	0,03730
Quindío	-0,19083	-0,34989	-0,03178	-0,32431	-0,05735
Risaralda	-0,12629	-0,27756	0,02498	-0,25324	0,00066
San Andrés	-0,14149	-0,32097	0,03799	-0,29211	0,00913
Santander	0,00933	-0,12865	0,14730	-0,10647	0,12512
Sucre	0,02401	-0,10283	0,15086	-0,08244	0,13047
Tolima	-0,12112	-0,25847	0,01622	-0,23639	-0,00586
Valle del Cauca	0,00986	-0,18678	0,20649	-0,15516	0,17488
Vaupés	0,07560	-0,09086	0,24205	-0,06410	0,21529
Vichada	-0,12042	-0,29257	0,05172	-0,26489	0,02404

Nota: El modelo muestra significatividad al 95% de confianza en los departamentos de Atlántico y Quindío, al 90% de confianza La guajira y Tolima también son significativos. Elaboración propia basado en el DANE y el DNP.

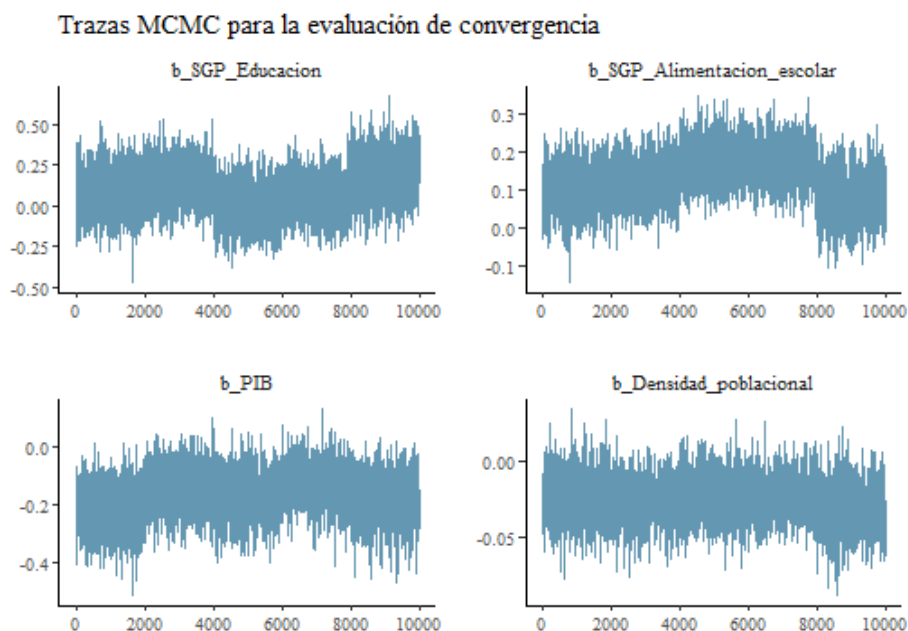
Rezago Escolar pruebas del modelo

- Gráfico de autocorrelación para las covariables del modelo de Rezago Escolar.



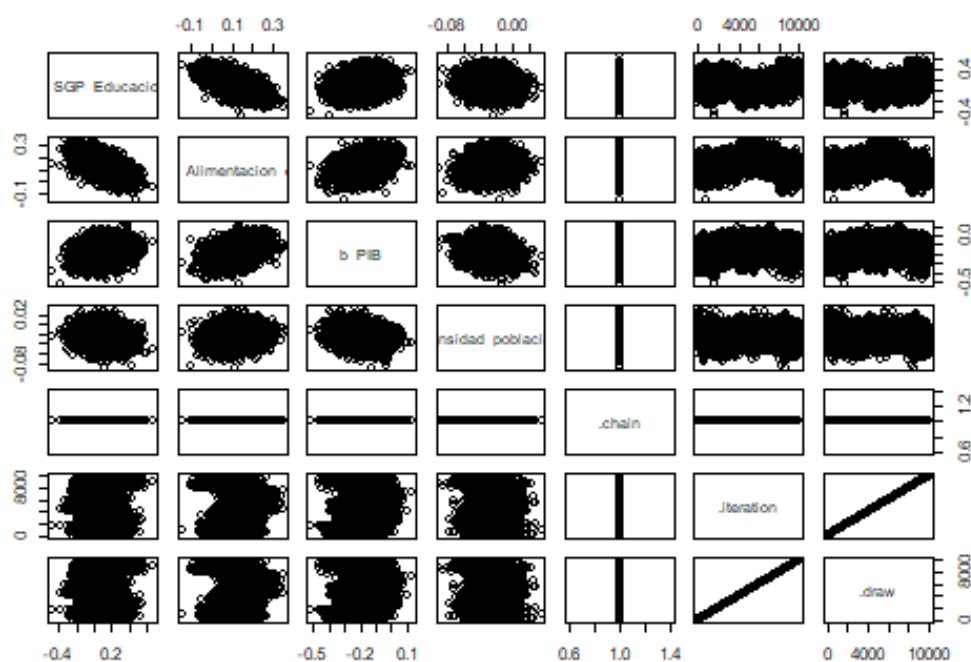
Nota: El modelo muestra convergencia hacia a distribución posterior de los parámetros del modelo. Elaboración propia basado en el DANE y el DNP.

- Gráfico de evaluación de convergencia para las covariables del modelo de Rezago Escolar.



Nota: Las covariables del modelo no evidencian una tendencia marcada, cada 2000 iteraciones representan los resultados para un modelo en una imputación. Elaboración propia basado en el DANE y el DNP.

- Prueba de Correlación y convergencia. Gráfico de pares para evaluar la relación entre las covariables del modelo de Rezago Escolar.



Nota: Las covariables del modelo no evidencian tendencias marcadas entre sus respectivas relaciones, por lo tanto, son independientes entre sí, el comportamiento de los valores a través de las iteraciones y las dos cadenas de Markov tampoco muestra tendencias marcadas, por lo cual, se logró una correcta convergencia. Elaboración propia basado en el DANE y el DNP.