Paradigma | Volumen 29 (2022) | Número 48

Paradigma
Revista de Investigación Educativa

Análisis Multinivel de Factores que Afectan el Rendimiento Escolar en Español Tercer Grado en Honduras

Multilevel Analysis of Factors Affecting School Performance in Third Grade Spanish in Honduras

Lesky Ibeth Rivas Martínez[a], **Cristian Andrés Cruz Torres**[b,*]

[a] rivaslesky@gmail.com. Instituto Gubernamental Mixto Hibueras, Honduras. https://orcid.org/0000-0003-1981-5237

[b] cristian.cruz@unah.edu.hn. Universidad Nacional Autónoma de Honduras. https://orcid.org/0000-0002-2185-5783

Resumen

En este trabajo, se hace un análisis de factores que afectan el rendimiento escolar en español tercer grado en Honduras, utilizando los datos del Tercer Estudio Regional Comparativo y Explicativo (TERCE, 2013), por medio de un modelo de regresión multinivel que permite realizar análisis con estimaciones por niveles; bajo los enfoques frecuentista y bayesiano. En el enfoque frecuentista se realizan varios modelos hasta obtener el modelo final, y se elaboran las pruebas para la verificación de los supuestos de homocedasticidad, ortogonalidad y normalidad, seleccionando el mejor modelo utilizando AIC, BIC, y Log-Verosimilitud. En el enfoque Bayesiano se comparan los modelos utilizando PSIS-LOO y WAIC. Dentro de los principales hallazgos resaltamos, que las variables de mayor impacto son: clima en el aula escolar y prácticas de recreación, las variables instalaciones y servicios básicos de la escuela, indican que unas mejores características de la escuela tienen un efecto positivo y significativo sobre el rendimiento en español de los estudiantes.

Palabras clave: rendimiento escolar, modelo multinivel, enfoque frecuentista, enfoque bayesiano

* **Autor para correspondencia**
http://doi.org/10.5377/paradigma.v29i48.15278
Recibido 18 de septiembre de 2022 | Aceptado 14 de noviembre de 2022
Disponible en línea Diciembre de 2022
Paradigma: Revista de Investigación Educativa | ISSN 1817-4221 | EISSN 2664-5033 |

Abstract

In this paper, is made an analysis of school factors that affect performance in spanish third grade in Honduras, using data from Third Comparative and Explanatory Regional Study TERCE 2013, through a model of multilevel regression that allows performing analysis with estimates by levels; by the frequentist and Bayesian approaches. In the frequentist approach several models are made until the final model is obtained, and tests are made to verify the assumptions of homoscedasticity, orthogonality and normality, selecting the best model using AIC, BIC, and Log-Likelihood criterion. In the Bayesian approach, the models are compared using PSIS-LOO and WAIC criterion. Among the main findings, the variables highlighted with the greatest impact are: climate in the school classroom and recreation practices, the variables facilities and basic services of the school, indicate that better school characteristics have a positive and significant effect on student performance in Spanish.

Keywords: school performance, multilevel model, frequentist approach, bayesian approach

Introducción

El rendimiento escolar se define como “los resultados obtenidos en pruebas o exámenes normalizados, que miden los conocimientos o competencias en una materia específica” (UNESCO, 2007, p. 430). Algunos países efectúan periódica y sistemáticamente mediciones que permiten determinar lo que los estudiantes han aprendido durante su experiencia escolar, ya sea en exámenes por clase, evaluaciones nacionales o evaluaciones internacionales.

El rendimiento escolar “es medido, en su mayor parte por tests especialmente diseñados, pero en algunos casos, el rendimiento escolar es medido a través de exámenes nacionales en lengua, matemática o ciencia” (Velez et al., 1994, p. 4). Dichas evaluaciones y sus resultados orientan la toma de decisiones en el sistema educativo.

Las agencias internacionales como: Tendencias en el Estudio Internacional de Matemáticas y Ciencias (TIMSS), Progreso en el Estudio Internacional de Alfabetización en Lectura (PIRLS), Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), y las nacionales como: Unidad Externa de Medición de la Calidad de la Educación, por sus siglas UMCE; realizan evaluaciones que monitorean las tendencias en el rendimiento estudiantil en matemáticas, ciencias y lectura. Gracias a una reciente serie de estudios nacionales e internacionales, se han adquirido nuevos conocimientos importantes sobre lo que saben y pueden hacer los estudiantes de las escuelas primarias de los países en desarrollo (UNESCO, 2000).

El uso de modelos en la investigación sobre eficacia escolar constituye una interesante propuesta realizada por Stringfield et al. (1994). Este expresa tres razones que considera de interés para elaborar modelos. En primer lugar, un modelo ofrece a los investigadores del área una fuente de información que permite considerar, relacionar entre sí, y tender puentes sobre la investigación previa. Posteriormente, permite abrir una serie de caminos, o alternativas, para desarrollar investigaciones futuras que, a posteriori, guiarán en la toma de las decisiones políticas relacionadas con la eficacia escolar. Por último, los modelos proporcionan una visión panorámica de las diferentes rutas de investigación posibles (Bustamante, 2015). Es por esto que las agencias dedicadas a realizar estudios sobre rendimiento escolar utilizan diferentes modelos en el análisis de los factores asociados al rendimiento escolar, aportando herramientas para la formulación de políticas públicas pertinentes en cada país.

El objetivo de este trabajo es identificar los factores que afectan el rendimiento escolar en español tercer grado en Honduras, por medio de un modelo de regresión multinivel, para contribuir a la mejora educativa, considerando que conocer los factores asociados al rendimiento ayuda la revisión de prácticas educativas tanto en la escuela como en el hogar.

Discusión Teórica

Aitkin y Longford (1986) propusieron una técnica de análisis que ha marcado la investigación educativa desde entonces: los modelos multinivel (o modelos lineales-jerárquicos). Éstos reconocen y manejan la organización jerárquica de los sistemas educativos (estudiantes en aula de clases, aulas en escuelas, escuelas en países) y ofrecen resultados con una menor incidencia de los errores de estimación.

Modelos Lineales Multinivel bajo el Enfoque Frecuentista

El modelado multinivel puede verse útil al realizar análisis de regresión en circunstancias específicas. Las circunstancias son aquellas en las que las observaciones, como los estudiantes, están anidadas o agrupadas en contextos identificables, como aulas, escuelas y distritos (Bickel, 2007). Los modelos multinivel son, en esencia, ampliaciones de los modelos de regresión lineal clásicos; ampliaciones mediante las cuales se elaboran varios modelos de regresión para cada nivel de análisis, con ello los modelos del primer nivel están relacionados por un modelo de segundo nivel en el que los coeficientes de regresión del nivel uno se regresa en un segundo nivel de variables explicativas, y así sucesivamente para los diferentes niveles (Torrecilla, 2008).

En un modelo multinivel hay dos tipos de parámetros: los parámetros fijos y los aleatorios. Los primeros corresponden a los efectos medios en la población. Los aleatorios corresponden a las varianzas y covarianzas de todos los niveles (Jurado, 2013).

1. Variables fijas y aleatorias: una variable aleatoria es una variable que toma sus valores de una distribución de probabilidad, por lo tanto, tiene una media y una varianza (que puede ser o no conocida). Una variable fija es aquella que sus valores son conocidos.
2. Coeficientes fijos y aleatorios: en los modelos de regresión lineal clásicos, se estiman los parámetros que especifican la recta de regresión, que son el intercepto y las pendientes; estos parámetros son los

mismos para todos los grupos, en otras palabras, estos coeficientes son fijos. Los coeficientes aleatorios son coeficientes que se distribuyen según una función de probabilidad. Una regla general para los coeficientes de regresión aleatorios es que solo pueden ser considerados aleatorios en el nivel superior en el que han sido medidos.

3. Efectos fijos y aleatorios: se habla de efectos fijos cuando se cuenta con una variable cuyo dominio representa todos los niveles posibles que son de interés teórico, por ejemplo, sexo, raza/etnia. Y de efectos aleatorios, cuando la variable representa una muestra de una población más amplia de valores potenciales: individuos, aulas, escuelas.

El modelo de regresión multinivel asume que existe un conjunto de datos jerárquicos, que a menudo consta de sujetos anidados dentro de grupos, con una única variable de resultado o respuesta que se mide en el nivel más bajo, y variables explicativas en todos los niveles existentes (Hox et al., 2018).

Modelo General dos Niveles

Tomando el siguiente modelo básico de regresión lineal simple:

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (1)$$

Aquí, la variable dependiente y se expresa como una función de una variable independiente x, multiplicada por un coeficiente de pendiente β_1, un intercepto β_0 y una variación aleatoria ε conocido como el error y se distribuye normalmente con una varianza constante σ^2. Se define el intercepto como la media condicional de y cuando el valor de x es 0.

En el contexto de un modelo de regresión de un solo nivel como este, el intercepto es común a todos los individuos de la población de interés. Sin embargo, cuando los individuos se agrupan de alguna manera, por ejemplo, estudiantes en aulas y escuelas, potencialmente habrá un intercepto separado para cada grupo, es decir, pueden existir diferentes medias para la variable dependiente para $x = 0$ en los diferentes grupos. Permitir interceptos y pendientes específicas de grupo conduce a la siguiente notación que se usa comúnmente para el modelo de nivel 1 o micro, en el modelado multinivel (Finch et al., 2014).

$$y_{ij} = \beta_{0j} + \beta_{1j}\, x + \varepsilon_{ij} \qquad (2)$$

Donde el subíndice ij se refiere al i-ésimo individuo del j-ésimo grupo, y_{ij} es la variable respuesta del i-ésimo individuo del j-ésimo grupo, β_{0j} es el intercepto de y del j-ésimo grupo, β_{1j} la pendiente promedio de la regresión y ε_{ij} es el error y se distribuye normalmente con una varianza constante.

Modelo multinivel más básico: Sirve para predecir el resultado a partir de interceptos que varían aleatoriamente para cada grupo:

$$y_{ij} = \beta_{0j} + \varepsilon_{ij} \qquad (3)$$

Al permitir que los interceptos difieran entre los grupos, como en la ecuación (3), conduce a los interceptos aleatorios expresarse como:

$$\beta_{0j} = \gamma_{00} + U_{0j} \qquad (4)$$

Donde γ_{00} representa un valor general o promedio que se mantiene en todos los grupos, mientras que U_{0j} es un efecto específico de grupo en el intercepto, con media cero y una varianza τ^2_{00}. γ_{00} es un efecto fijo porque permanece constante en todos los grupos, y U_{0j} es un efecto aleatorio porque varía de un grupo a otro. Por lo tanto, para un modelo lineal multinivel se está interesado no solo en algún valor medio general para y cuando x es 0 para todos los individuos de la población γ_{00}, sino también en la desviación entre la media general y los efectos específicos del grupo para U_{0j}.

Si sustituimos los dos componentes de los interceptos aleatorios en el modelo de regresión, obtenemos:

$$y = \gamma_{00} + U_{0j} + \beta_1 x + \varepsilon \qquad (5)$$

La ecuación (5) se denomina modelo completo o compuesto en el que los niveles múltiples se combinan en una ecuación unificada. A menudo, en el modelo lineal multinivel, se comienza el análisis de un conjunto de datos con este modelo de intercepción aleatoria simple conocido como el modelo nulo que toma la forma:

$$y_{ij} = \gamma_{00} + U_{0j} + \varepsilon_{ij} \qquad (6)$$

Si bien este modelo nulo no proporciona información sobre los impactos de variables independientes específicas sobre la dependiente, sí proporciona información importante sobre cómo la variación en y se divide entre la varianza, entre los valores individuales de σ^2 y la varianza entre los niveles τ^2. La varianza total de y es simplemente la suma de σ^2 y τ^2.

Se puede expandir el modelo de interceptos aleatorios en la ecuación (4) para acomodar una o más variables predictoras independientes. Como ejemplo, si agregamos un solo predictor x_{ij} a nivel individual (nivel 1) al modelo, se obtiene:

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + U_{0j} + \varepsilon_{ij} \qquad (7)$$

Este modelo también se puede expresar en dos niveles separados:

Nivel 1:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij} \qquad (8)$$

Nivel 2:

$$\beta_{0j} = \gamma_{00} + U_{0j} + \varepsilon_{ij} \qquad (9)$$

$$\beta_{1j} = \gamma_{10} \qquad (10)$$

El modelo ahora incluye el predictor y la pendiente que lo relaciona con la variable dependiente γ_{10}, que es el nivel *1* por el subíndice 10. La interpretación para γ_{10} es de la misma manera que β_1 en el modelo de regresión lineal, es decir, como una medida del impacto en y de un cambio de una unidad en x.

Correlación Intraclase (ICC)

Es un indicador de la homogeneidad interna de los grupos. Es una medida de la similitud de las unidades del nivel individual y de las diferencias entre las unidades del nivel macro (Fernández, 2012). Una correlación baja o cercana a cero significará que los sujetos dentro del mismo grupo son tan diferentes entre sí como los que pertenecen a otros grupos.

En un modelo de dos niveles, la correlación intraclase se calcula de la siguiente manera:

$$\rho = \frac{\sigma^2}{\tau^2 + \sigma^2} \qquad (11)$$

Donde σ^2 es la varianza dentro de los niveles y τ^2 es la varianza entre las medias de los distintos niveles.

La correlación intraclase ρ indica la proporción de la varianza total explicada por la estructura de agrupamiento en la población. La ecuación (11) establece que la correlación intraclase es la proporción de la varianza a nivel de grupo en comparación con la varianza total. La correlación ρ también se puede interpretar como la correlación esperada entre dos unidades que están en el mismo grupo (Hox et al., 2018).

Supuestos del Modelo Multinivel

Los modelos multinivel tienen algunos supuestos de partida, sin cuyo cumplimiento las estimaciones obtenidas no son correctas. Los principales supuestos recaen sobre el error del modelo, y su certificación se realiza a través del análisis de los residuos. Estos supuestos son los siguientes (Torrecilla, 2008):

1. Homocedasticidad: el error tiene media nula y varianza constante.
2. Ortogonalidad: los componentes aleatorios y el valor previsto son ortogonales, es decir que si por ejemplo se llaman a dos variables aleatorias X y Y entonces $E[X^T Y] = 0$, donde X^T representa la transpuesta de X, y $E[.]$ representa el valor esperado.
3. Normalidad: el error debe tener una distribución normal para que se puedan inferir los resultados de la muestra a la población.

Para verificar si tales supuestos se cumplen o no, se realizarán pruebas gráficas partiendo de los residuos, y como prueba adicional de normalidad se aplicará la prueba de Shapiro-Wilk, con un nivel α de significancia del 5%.

Calidad del Modelo

Es posible estimar la proporción de varianza en la variable de resultado contabilizada en cada nivel del modelo. En el contexto del modelado multinivel, los valores de R^2 se pueden estimar para cada nivel del modelo (Finch et al., 2014). Lo que se desea conocer es cuánta varianza de la escuela y del estudiante es explicada por el modelo, sería un valor de su capacidad explicativa, R^2 toma valores entre *0* y *1*.

Para el nivel 1, se puede calcular de la siguiente manera:

$$R_1^2 = 1 - \frac{\sigma_{MF}^2}{\sigma_{M0}^2} \qquad (12)$$

Para el nivel 2:

$$R_2^2 = 1 - \frac{\tau_{MF}^2}{\tau_{M0}^2} \qquad (13)$$

Si el intercepto apenas tiene varianza aleatoria, la varianza total será la suma de las varianzas de los niveles 1 y 2. De esta forma, podremos estimar el coeficiente de determinación total R^2:

$$R^2 = 1 - \frac{\sigma_{MF}^2 + \tau_{MF}^2}{\sigma_{M0}^2 + \tau_{M0}^2} \qquad (14)$$

Donde σ_{MF}^2 es la varianza entre las medias de los distintos niveles del modelo final, $\sigma_{MF}^2 + \tau_{MF}^2$ es la varianza dentro de los niveles del modelo final, σ_{M0}^2 es la varianza entre las medias de los distintos niveles del modelo nulo, τ_{M0}^2 es la varianza dentro de los niveles del modelo nulo.

Inferencia Bayesiana

El enfoque bayesiano para el análisis de datos se diferencia del frecuentista en que cada parámetro del modelo se considera una variable aleatoria, contrariamente al enfoque frecuentista que considera los valores de los parámetros como cantidades desconocidas y fijas, y por el uso explícito de la probabilidad para modelar la incertidumbre. Los dos enfoques también difieren en su concepción de lo que es la probabilidad. En el marco bayesiano, la probabilidad se refiere a la experiencia de la incertidumbre, mientras que en el marco frecuentista se refiere al límite de una frecuencia relativa, es decir, la frecuencia relativa de un evento cuando el número de ensayos se acerca al infinito. Una consecuencia directa de estas dos diferencias es que el análisis de datos bayesianos permite a los investigadores discutir la probabilidad de un parámetro o un vector de parámetros θ, dado un conjunto de datos y usando el teorema de Bayes, se puede derivar una distribución de probabilidad $p(\theta|y)$ llamada distribución posterior, que refleja el conocimiento sobre el parámetro, dados los datos y la información previa. Esta distribución es el objetivo de cualquier análisis bayesiano y contiene toda la información necesaria para la inferencia (Nalborczyk et al., 2019).

El teorema de Bayes está dado por:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} \qquad (15)$$

El término $p(\theta)$ corresponde a la distribución previa, que especifica la información previa sobre los parámetros, es decir, lo que se conoce sobre θ antes de observar los datos como una distribución de probabilidad. El lado derecho del numerador $p(y|\theta)$ representa la verosimilitud también llamada distribución muestral o modelo generativo, y es la función a través de la cual los datos afectan la distribución posterior, y finalmente, $p(y)$ se llama probabilidad marginal.

El resultado de un análisis bayesiano, es decir, la distribución posterior $p(\theta|y)$, está dado por el producto de la información contenida en los datos, es decir, la verosimilitud y la información disponible antes de observar los datos, es decir, la previa. Este constituye el principio crucial de la inferencia bayesiana, que puede verse como un mecanismo de actualización. En resumen, el teorema de Bayes permite actualizar un estado de conocimiento previo a un estado de conocimiento posterior, lo que representa un compromiso entre el conocimiento previo y la evidencia empírica.

Modelos Bayesianos Multinivel

Se construirá un modelo lineal multinivel a partir del modelo de regresión lineal ordinario y se intentará predecir un resultado y_i mediante una combinación lineal de un intercepto y una pendiente que cuantifica la influencia de un predictor x_i (Nalborczyk et al., 2019):

$$y_i \sim N(\mu_i, \sigma_e^2) \qquad (16)$$

$$\mu_i = \alpha + \beta x_i \qquad (17)$$

En términos bayesianos, estas dos líneas (16) y (17) describen la verosimilitud del modelo, que es la suposición hecha sobre el proceso generativo a partir del cual se emiten los datos. Suponemos que los resultados y_i se distribuyen normalmente alrededor de una media μ_i con algún error σ_e^2. Esto equivale a decir que los errores se distribuyen normalmente alrededor de 0, como lo ilustra la equivalencia anterior. Luego, podemos extender este modelo al siguiente modelo multinivel, agregando un intercepto variable:

$$y_i \sim N(\mu_i, \sigma_e^2) \qquad (18)$$

$$\mu_i = \alpha_{j[i]} + \beta x_i \qquad (19)$$

$$\alpha_{j[i]} \sim N(\alpha, \sigma_\alpha^2) \qquad (20)$$

Donde $\alpha_{j[i]}$ indica que a cada grupo j se le da un intercepto único, emitido a partir de una distribución gaussiana centrada en α, el gran intercepto, lo que significa que puede haber diferentes puntuaciones medias para cada clase. A partir de esta notación se puede ver que además de la varianza residual σ_e^2, ahora se esta estimando un componente más σ_α^2 que es la varianza de la distribución de intercepciones variables. Se puede interpretar la variación del parámetro α entre los grupos j considerando la correlación intraclase (ICC), que va a *0*, si la agrupación no transmite información, y a *1*, si todas las observaciones en un grupo son idénticas.

La tercera línea (20) se denomina distribución previa en el marco bayesiano. Esta distribución previa describe la población de interceptos, modelando así la dependencia entre estos parámetros. Siguiendo la misma estrategia, se puede agregar una pendiente variable, que puede variar según el grupo j:

$$y_i \sim N(\mu_i, \sigma_e^2) \quad (21)$$

$$\mu_i = \alpha_{j[i]} + \beta_{j[i]} x_i \quad (22)$$

$$\alpha_{j[i]} \sim N(\alpha, \sigma_\alpha^2) \quad (23)$$

$$\beta_{j[i]} \sim N(\beta, \sigma_\beta^2) \quad (24)$$

A estas pendientes variables se les asigna una distribución previa centrada en la gran pendiente β, y con varianza σ_β^2.

En el marco bayesiano, toda incógnita se considera una variable aleatoria que se puede describir mediante distribuciones de probabilidad. Como consecuencia, no existe tal cosa como un “efecto fijo” o una “distribución de efectos aleatorios”. Sin embargo, estas disputas semánticas desaparecen cuando se escribe el modelo (Nalborczyk et al., 2019).

Supongamos que tenemos una variable continua dependiente y y un predictor categórico dicotómico x. Sea y_{ij} la puntuación del i-ésimo participante en la j-ésima condición. Se puede escribir un modelo de “efectos mixtos” que contiene tanto efectos fijos como aleatorios de la siguiente manera:

$$y_{ij} = \alpha + \alpha_i + \beta x_i + \varepsilon_{ij} \text{ con } \varepsilon_{ij} \sim N(0,\sigma_e^2) \text{ y } \alpha_i \sim N(0,\sigma_\alpha^2) \quad (25)$$

Donde los términos α y β representan los “efectos fijos” y denotan la respuesta media general y la diferencia de condición en la respuesta, respectivamente. Además, ε_{ij} son errores aleatorios que se supone que se distribuyen normalmente con varianza desconocida σ_e^2, y los α_i son efectos aleatorios específicos individuales normalmente distribuidos en la población con varianza desconocida σ_α^2.

Se puede reescribir el modelo anterior de la siguiente manera:

$$y_{ij} \sim N(\mu_{ij}, \sigma_e^2) \quad (26)$$

$$\mu_{ij} = \alpha_i + \beta x_i \quad (27)$$

$$\alpha_i \sim N(\alpha, \sigma_\alpha^2) \quad (28)$$

Comparación de los Modelos Bayesianos Multinivel

Hay dos familias de estrategias para la comparación de los modelos: validación cruzada y criterios de información. Estas estrategias tratan de visualizar cómo funcionarán bien los modelos, en promedio, en la predicción de nuevos datos.

Validación Cruzada

Una estrategia popular para estimar la precisión predictiva es probar la precisión predictiva del modelo en otra muestra. Esto se conoce como validación cruzada, dejando fuera una pequeña parte de las observaciones de la muestra y devaluando el modelo en las observaciones que quedaron fuera. Por supuesto, no se quiere omitir datos. Entonces, lo que se suele hacer es dividir la muestra en varios trozos, llamados "pliegues". Se le pide al modelo que prediga un grupo, después del entrenamiento en todos los demás pliegues para obtener una estimación de la precisión fuera de la muestra. El número mínimo de pliegues es *2* (McElreath, 2020). Una de las más utilizadas es validación cruzada suavizada por Pareto (PSIS-LOO).

Criterios de Información

El segundo enfoque es el uso de criterios de información para calcular una puntuación esperada de la muestra, estimación de la divergencia KL (Kullback-Leibler) relativa fuera de la muestra (Seghouane y Amari, 2007). Entre los criterios de información se utiliza: Criterio de información de Akaike (AIC), el criterio de información ampliamente aplicable (WAIC).

Criterios de Calidad del Modelo Bayesiano

Los modelos multinivel se basan en relaciones de regresión en diferentes niveles, con el primer nivel correspondiente a los datos individuales y los niveles subsiguientes correspondientes a regresiones entre grupos de los efectos de los predictores individuales sobre las variables de la unidad de agrupación.

Se utiliza un enfoque bayesiano para definir R^2 en cada nivel del modelo multinivel, en lugar de intentar crear una sola medida de resumen de ajuste, este método se basa en comparar varianzas en un solo modelo ajustado en lugar de un modelo nulo. También se realiza una comparación de varianza relacionada para resumir el grado en que las estimaciones en cada nivel del modelo se agrupan en función de la relación de regresión específica del nivel, en lugar de estimarse por separado. Este factor de agrupación está relacionado con el concepto de contracción en modelos jerárquicos simples (Gelman y Pardoe, 2012).

El modelo jerárquico descrito permite determinar las comparaciones de varianza apropiadas en cada uno de los dos niveles del modelo, esto puede considerarse un modelo multinivel con solo el predictor constante a nivel de grupo. Para simplificar, se supone que los tamaños de muestra dentro del grupo n_j son todos iguales a un valor común n, de modo que el tamaño de muestra total es $N = n_J$.

***Varianza Explicada** R^2*

Para el modelo de nivel de datos condicionada a σ_e^2 y σ_α^2 la proporción de la varianza explicada es (Gelman y Pardoe, 2012):

$$R^2 = 1 - \frac{\frac{\hat{\sigma}_\alpha^2}{n}}{\frac{\hat{\sigma}_\alpha^2}{n+1} + \frac{\hat{\sigma}_e^2}{n}} \qquad (29)$$

Factor de Agrupación λ

El factor de agrupación λ permite resumir el grado en que las estimaciones en cada nivel del modelo se agrupan en función de la relación de regresión específica del nivel.

Para el modelo de nivel de datos condicionada a σ_e^2 y σ_α^2, el factor de agrupación a nivel de datos es (Gelman y Pardoe, 2012):

$$\lambda = 1 - \frac{\frac{n}{(n+1)}\hat{\sigma}_\alpha^2 + \frac{\hat{\sigma}_e^2}{n}}{\hat{\sigma}_\alpha^2 + \frac{\hat{\sigma}_e^2}{n}} \qquad (30)$$

El factor de agrupación a nivel de grupo es:

$$\lambda = 1 - \frac{\hat{\sigma}_e^2}{\hat{\sigma}_\alpha^2 + \frac{\hat{\sigma}_e^2}{n}} \qquad (31)$$

Al interpretar el factor de agrupación λ, 0,5 es el punto de referencia. Un factor de agrupación $\lambda < 0{,}5$ sugiere un mayor grado de información dentro del grupo que la información a nivel de población. Por el contrario, un factor de agrupación $\lambda > 0{,}5$ sugiere un mayor grado de información a nivel de población que información dentro del grupo (Gelman y Pardoe, 2012).

Métodos y Materiales

Con el propósito de analizar los factores que afectan el rendimiento escolar en español tercer grado en Honduras por medio de un modelo de regresión multinivel, esta investigación se desarrolló desde un método cuantitativo de investigación.

Entorno

En el año 2013 se publicó por la UNESCO los resultados del Tercer Estudio Regional Comparativo y Explicativo (TERCE), en cooperación con las coordinaciones nacionales de los países participantes (UNESCO, 2016). El propósito principal de este estudio fue evaluar la calidad de la educación en términos de logro de aprendizaje en los países participantes de América Latina y el Caribe, e identificar factores asociados a este logro, tales como: antecedentes escolares, prácticas educativas del hogar, características socioeconómicas, demográficas y culturales, formación docente, asistencia y puntualidad docente, recursos del aula, prácticas del aula, desigualdad en los resultados académicos entre escuelas y al interior de estas, población que atiende las escuelas, tipo de escuela y entorno social, recursos escolares, procesos en las escuelas.

Población y Muestra

Debido a los esfuerzos realizados por la UNESCO y el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), la Secretaría de Educación de Honduras ratificó un acuerdo en el año de 1999 que permite a LLECE la elaboración de informes de lenguaje, matemática y factores asociados a la calidad de la educación (Secretaría de Educación, 1999), esta es una de las razones por las cuales Honduras participó en el TERCE. Al no tener acceso público a las bases de datos de estudios realizados en Honduras sobre rendimiento escolar, se tomó la decisión de trabajar con las bases de datos públicas del TERCE (UNESCO, 2016), seleccionado únicamente los datos correspondientes a Honduras para realizar este estudio.

La base de datos de español tercer grado consta de 167 variables, contiene preguntas relativas a aspectos educativos del hogar, procesos y estrategias de aprendizaje dentro del aula de clases, actividades recreativas y disponibilidad de materiales (UNESCO, 2016).

Los Experimentos

Para este estudio se utilizó los datos recopilados en el TERCE, y se realizó un análisis de componentes principales para reducir la dimensionalidad de los datos.

El análisis de componentes principales (ACP) es particularmente útil para reducir la dimensionalidad de un grupo de datos. Un ACP se ocupa de explicar la estructura de varianza-covarianza de un conjunto de variables a través de algunas combinaciones lineales de estas. A menudo revela relaciones que no se sospechaban anteriormente y, por lo tanto, permite interpretaciones que normalmente no resultarían (Johnson y Wichern, 2007).

Algebraicamente, los componentes principales son combinaciones lineales particulares de p variables aleatorias, $x_1, x_2, \ldots, x_p$. Geométricamente, estas combinaciones lineales representan la selección de un nuevo sistema de coordenadas obtenida mediante la rotación del sistema original, con $x_1, x_2, \ldots, x_p$ como ejes de coordenadas. Los nuevos ejes representan las direcciones con la máxima variabilidad y proporcionan una descripción más simple y parsimoniosa de la estructura de covarianza.

Se realizó el análisis de componentes principales para la base de datos de español tercer grado, estableciendo un umbral del 90% de la varianza total, lo que permitió seleccionar 33 componentes. La base de datos de español tercer grado consta de 167 variables cada una. Las cuales están contenidas en los diferentes índices, que de ahora en adelante llamaremos variables en los modelos.

Análisis Estadístico

Se seleccionaron las variables que serían analizadas en el modelo, aplicando la técnica estadística de análisis de componentes principales (ACP), el proceso consistió en realizar combinaciones con las variables seleccionadas, más la ayuda de estudios anteriores a fin de obtener un modelo explicativo. Vale mencionar que toda la información de las bases de datos se procesó con el paquete estadístico R (Team, 2020). La elaboración de los índices que en este caso representan las variables, se dividieron en, estudiantes: que contiene todos los índices relacionados con los estudiantes y sus familias. Director: que contiene todos los índices relacionados con la escuela y el director. Y docentes: contiene todos los índices relacionados con el aula y los docentes.

Para realizar el análisis bajo el enfoque frecuentista toda la información de la base de datos se procesó con el paquete estadístico R (Team, 2020) con la librería lme4 (Bates et al., 2014). Y para el análisis bajo el enfoque bayesiano utilizando siempre el paquete estadístico R, y para este caso la librería brms y rstan (Bürkner, 2017).

Resultados

Los resultados obtenidos en este trabajo indican que tan bien se ajusta la variable dependiente y las variables independientes en los modelos utilizando regresión múltiple. Permitiendo identificar los factores que afectan el rendimiento escolar en español tercer grado Honduras.

Resultados del Enfoque Frecuentista

Se trató de obtener el modelo que mejor se ajuste a los datos utilizando regresión multinivel con un enfoque frecuentista. Toda la información de la base de datos se procesó con el paquete estadístico R (Team, 2020) con la librería lme4 (Bates et al., 2014).

Se trabajó con un modelo multinivel de 2 niveles, estudiantes anidados dentro de escuelas. A continuación, se describen las variables según el nivel: Nivel 1 estudiante: Género, edad, nivel educativo de los padres, servicios con los que cuenta el hogar, bienes con los que cuenta el hogar, clima en el salón de clases, puntualidad del docente, actitud del docente en el salón de clases, preparación docente en la materia, metodología docente, liderazgo en el salón de clases, hostilidad en el salón de clases, razón por la que lee el niño, prácticas de recreación, trabajo infantil, uso del computador, acceso a internet, uso del computador en el tiempo libre, nivel educativo de los padres según los padres, ingresos en el hogar, alimentación en el hogar, servicios con los que cuenta la comunidad, clima negativo del barrio o comunidad, clima positivo del barrio o comunidad, gusto por la lectura, asistencia al centro educativo, supervisión de estudios en el hogar, comunicación con la escuela, técnicas de evaluación, clima del aula según el docente.

Nivel 2 escuela: Instalaciones en la escuela, infraestructura del aula, servicios básicos de la escuela, programas escolares adicionales, gestión directiva interna, satisfacción laboral según el director, evaluación desempeño docente, preparación docente, actualización docente, ambiente laboral según el docente, acompañamiento docente, gestión directiva interna según el docente, monitoreo de las prácticas docentes.

Modelo Nulo

El modelo nulo contiene únicamente la variable respuesta y el intercepto, sin ninguna variable predictora. Este modelo es útil para obtener estimaciones de la varianza residual y de intercepción cuando solo se considera el agrupamiento por escuela (Torrecilla, 2008). Los valores AIC y BIC, Log-Verosimilitud que son de interés primordial, en este caso serán útiles para comparar este modelo con otros que incluyen una o más variables independientes, y seleccionar el modelo que mejor se ajuste, como se verá más adelante.

Para el caso del modelo nulo con dos niveles para la base de datos tercer grado español con 3,647 estudiantes y 196 escuelas, obtuvo como resultado un AIC de -5463,5, BIC de -5444,9 y Log-Verosimilitud de 2734,7 con varianza τ^2_{M0} = 0,006413 y σ^2_{M0} = 0,011568, con una correlación intraclase ICC de 0,35 lo que indica que la correlación de los puntajes de las pruebas de español entre los estudiantes dentro de las mismas escuelas es 35%.

Modelo solo Variables de Estudiantes

La expansión del modelo se realiza incluyendo variables relacionadas con los estudiantes al modelo nulo y verificar la significancia de cada nueva variable comparando los valores del estadístico del modelo nulo y los modelos siguientes que contienen variables relacionadas con los estudiantes, lo que permite evaluar el aporte de cada variable; de igual manera la varianza de cada nivel permite analizar cómo la inclusión de cada variable afecta los diferentes niveles (Jurado, 2013).

De las 20 variables incluidas en el modelo tercer grado español, se obtuvieron los siguientes resultados: un AIC de -5597,5, BIC de -5454,9 y Log-Verosimilitud de 2821,8, estos resultados indican una mejoría en el modelo, y se obtuvieron varianzas τ^2_{M1} = 0,004761 y σ^2_{M1} = 0,011163.

De las 20 variables incluidas en este modelo solo 9 resultaron significativas, se realiza un nuevo análisis y se obtuvieron los resultados de un AIC de -5612,6, BIC de -5538,2 y Log-Verosimilitud de 2818,3, estos resultados indican que estas variables proporcionan un mejor ajuste al modelo, esto significa que se deben incluir en el modelo final, y la estimación de la variación en las intersecciones entre escuelas es τ^2_{M2} = 0,004858, mientras que la variación dentro de la escuela se estima en σ^2_{M2} = 0,011175.

Modelo solo Variables de Escuela

Se continúa con el modelo para mostrar el efecto de la adición de cada una de las variables del nivel escuela al modelo nulo y verificar la significancia de cada nueva variable comparando los valores del estadístico del modelo nulo, se incorporaron las variables correspondientes al nivel escuela.

De las 15 variables incluidas en el modelo de español tercer grado se obtuvieron los siguientes resultados: un AIC de -5571,7, BIC de -5460,0 y Log-Verosimilitud de 2803,8, estos resultados indican una mejoría en el modelo, y se obtuvieron varianzas τ^2_{M1} = 0,002461 y σ^2_{M1} = 0,011616.

De las 15 variables incluidas en este modelo solo 4 resultaron significativas, se realiza un nuevo análisis y para este modelo se obtuvieron los resultado de un AIC de -5574,5, BIC de -5531,1 y Log-Verosimilitud de 2794,3, estos resultados indican que estas variables proporciona un mejor ajuste al modelo, esto significa que se deben incluir en el modelo final, y la estimación de la variación en las intersecciones entre escuelas es τ^2_{M2} = 0,002817, mientras que la variación dentro de la escuela se estima en σ^2_{M2} = 0,011614.

Modelo Final

Finalmente se quiere analizar el efecto que tienen de manera conjunta las variables de los dos niveles sobre el rendimiento de los estudiantes. Para el modelo español tercer grado se obtuvieron los resultados un AIC de -5711,6, BIC de -5612,3 y Log-Verosimilitud de 2871,8, estos resultados indicaron una mejoría en el modelo, y se obtuvieron varianzas τ^2_{M1} = 0,002817 y σ^2_{M1} = 0,011614, las variables significativas se presentan en la Tabla 1.

Tabla 1

Variables significativas del modelo final español tercer grado

Nombre de las variables	Estimación	Error estándar	Valor t	Pr(>|t|)
Género	-0,0078	0,0036	-2,1640	0,0305
Edad	0,0085	0,0038	2,2560	0,0241
Trabajo infantil	0,0041	0,0019	2,1780	0,0295
Clima en el aula escolar	0,0070	0,0012	5,8330	0,0000
Prácticas de recreación	0,0023	0,0004	5,4530	0,0000
Ingresos en el hogar	0,0009	0,0003	3,2540	0,0011
Supervisión de estudios en el hogar	0,0017	0,0006	2,6850	0,0073
Instalaciones en la escuela	0,0051	0,0019	2,6740	0,0083
Servicios básicos de la escuela	0,0122	0,0038	3,1690	0,0018
Disponibilidad de libros y materiales en la escuela	-0,0180	0,0046	-3,9090	0,0001
Consejo docente	0,0027	0,0010	2,6000	0,0100

Nota. Fuente: elaboración propia, en base a datos del TERCE, 2013.

De las 13 variables incluidas en este modelo final solo 11 resultaron significativas. Estas variables proporcionan el mejor ajuste a los datos, se puede observar la disminución de los AIC = -5712,1, BIC = -5625,2 en comparación con el modelo nulo.

La estimación de la variación en las intersecciones entre escuelas es 0,002368, y la variación dentro de la escuela se estima en 0,011203, con una correlación intraclase ICC de 0,18 lo que indica que la correlación de los puntajes de las pruebas de español entre los estudiantes dentro de las mismas escuelas es 18%.

Quedando el modelo final completo español tercer grado de la siguiente manera:

Nivel 1: Estudiantes

$$\begin{aligned} log(PE_{ij}) = {} & \beta_{0j} + \beta_{1j}\, Género_{ij} + \beta_{2j}\, Edad_{ij} + \beta_{3j}\, TrabInfL3_{ij} + \beta_{4j}\, ClimAulaL3_{ij} \\ & + \beta_{5j}\, PracRecreL3_{ij} + \beta_{6j}\, IngreHogarL3_{ij} + \beta_{7j}\, ISupEstHogarL3_{ij} + \varepsilon_{ij} \end{aligned}$$

Nivel 2: Escuela

$$\begin{aligned} \beta_{0j} = {} & \gamma_{00} + \gamma_{01}\, InstEscL3_{0j} + \gamma_{02} ServBasEscL3_{0j} + \gamma_{03}\, LibyMatDirL3_{0j} \\ & + \gamma_{04}\, ConsDocL3_{0j} + U_{0j} \end{aligned}$$

Selección del Modelo

Se utilizó el método paso a paso, para determinar si el modelo, mejora o no, con cada incorporación o extracción de variables. Como se usa paquete estadístico R con la librería lme4, esta librería muestra los valores AIC, BIC, logLik para cada modelo que se ejecuta, y los resultados se van comparando con el modelo nulo.

Tabla 2

Resultados del modelo español tercer grado

Modelo	AIC	BIC	logLik
Modelo nulo	-5463,5	-5444,9	2734,7
Modelo 1	-5612,6	-5538,2	2818,3
Modelo 2	-5574,5	-5531,1	2794,3
Modelo final	-5712,1	-5625,2	2870,0

Nota. Fuente: elaboración propia, en base a los resultados obtenidos por la librería lme4.

El modelo final que contiene las variables de los niveles estudiantes y escuela proporcionan el mejor ajuste a los datos, se puede observar la disminución de los AIC y BIC en comparación con el modelo nulo, modelo 1 que solo contiene las variables a nivel de estudiante y el modelo 2 que contiene solamente las variables a nivel escuela. También se observó el ligero incremento de log-verosimilitud en comparación con el modelo Nulo (ver Tabla 2).

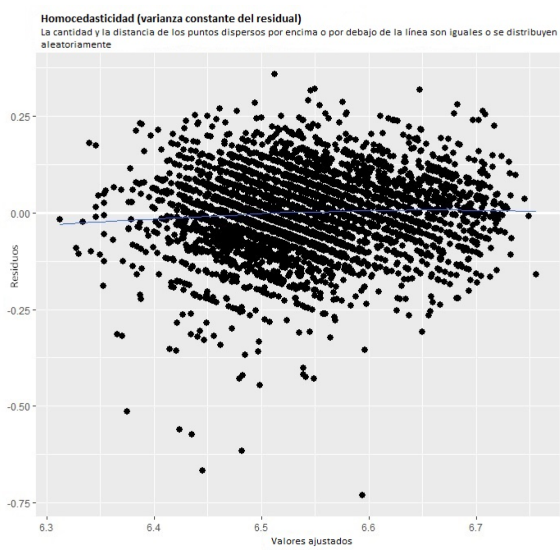
Verificación de los Supuestos

Homocedasticidad. Los modelos de regresión multinivel asumen que la varianza de los residuos es igual en todos los grupos. Se realizó un análisis descriptivo de los residuos (ANOVA) del modelo final de español tercer grado y se obtuvo un p-value de 0,8244 mayor que 0,05, se puede concluir que la varianza de los residuales es igual y, por lo tanto, se cumple el supuesto de homocedasticidad.

Una solución visual. Esta hipótesis es importante debido a que al graficar los residuos frente a valores estimados se pueden ver la variabilidad cuando tiende a crecer o a decrecer con las variables predictoras del modelo.

Figura 1

Varianza constante de los residuos modelo final de español tercer grado



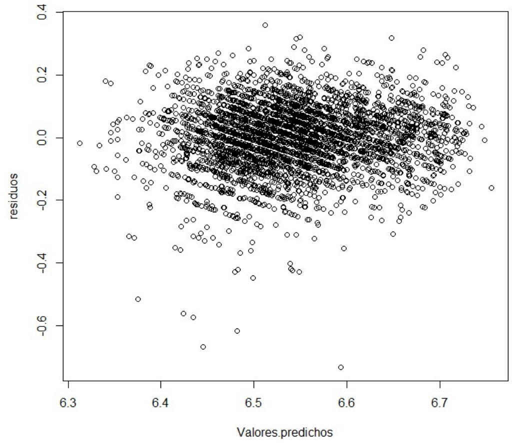
Nota. Fuente: Elaboración propia.

Se observa en la Figura 1 que la varianza permanece constante para todas las puntuaciones de los estudiantes en español tercer grado, con lo cual se confirma la homocedasticidad.

Ortogonalidad de los Componentes Aleatorios y Valores Previstos. Una estrategia comúnmente utilizada para verificar este supuesto es mediante el estudio del gráfico de los valores predichos frente a los residuos estandarizados.

Figura 2

Ortogonalidad modelo final de español tercer grado



Nota. Fuente: elaboración propia.

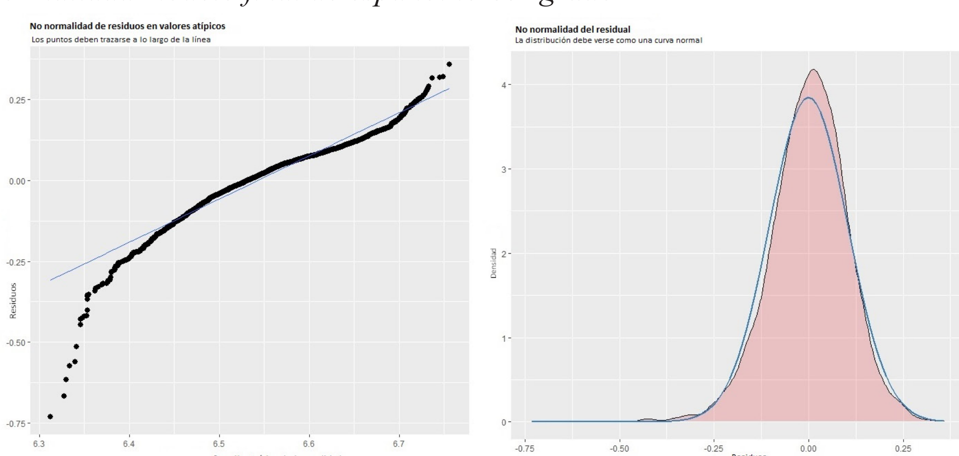
Debido a que se observó una nube de puntos sin tendencia en la Figura 2, se puede afirmar que el supuesto de ortogonalidad se cumple para el modelo de español tercer grado.

Normalidad. En los modelos multinivel se asume que los residuos del análisis están distribuidos normalmente. Los gráficos de normalidad pueden proporcionar una estimación de dónde se encuentran los residuales estandarizados con respecto a los cuantiles normales. Una fuerte desviación de la línea proporcionada indica que los propios residuos no se distribuyen normalmente.

Se puede observar en la Figura 3 que hay una desviación de la línea normal en las colas, sobre todo en la cola inferior, lo que indica la violación del supuesto de normalidad. Se realizó la prueba de normalidad de Shapiro-Wilk y se obtuvo como resultado un valor de W = 0,97683 y p_{value} = 0,00 se rechazó la hipótesis nula que establece que los errores se distribuyen normalmente. Sin embargo, al comparar la densidad de los residuos, notamos que no es muy diferente que la densidad de la distribución Normal, como podemos ver en la Figura 3.

Figura 3

Gráficos de normalidad modelo final de español tercer grado



Nota. Fuente: elaboración propia.

Calidad del Modelo Frecuentista

Finalmente, es posible estimar la proporción de varianza en la variable de resultado contabilizada en cada nivel del modelo. En el contexto del modelado multinivel, los valores de R^2 se estiman para cada nivel del modelo.

Para el modelo final de español tercer grado el R^2 para nivel 1 es de 0,031552559, R^2 para el nivel 2 es 0,630750039 y R^2 total es de 0,245258884; por lo tanto, el modelo final explica el 25% de la varianza total, el 63% de la varianza entre centros y apenas el 3% de la varianza entre estudiantes.

Resultados Bajo el Enfoque Bayesiano

Se trata de obtener el modelo que, mejor se ajuste a los datos utilizando regresión multinivel con un enfoque bayesiano. La información de la base de datos se procesó con el paquete estadístico R (Team, 2020) con la librería brms y rstan (Bürkner, 2017).

Los modelos se ajustaron utilizando 2 cadenas, cada una con 2000 iteraciones, de las cuales las primeras 1000 son para calibrar el muestreador, lo que lleva a un total de 2000 muestras posteriores. Por supuesto, cada iteración es más computacionalmente intensiva y consume más tiempo que las iteraciones de otros algoritmos. Como se transformó la variable dependiente en logaritmo, se utilizó una familia lognormal, lo que implica un modelo logarítmico normal para la variable dependiente puntaje. Se estableció una priori normal con media 0 y desviación estándar 5, para todos los efectos a nivel de población, esto conduce a un muestreo más rápido ya que en este caso se pueden vectorizar los datos a priori.

Se utilizaron los mismos niveles y variables que en el enfoque frecuentista.

Convergencia y Eficiencia para Cadenas de Markov

Los métodos de Montecarlo vía cadenas de Markov (MCMC) son importantes en las estadísticas computacionales, especialmente en las aplicaciones bayesianas donde el objetivo es representar la inferencia posterior utilizando una muestra de sorteos posteriores. Para el caso del paquete Stan en el paquete estadístico R se utilizan algoritmos dinámicos hamiltonianos de Monte Carlo (HMC) (Vehtari et al., 2021).

A continuación, se explican las medidas del diagnóstico proporcionadas por el paquete stan y la librería brms:

1. La función Rhat produce un diagnóstico de convergencia R-hat, que compara las estimaciones entre y dentro de la cadena para los parámetros del modelo y otras cantidades univariadas de interés. Si las cadenas no se han mezclado bien, es decir, las estimaciones entre y dentro de la cadena no concuerdan, R-hat es mayor que 1, es aceptable hasta 1,05.

2. Bulk-ESS es una medida útil para la eficiencia del muestreo en la mayor parte de la distribución, relacionada, por ejemplo, con la eficiencia de las estimaciones de la media y la mediana, y está bien definida incluso si las cadenas no tienen una media o varianza finitas.
3. Tail-ESS es una medida útil para la eficiencia de muestreo en las colas de la distribución, relacionada, por ejemplo, con la eficiencia de la varianza y las estimaciones de cuantiles de cola.

Tanto el Bulk-ESS como el Tail-ESS deben ser al menos 100 (aproximadamente) por Cadena de Markov para que sea confiable e indique que las estimaciones de los respectivos cuantiles posteriores son confiables (Vehtari et al., 2021).

Modelo solo Variables de Estudiantes

El modelo se realizó incluyendo variables relacionadas con los estudiantes, para verificar la significancia de los coeficientes de regresión bayesiana se comprueba si el intervalo de credibilidad correspondiente contiene a cero o no, si no contiene a cero, entonces este coeficiente es significativo. Esto permite evaluar el aporte de cada variable al modelo.

Se obtuvieron los resultados de las 20 variables incluidas en el modelo en tercer grado español de varianzas $\sigma^2_{\alpha 1} = 0{,}010786$ y $\sigma^2_{e1} = 0{,}016314$. De las 20 variables incluidas en este modelo solo 7 resultaron significativas. Se realizó de nuevo el análisis y se obtuvieron los resultados de la estimación de la variación en los interceptos entre escuelas es $\sigma^2_{\alpha 2}$ = 0,010710, Bulk-ESS = 511 y Tail-ESS = 829, mientras que la variación dentro de la escuela se estima en σ^2_{e2} = 0,016319 Bulk-ESS = 3803 y Tail-ESS = 1548, con Rhat = 1 lo que indica que las cadenas convergen.

Modelo solo Variables de Escuela

En este modelo se incorporan las variables correspondientes a nivel de escuela para evaluar el aporte de las mismas al modelo.

Se obtuvieron los resultados de las 15 variables incluidas en el modelo en tercer grado español de varianzas $\sigma^2_{\alpha 1} = 0{,}008142$ y $\sigma^2_{e1} = 0{,}016595$. De las 15 variables incluidas en este modelo solo 4 resultaron significativas y se obtuvieron los resultados de la estimación de la variación en los interceptos entre escuelas es σ^2_{e2} = 0,008439, Bulk-ESS = 648 y Tail-ESS = 1034, mientras que la variación dentro de la escuela se estima en $\sigma^2_{\alpha 1}$ = 0,016599, Bulk-ESS = 4652 y Tail-ESS = 1380, con Rhat = 1 lo que indica que las cadenas convergen.

Modelo Final

Finalmente se desea analizar el efecto que tienen de manera conjunta las variables de los dos niveles sobre el rendimiento de los estudiantes.

Se obtuvieron los resultados en tercer grado español de varianzas $\sigma^2_{\alpha 1} = 0{,}007835$ y $\sigma^2_{e1} = 0{,}016322$, las variables significativas resultantes se presentan en la Tabla 3.

Tabla 3

Variables significativas del modelo final español tercer grado

Nombre de las variables	Estimación	Error Estándar	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Género	-0.0012	0.0006	-0.0023	-0.0001	1.00	2097	1063
Edad	0.0015	0.0006	0.0003	0.0026	1.00	2117	1518
Trabajo infantil	0.0006	0.0003	0.0001	0.0012	1.00	2158	1199
Clima en el aula	0.0011	0.0002	0.0007	0.0014	1.00	2008	1302
Prácticas de recreación	0.0004	0.0001	0.0002	0.0005	1.00	2094	1589
Servicios en el hogar	0.0006	0.0002	0.0003	0.0010	1.00	2072	1368
Servicios básicos en la escuela	0.0027	0.0005	0.0017	0.0037	1.00	494	763
Disponibilidad de libros en la escuela	-0.0029	0.0007	-0.0044	-0.0015	1.01	451	640
Consejo docente	0.0005	0.0002	0.0002	0.0008	1.00	724	1012

Nota. Fuente: elaboración propia, en base a datos del TERCE, 2013.

Se puede observar en la Tabla 3 que de las 11 variables incluidas en este modelo solo 9 resultaron significativas. Se realizó un nuevo análisis y se obtuvieron los resultados de la estimación de la variación en las intersecciones entre escuelas es $\sigma^2_{\alpha 2}$ = 0,007902, Bulk-ESS = 600 y Tail-ESS = 1023, mientras que la variación dentro de la escuela se estima en σ^2_{e2} = 0,016336, Bulk-ESS = 1997 y Tail-ESS = 1299, con Rhat = 1 lo que indicó que las cadenas convergen. Y además con una correlación intraclase ICC de 0,33 lo que indicó que la correlación de los puntajes de las pruebas de español entre los estudiantes dentro de las mismas escuelas es 33%.

Modelo final completo español tercer grado

Nivel 1: Estudiantes

$$log(PE_{ij}) = \beta_{0j} + \beta_{1j}\ Género_{ij} + \beta_{2j}\ Edad_{ij} + \beta_{3j}\ TrabInfL3_{ij} + \beta 4j\ ClimAulaL3_{ij}$$
$$+ \beta_{5j}\ PracRecreL3_{ij} + \beta 6j\ ServiHogarL3ij + \varepsilon ij$$

Nivel 2: Escuela

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\ ServBasEscL3_{0j} + \gamma_{02}\ LibyMatDirL3_{0j} + \gamma_{03}\ ConsDocL3_{0j} + U_{0j}$$

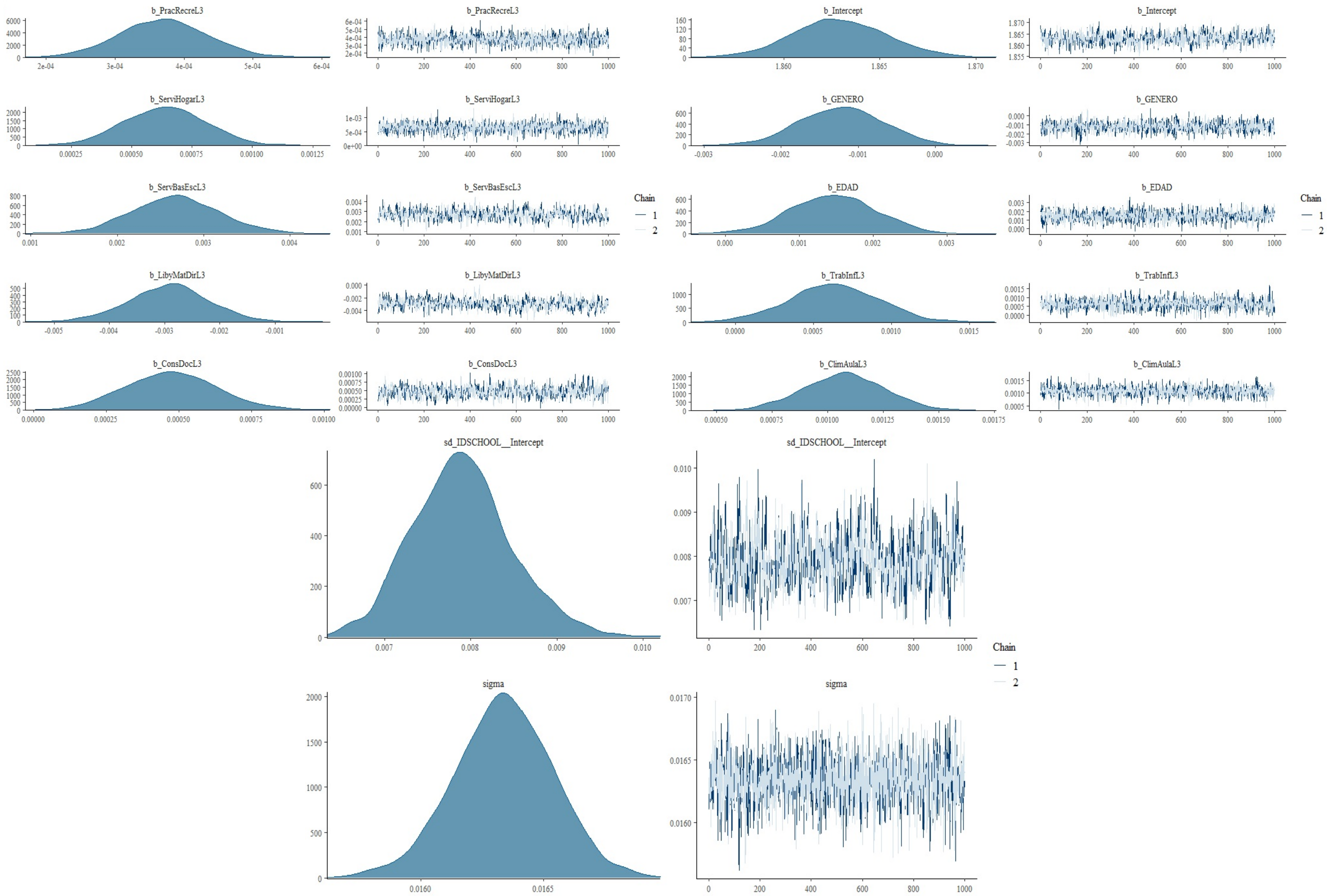
$$log(PE_{ij}) \sim N(\mu_{ij}, \sigma_e^{\ 2})$$

$$\varepsilon_{ij} \sim N(0,5)$$

$$U_{0j} \sim N(0,5)$$

Figura 4

Gráficos de cadenas de Markov de todos los parámetros significativos en el modelo final



Nota. Fuente: elaboración propia.

Según los gráficos de la Figura 4, se puede observar que las cadenas de Markov Monte Carlo convergieron bien y hacia la misma media posterior.

Estimaciones Bajas

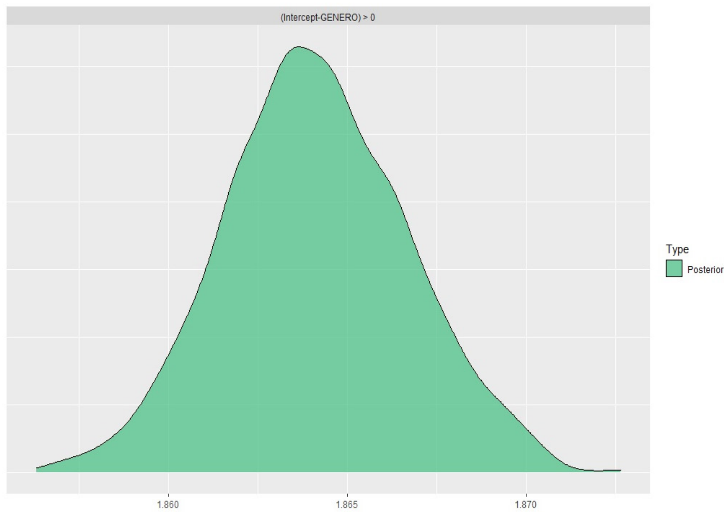
Al observar los efectos fijos, los resultados de los parámetros de las variables son sospechosamente pequeños. Para probar si es más pequeño que el parámetro de desviación estándar de intercepción, se aplicó el método de prueba de hipótesis lineal general (Bürkner, 2017). Este método dio como resultado que el intervalo de credibilidad del 95% unilateral no contiene cero, lo que indica que las desviaciones estándar difieren entre sí en la dirección esperada. Se realizó con cada variable y el resultado fue el mismo.

Por ejemplo: para la variable género en el modelo español tercer grado la prueba de hipótesis indicó que existe una probabilidad 1 de que la hipótesis Intercepto-género sea mayor que 0, se obtuvo el gráfico (Figura 5) donde se observó la proporción de muestras de la parte posterior que son mayores que cero, ver Figura 5.

Es importante notar que este tipo de comparación no es posible fácilmente cuando se aplican métodos frecuentistas, porque en este caso solo están disponibles estimaciones puntuales para las desviaciones estándar y las correlaciones a nivel de grupo.

Figura 5

Variable género en el modelo español tercer grado, se observa la proporción de muestras de la parte posterior que son mayores que cero



Nota. Fuente: Elaboración propia.

Selección del Modelo

Se utilizó el método paso a paso, este requiere de criterio matemático para determinar si el modelo, mejora o no, con cada incorporación o extracción. Hay dos familias de estrategias para la comparación de los modelos: validación cruzada utilizando validación cruzada suavizada por Pareto PSIS con LOO para paquetes stan y criterios de información con WAIC. Estas estrategias tratan de visualizar cómo funcionarán bien los modelos, en promedio, en la predicción de nuevos datos, donde elpd-diff es la diferencia en elpd (logaritmo esperado predictivo densidad puntual) para dos modelos, si se comparan más de dos modelos, la diferencia se calcula en relación con el modelo con elpd más alto y se-diff es el error estándar de la diferencia, ver Tabla 4.

Tabla 4

Resultado del modelo español tercer grado usando PSIS-LOO

Modelo	elpd-diff	se-diff	LOO
Modelo final	0	0	-5801.1
Modelo 2	3.4	6.1	-5794.3
Modelo 4	-58.1	12.4	-5684.9

Nota. Fuente: elaboración propia, en base a los resultados obtenidos por las librerías brms y rstan.

El modelo final que obtuvo como resultado el valor de LOO más pequeño y es el que contiene las variables de los niveles estudiantes y escuela es ligeramente mejor que el modelo 2 que solo contiene las variables a nivel de estudiante. El modelo 4 que contiene solamente las variables a nivel escuela no es el mejor de los tres modelos. Todas las estimaciones en los modelos están bien ya que k<0,7 en distribución de Pareto.

Tabla 5

Resultado del modelo español tercer grado usando WAIC

Modelo	elpd-diff	se-diff	WAIC
Modelo final	0	0	-5802.2
Modelo 2	-3.1	6.1	-5795.9
Modelo 4	-58.1	12.4	-5684.9

Nota. Fuente: elaboración propia, en base a los resultados obtenidos por las librerías brms y rstan.

WAIC confirma lo que indicó PSIS-LOO, el modelo final es el mejor ya que contiene los valores más pequeños, ver Tabla 5.

Calidad del Modelo Bayesiano

Se utilizó un enfoque bayesiano para definir R^2 en cada nivel del modelo multinivel es decir varianza explicada, no se ajusta un modelo nulo si no que se compararon varianzas en un solo modelo ajustado y se calcula el factor de agrupación λ para resumir el grado en que las estimaciones en cada nivel del modelo se agrupan en función de la relación de regresión específica del nivel.

Para el modelo final de español tercer grado el R^2 para nivel 1 es de 0,2930936, R^2 para el nivel 2 es 0,26621 y R^2 total es de 0,3555721; por lo tanto, el modelo final explica el 36% de la varianza total, el 27% de la varianza entre centros y el 29% de la varianza entre estudiantes. El valor aproximado de λ para este modelo a nivel de estudiantes es 0,11 y para escuela es 0,41 lo que sugiere que hay mayor nivel de información dentro de grupo es decir a nivel de escuela.

Conclusiones

Conclusiones sobre los Modelos

1. Se identifica un fuerte impacto positivo de las variables, clima en el aula escolar y prácticas de recreación, lo que permite un alto rendimiento de los estudiantes.
2. El hecho de no disponer de libros y materiales en la escuela representa un efecto negativo en rendimiento de los estudiantes en español tercer grado.
3. Se puede verificar por medio de las variables instalaciones en la escuela y servicios básicos de la escuela, que unas mejores características de la escuela tienen un efecto positivo y significativo sobre el

rendimiento de los estudiantes. De igual forma, aquellos estudiantes que no trabajan y tienen supervisión de estudios en el hogar, tendrán un mejor rendimiento.

4. Por medio de los ingresos en el hogar se ha constatado que los estudiantes pertenecientes a familias más favorecidas económicamente tienen un mejor rendimiento.
5. El género del estudiante muestra una relación de menor magnitud, pero igualmente significativa, en este caso es posible señalar que las niñas alcanzan mayor logro en lectura que los niños.

Conclusión General

Se puede verificar la fortaleza del uso de los modelos multinivel para el análisis de datos con estructura jerárquica, ya que los resultados de los modelos muestran la importancia del nivel escuela, al encontrar que los logros alcanzados por los estudiantes en gran medida se explican por las características propias de las instituciones. El análisis de las características de las escuelas muestra los distintos factores que ayudan a comprender el desempeño académico de los estudiantes en español tercer grado, como ser: clima en el aula escolar, disponibilidad de libros y materiales en la escuela, instalaciones en la escuela y servicios básicos de la escuela.

Consideraciones Finales

Debido al poder del análisis multinivel para identificar la variabilidad de los grupos dentro y entre ellos, se recomienda el uso de los mismos por el eficiente resultado como herramienta de análisis.

Para futuros estudios, se podría ampliar el número de niveles utilizando otras bases de datos más actualizadas, con los resultados del rendimiento en español tercer grado en los años de la pandemia de la COVID-19, así como los desastres naturales Eta e Iota, que permitirán observar otros factores relacionados con el entorno de la comunidad del estudiante. Además, que se pueda verificar si los resultados numéricos en cuanto a las estimaciones de los parámetros y la precisión de las mismas se mantiene similares entre enfoques frecuentista y bayesiano.

Referencias Bibliográficas

Aitkin, M., y Longford, N. (1986). *Statistical Modelling Issues in School Effectiveness Studies. Journal of the Royal Statistical Society*. Series A (General), Vol. 149, No. 1, 1-43. doi:https://doi.org/10.2307/2981882

Bates, D., Mächler, M., Bolker, B., y Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bickel, R. (2007). *Multilevel analysis for applied research: it's just regression!*. New York: The Guilford Press.

Bürkner, P. C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1). doi:https://doi.org/10.18637/jss.v080.i01

Bustamante, A. M. (2015). *Dirección escolar exitosa en España*. [Tesis doctoral, Universidad Autónoma de Madrid]. https://repositorio.uam.es/handle/10486/666722

Fernández, V. P. (2012). *Los modelos multinivel en el análisis de factores de riesgo de sibilancias recurrentes en lactantes. Enfoques frecuentista y bayesiano*. [Tesis doctoral, Universidad de Murcia]. https://tesisenred.net/handle/10803/109213

Finch, W.H., Bolin, J.E., y Kelley, K. (2019). *Multilevel Modeling Using R (2nd ed.)*. Chapman and Hall/CRC. https://doi.org/10.1201/9781351062268

Gelman, A., y Pardoe, I. (2012). *Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models*. Technometrics, 241-251. doi:https://doi.org/10.1198/004017005000000517

Hox, J., Moerbeek, M., y Schoot, R. V. (2018). *Multilevel analysis techniques and applications*. New York: Taylor & Francis Group. https://lccn.loc.gov/2017013032

Johnson, R., y Wichern, D. (2007). *Applied Multivariate Statistical Analysis*. Prentice Hall. https://libgen.li/file.php?md5=43c0d654a6ec0d23aa6697c7619daf27

Jurado, J. C. (2013). Análisis multinivel del rendimiento escolar en matemáticas para cuarto grado de Educación Básica Primaria en Colombia. *Sociedad y Economía*, (25), 205-236. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-63572013000200010

McElreath, R. (2020). *Statistical Rethinking A Bayesian Course with Examples in R and Stan*. Columbia: Taylor & Francis Group. Obtenido de https://xcelab.net/rm/statistical-rethinking/

Nalborczyk, L., Batailler, C., Loevenbruck, H., Vilain, A., y Bürkner, P. (2019). *An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian*. University Grenoble Alpes. Obtenido de https://osf.io/dpzcb/

Secretaría de Educación. (1999). *Carta/acuerdo entre la Secretaría de Educación y la UNESCO*. Honduras.

Seghouane, A. K., y Amari, S. I. (2007). The AIC Criterion and Symmetrizing the Kullback–Leibler Divergence. *IEEE Transactions on Neural Networks*, 18(1) . doi:10.1109/TNN.2006.882813

Stringfield, S., Reynolds, D., Creemers, B., Nesselrodt, P., Schanffer, E., y Teddlie, C. (1994). A model of elementary school effects. *Advances in School Effectiveness Research and Practice*, 153-187. https://doi.org/10.1016/B978-0-08-042392-0.50013-9

Team, R. C. (2020). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Torrecilla, J. M. (2008). Los modelos multinivel como herramienta para la investigación educativa. *Revista Internacional de Investigación en Educación*, 1(1), 45-62. http://www.redalyc.org/articulo.oa?id=281021687004

UNESCO. (2000). *Situación y tendencias 2000 Evaluación del aprovechamiento escolar*. Foro consultivo internacional sobre educación para todos.

UNESCO. (2007). *Informe de seguimiento de la educación para todos en el mundo*. Paris: Fortemps.

UNESCO. (2016). *Tercer estudio comparativo y explicativo: reporte técnico. Oficina regional de educación para América Latina y el Caribe*. https://unesdoc.unesco.org/ark:/48223/pf0000247123

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., y Bürkner, P. C. (2021). Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2), 667–718. doi:https://doi.org/10.1214/20-BA1221

Velez, E., Schiefelbein, E., y Valenzuela, J. (1994). Factores que Afectan el Rendimiento Académico en la Educación. *Revista latinoamericana de Innovaciones Educativas*, (17). https://core.ac.uk/download/pdf/143614621.pdf