

Resumen

Este estudio aborda la creciente importancia de determinar los factores que influyen en los puntajes obtenidos por estudiantes latinoamericanos en las pruebas PISA, en donde se ha evidenciado bajas calificaciones para la región. Por medio del modelo de regresión logística y el uso de datos en competencias matemáticas para los estudiantes latinoamericanos, género y variables de índole socioeconómica se busca determinar las probabilidades de pertenecer a los niveles de alfabetización matemática 2, 3 y 4. Los resultados mostraron que el género toma una relevancia creciente conforme los estudiantes se acercan a la cola derecha de la distribución de calificaciones; además, las variables estatus migratorio y nivel académico de la madre no son determinantes en el desempeño de los estudiantes de alto rendimiento.

Palabras clave: modelo de regresión logístico ordinal, matemáticas, pruebas PISA, nivel académico

Abstract

This study addresses the growing importance of determining the factors that influence the scores obtained by Latin American students in the PISA tests, where low scores have been evidenced for the region. By using the logistic regression model and the use of data on mathematical competencies for Latin American students, gender, and socioeconomic variables, we sought to determine the probabilities of belonging to levels of mathematical literacy 2, 3 and 4. The results showed that the gender takes an increasing relevance as students approach the right tail of the grade distribution. Furthermore, the variables migratory status and academic level of the mother are not determinants in the performance of high-achieving students.

Keywords: ordinal logistic model, mathematics, PISA tests, academic level

Introducción

Gurría (2014), Secretario General de la OCDE, explicó que el Programa de Evaluación Internacional de Alumnos (PISA) consiste en un proyecto a cargo de la Organización para la Cooperación y el Desarrollo Económico (OCDE) cuyo objetivo es evaluar la formación de los alumnos cuando llegan al final de la etapa de la enseñanza obligatoria a los 15 años por medio de exámenes estandarizados. Las evaluaciones de PISA se centran en un área temática concreta en cada una de sus ediciones: la lectura, las matemáticas y las ciencias. Para el año 2012, el enfoque de la prueba fue el área de matemática.

La alfabetización matemática de PISA se define como “la capacidad de un individuo para formular, emplear e interpretar las matemáticas en una variedad de contextos. Incluye razonamiento matemático y el uso de conceptos matemáticos, procedimientos, hechos y herramientas para describir, explicar y predecir fenómenos” (OCDE, 2017, p.134). Según Castro et al. (2017), en PISA 2012, los países latinoamericanos se ubicaron dentro del tercio de economías con peores resultados a nivel mundial. Siete países de la región obtuvieron un puntaje promedio en matemática por debajo del nivel de alfabetización matemática 2, de seis niveles posibles.

Castro et al. (2017), destaca que en la región persiste inequidad a lo interno de los países, donde los estudiantes de regiones favorecidas suelen obtener calificaciones más altas respecto a sus contrapartes. Las explicaciones respecto a la diferencia en la calificación de los estudiantes en pruebas estandarizadas son abundantes, y estudian principalmente el impacto de factores sociales, familiares y económicos sobre el éxito académico que obtienen los individuos. El sistema educativo debería estar enfocado en reducir las brechas entre los estudiantes y promover la movilidad social, por lo que entender el impacto de estos factores es una prioridad para la mejora de la calidad educativa en la región.

Por lo tanto, el objetivo principal de este estudio radica en determinar la influencia que presentan las variables de género, migración, índice económico y educación de los padres sobre la probabilidad de estar en el segundo, tercer y cuarto nivel de alfabetización matemática en el año 2012 para los estudiantes latinoamericanos, por medio de la utilización de la regresión logística. Además, como segundo objetivo se busca estimar los efectos marginales promedio para medir el efecto en las probabilidades de pertenecer a los niveles de alfabetización derivados del aumento en un nivel o unidad en las variables mencionadas.

Bajo esta línea, se espera que conforme se avance en el nivel de alfabetización matemática obtenido (mientras más se acerquen los sujetos a la cola derecha de la distribución), el ser hombre, tener un índice socioeconómico o padres con niveles educativos elevados otorgue una probabilidad más alta de estar en el grupo de alto rendimiento. Por otro lado, ser migrante de primera o segunda generación disminuiría estas probabilidades. El desarrollo de este estudio se divide en cuatro partes; revisión de literatura donde se justifican las variables seleccionadas, descripción de los datos y el método estadístico empleado, sección de resultados donde se busca dar respuesta a los objetivos y, por último, una sección de discusión donde se resumirá las implicaciones de los principales hallazgos.

Revisión de literatura

¿Por qué existe una brecha de rendimiento entre los estudiantes latinoamericanos? A pesar de que la prueba estandarizada de PISA aplicada en el 2012 aporta una imagen en cuanto a los retos que enfrenta el sistema educativo en capacidad matemática; es necesario tomar en cuenta que el desempeño en las pruebas no cuenta toda la historia y las explicaciones respecto al bajo desempeño de los países latinoamericano no son totalmente claras.

Las teorías alrededor de este fenómeno abordan a los sistemas educativos, variables de escolaridad (calidad de los maestros, tipo de colegio), variables socioeconómicas (estatus socioeconómico, ocupación de los padres), demográficas, biológicas, entre muchas otras. A continuación, se realiza un breve recorrido de algunas investigaciones que abordan posibles explicaciones a la brecha en rendimiento matemático que atañe a los países de la región que participan en las evaluaciones PISA.

Smith et al., (2008) considera que el nivel educativo de los padres está ligado al éxito académico de los hijos. Según Rodríguez y Guzmán (2019) se ha demostrado que en contextos sociales y familiares favorables (progenitores con estudios universitarios, de estratos ocupacionales medios y privilegiados, cuyas relaciones familiares se caracterizan por el apoyo afectivo) las probabilidades de éxito académico se incrementan; y, consecuentemente, disminuyen cuando los estudiantes viven en contextos desfavorables o con menos oportunidades (pobreza, marginalidad, disfuncionalidad familiar, baja educación de la madre, familia monoparental y hacinamiento en el hogar).

Aunado a lo anterior, León y Youn (2016), analizaron la influencia del estatus socioeconómico en el desempeño matemático en una muestra de estudiantes peruanos provenientes de instituciones educativas públicas y encontró que existe una relación positiva y significativa entre el rendimiento de los estudiantes y su nivel socioeconómico. Por lo tanto, el

nivel socioeconómico de las familias de los alumnos estaría jugando un rol importante para explicar las diferencias en el rendimiento en Matemática entre estudiantes. Esto, según Feinstein et al., (2008) también implica que el nivel educativo de los padres tiene un efecto directo en el nivel económico de la familia. Además, la composición social del alumno como la de su escuela se asocia a factores contextuales como la repetición, el ausentismo, la matrícula oportuna, entre otros. Como resultado, esto implica que los estudiantes presenten pérdida en la motivación y ansiedad hacia las pruebas.

En adición, se ha documentado la existencia de una brecha de género en las pruebas estandarizadas de matemática. De acuerdo con González y Goiricelaya (2012), las estudiantes mujeres tienen un rendimiento inferior a los hombres en casi todos los países que participan en las evaluaciones PISA. Liu y Wilson (2009), utilizaron un modelo logístico multinomial de coeficientes aleatorios multidimensionales con lo que obtuvieron que los hombres sostienen una ventaja pequeña en las pruebas PISA en los años 2000 y 2003.

Hay muchas explicaciones posibles para la existencia de esta brecha de género. En primer lugar, se encuentran los factores personales, como la percepción de los padres de familia y el trato de los maestros hacia las estudiantes mujeres. Rätty et al., (2002), encontró que el éxito de las mujeres normalmente se atribuye a la disciplina y trabajo duro, mientras que el de los hombres se asocia con talento e inteligencia. Mientras que Freeman (1991) y Furnham et al., (2002), encontraron que los padres de familia asumen que la capacidad matemática de sus hijos es mayor que la de sus hijas.

Además, la percepción de la capacidad personal juega un rol importante en el desempeño de las mujeres en estas pruebas. El estudio de Chipman et al., (1992), pudo demostrar una fuerte influencia de la ansiedad y confianza en matemáticas, independientemente de los efectos de las puntuaciones cuantitativas del examen estandarizado SAT. Es necesario mencionar la existencia

de evidencia que muestra que las diferencias entre hombres y mujeres se hacen más notables dentro de los alumnos que se encuentran en el extremo derecho de la distribución, tal como documenta Zhou et al., (2017).

Asimismo, la influencia del estatus migratorio del estudiante no debe ser despreciado dentro del estudio del rendimiento académico. Dustmann et al., (2014) mostraron que, en promedio, estudiantes inmigrantes en la mayoría de los países europeos tienden a obtener puntuaciones significativamente más bajas en las pruebas PISA que los estudiantes nativos. Por lo tanto, existen factores que afectan el riesgo académico de estudiantes en estatus migratorio como lo son el sentido de pertenencia a la institución y la etapa de llegada a ésta.

Metodología

Muestra de datos

El origen de la base de datos utilizada en este estudio proviene del cuestionario de estudiantes del Programa Internacional de Evaluación de Estudiantes (PISA) durante el 2012. La base total cuenta con 485490 observaciones de 65 países. Para realizar este estudio se seleccionó las 11350 observaciones correspondientes a estudiantes latinoamericanos provenientes de Colombia, Uruguay, Costa Rica, Chile, Ecuador, Perú, México, Argentina y Brasil.

En la siguiente sección se explica el tratamiento de las variables seleccionadas para obtener los resultados finales, adicionalmente (ver anexo en la Tabla 1.a), se pueden encontrar las principales estadísticas descriptivas.

Variables

Con estos precedentes se hablará de dos tipos de variables en el estudio:

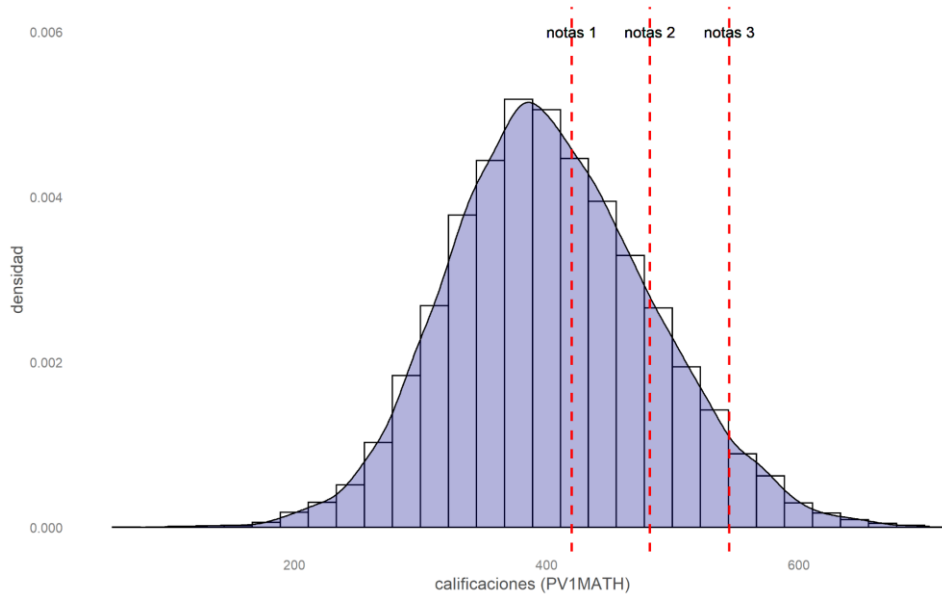
Variable dependiente: Calificaciones

Se toma la variable PV1MATH para medir el rendimiento estudiantil, la cual corresponde al primer valor plausible en la escala matemática. De acuerdo con Huslin (2014) los valores plausibles incluidos en PISA (PV1MATH...PV5MATH) son el principal predictor del interés para el rendimiento en matemática, que se mide a nivel individual. En el caso de los estudiantes latinoamericanos, el valor mínimo que toma esta variable es 59,28 y el máximo es 703,54.

PISA establece la existencia de seis niveles de alfabetización matemática. En el caso de los estudiantes latinoamericanos, en la figura 1 muestra la distribución de la variable PV1MATH a partir de la que se determinan los niveles de alfabetización matemática para la prueba PISA. La calificación de 420 (percentil 56) se considera la mínima para pertenecer al nivel de alfabetización matemática 2. De la misma manera, se determina el nivel de alfabetización matemática 3, cuando un estudiante obtiene una nota mínima de 480 (percentil 83). Por último, el nivel de alfabetización matemática 4, se ubica a partir de una calificación de 545 (percentil 95). Cabe resaltar que no se cuenta con suficientes observaciones de estudiantes latinoamericanos que pertenezcan a niveles de alfabetización 5 y 6, por lo que estos son excluidos dentro de este estudio.

Figura 1

Gráfico de distribución de la variable PV1MATH para la prueba de matemática PISA 2012



Fuente: Elaboración propia con datos de PISA 2012.

Variables independientes

- **Género:** El género se basa en la codificación (1) mujer u (2) hombre.
- **Nivel educativo de la adres:** Para el nivel educativo de la madre (MISCED) se tomó en cuenta la Clasificación Internacional Normalizada de la Educación (ISCED), las categorías son las siguientes: (0) Ninguna, (1) educación primaria, (2) secundaria baja, (3) dividida en nivel 3B o 3C que es vocacional/ vocacional secundaria alta, (4) en sus niveles 3A que es secundaria alta general y postsecundaria no terciaria, (5) que es vocacional terciaria y (6) corresponde a educación terciaria y posgrado. (OCDE, 2014)

Para efectos de este trabajo se recodificó la variable de educación de la madre en 6 niveles: 0-Ninguna, 1: Educación primaria, 2(2 y 3): Educación secundaria, 3(4): postsecundaria no terciaria, 4(5): Educación terciaria profesional, 5(6) terciario y postgrado con orientación teórica.

- **Estatus migratorio:** También se recopiló información sobre el país de nacimiento de los estudiantes y sus padres. Lo incluido en la base de datos fueron tres variables específicas de cada país relacionadas con el país de nacimiento del alumno, su madre y su padre. Para realizar la investigación se trabajó en las tres categorías originales de PISA: (1) El país de nacimiento es el mismo que el país de evaluación, (2) estudiantes de segunda generación (los nacidos en el país de la evaluación, pero cuyos padres nacieron en otro país) y (3) estudiantes de primera generación (aquellos estudiantes nacidos fuera del país de evaluación y cuyos padres también nacieron en otro país). (OCDE, 2014)
- **Nivel socioeconómico:** Las variables que componen el ESCS para PISA 2012 incluyeron: posesiones del hogar (HOMEPOS) - que comprendía todos los ítems de las escalas WEALTH, CULTPOS y HEDRES, así como libros en el hogar (ST28Q01) recodificado en una variable categórica de cuatro niveles (menor o igual a 25 libros, 26-100 libros, 101-500 libros, más de 500 libros); la ocupación parental más alta (HISEI); y la educación de los padres más alta expresada como años de escolarización (PARED). (OCDE, 2014).

Para los países participantes, las puntuaciones de la ESCS se obtuvieron con la expresión:

$$ESCS = \frac{\beta_1 HISEI' + \beta_2 PARED' + \beta_3 HOMEPOS'}{\varepsilon_f}$$

Donde $\beta_1, \beta_2, \beta_3$ son las cargas factoriales de la OCDE, HISEI', PARED' y HOMEPOS' las variables " estandarizadas por la OCDE " y ε_f es el valor propio del primer componente principal. (OCDE, 2014).

Seguidamente, se muestra en la tabla 1 la distribución de hombres y mujeres para cada nivel de alfabetización matemática considerado dentro de este estudio. Es notoria la existencia de la brecha de género para los niveles de alfabetización, ya que en ninguno se puede encontrar una repartición igualitaria entre géneros. La cantidad de mujeres conforme se avanza en los niveles de alfabetización es decreciente, en contraste con los hombres, que pasan de representar el 53,4% de la población en el nivel 2 hasta el 64,9% en el nivel 4.

A partir de esta tabla se puede intuir una concentración mayor de estudiantes hombres en el nivel de alfabetización más alto. Respecto a la condición de migrantes se vislumbra cómo la mayoría de los estudiantes latinoamericanos se identifican como nacidos en el país en el cual se aplica la prueba, con lo cual el peso de los estudiantes migrantes es muy bajo respecto a la muestra total.

Tabla 1

Distribución de estudiantes por género y condición migratoria según nivel de alfabetización PISA 2012

	Nivel de alfabetización 2		Nivel de alfabetización 3		Nivel de alfabetización 4	
	No pertenece	Pertenece	No pertenece	Pertenece	No pertenece	Pertenece
Género						
Hombres	2961 (43,5)	2427 (54,4)	4291 (45,4)	1097 (57,8)	5042 (46,6)	346 (64,9)
Mujer	3841 (56,5)	2121 (46,6)	5160 (54,6)	802 (42,2)	5775 (53,4)	187 (35,1)
Total	6802	4548	9451	1899	10817	533
Condición migratoria						
Primera gen	54 (0,83)	17 (0,38)	65 (0,71)	6 (0,32)	70 (0,67)	1 (0,19)
Segunda gen	63 (0,96)	21 (0,47)	78 (0,85)	6 (0,32)	83 (0,79)	1 (0,19)
Nativos	6428 (98,21)	4467 (99,16)	9015 (98,44)	1880	10367 (98,55)	528 (99,62)
Total	6545	4505	9158	1892	10520	530

Fuente: Elaboración propia con datos de PISA 2012

Por otra parte, en la tabla 2 se repite el análisis según el nivel educativo de los estudiantes. Resulta lógico suponer que los estudiantes cuyos padres poseen algún título de educación terciaria se aglomeran en los niveles de alfabetización más alto; sin embargo, se puede observar un desbalance en el nivel dos y tres para los estudiantes de América Latina. Adicionalmente, se puede notar que para la no tenencia de grados académicos es mayor la cantidad de estudiantes que califican dentro del nivel de calificación 4 en comparación a aquellos que quedan por fuera.

Tabla 2

Distribución de estudiantes por grado de escolaridad de la madre según nivel de alfabetización PISA 2012

	Nivel de alfabetización 2		Nivel de alfabetización 3		Nivel de alfabetización 4	
	No pertenece	Pertenece	No pertenece	Pertenece	No pertenece	Pertenece
Grado de escolaridad						
Ninguno	43,5	54,4	45,4	57,8	46,6	64,9
Educación primaria	56,5	46,6	54,6	42,2	53,4	35,1
Educación secundaria	22,70	21,51	22,74	19,65	22,64	13,75
Postsecundaria no	21,02	22,96	21,70	22,31	21,91	19,77
Terciaria profesional	8,21	12,39	9,07	13,95	9,77	12,62
Terciaria y posgrado	11,48	25,62	13,88	33,49	15,72	46,89
Total	100	100	100	100	100	100

Nota: Cifras en porcentaje

Fuente: Elaboración propia con datos de PISA 2012

Explicación teórica del modelo

Regresión logística ordinal

El modelo de regresión logística analiza variables dependientes, también nombradas variables latentes, distribuidas de forma binomial. De acuerdo con Wooldridge (2010), la variable latente se define por la siguiente función indicadora.

$$y^* = \beta_0 + \mathbf{X}\beta + \varepsilon, \quad y = 1[y^* > 0],$$

Donde, la variable dependiente y , toma valores 0 cuando no se presenta la característica de interés y 1 cuando sí se presenta. En nuestro caso, la variable latente toma el valor de 1 cuando el sujeto obtiene una calificación que lo sitúa por encima del percentil 56, percentil 83 o percentil 95.

El modelo de regresión logística se define de la siguiente manera:

$$\log\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

- $\log\left(\frac{P_i}{1-P_i}\right)$: Log odds o likelihood
- $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$: Variables independientes

Vitola (2015), explica que los coeficientes asociados a las variables independientes en el modelo logístico se calculan con el método de máxima verosimilitud, en contraste con los modelos lineales. El tipo de relación que tienen las variables explicativas con la variable dependiente se interpreta de acuerdo con el signo que tienen sus coeficientes. Si el signo es positivo, la relación entre las variables es directa; si es negativo la relación entre es inversa.

Sin embargo, tal como se explica los modelos de tipo logístico son muy difíciles de interpretar. Por lo que, la magnitud de los coeficientes puede entenderse con mayor facilidad por medio del uso de la Razón de probabilidades u *Odds Ratio*. Según Alderete (2006), los *Odds Ratio* corresponden, a la razón entre probabilidades de la variable dependiente para cada uno de los valores de la variable independiente.

Para el modelo de regresión logístico, se toman los coeficientes estimados, y mediante la aplicación de la función exponencial de estos coeficientes se obtienen los *Odds Ratio*. La obtención de un valor de 1 indica ausencia de asociación entre las variables. Cuando es inferior a

1, señala la existencia de una relación negativa; además conforme más se aleje de 1 más fuerte se considera la asociación existente entre las variables. Los *Odds Ratio* se obtienen por medio de la expresión:

$$\exp(\beta_i) = \frac{\frac{P(Y = 1|X = 2)}{1 - P(Y = 1|X = 2)}}{\frac{P(Y = 1|X = 1)}{1 - P(Y = 1|X = 1)}}$$

Donde:

- $\frac{P(Y=1|X=2)}{1-P(Y=1|X=2)}$: *Odds en el grupo de interés*
- $\frac{P(Y=1|X=1)}{1-P(Y=1|X=1)}$: *Odds en el grupo de no interés*

Es necesario aclarar que, al realizar la regresión logística, el modelo toma como referencia la primera clase de cada variable categórica. Para la base de datos considerada en este estudio, los *Odds Ratio* de género se deben interpretar respecto a las mujeres. Nivel educativo se compara respecto al nivel 0 que es no contar con grados académicos; estatus migratorio se interpreta respecto a la categoría 1 que corresponde a los estudiantes cuyos padres y ellos mismos son residentes.

Asimismo, en este estudio se adiciona el uso de los efectos marginales promedio para describir el impacto de cada variable dependiente sobre la probabilidad de la variable respuesta manteniendo constantes todas las demás variables del modelo. De acuerdo con Trenton (2019), los efectos marginales promedio se estiman al calcular los efectos marginales para cada observación en la muestra, a los cuales se les aplica el promedio total para cada efecto marginal observado en la muestra.

Por medio de esta metodología y las variables anteriormente descritas, se plantean los siguientes tres modelos logísticos para encontrar la influencia del género y sus variables

socioeconómicas en el rendimiento de matemáticas en los estudiantes participantes en la prueba PISA 2012. En el modelo 1, la variable latente toma el valor de 1 para los estudiantes pertenecientes al nivel 2 de alfabetización; bajo la misma lógica, el modelo 2 y 3 se asocia a los estudiantes de alfabetización 3 y 4, respectivamente.

Modelo 1

$$P(\text{notas.1}|x) = f(b_0 + b_1\text{género} + b_2\text{indice.economico} + b_3\text{nv1.academico} + b_4\text{status.migra}) + \varepsilon_i$$

Modelo 2

$$P(\text{notas.2}|x) = f(b_0 + b_1\text{género} + b_2\text{indice.economico} + b_3\text{nv2.academico} + b_4\text{status.migra}) + \varepsilon_i$$

Modelo 3

$$P(\text{notas.3}|x) = f(b_0 + b_1\text{género} + b_2\text{indice.economico} + b_3\text{nv3.academico} + b_4\text{status.migra}) + \varepsilon_i$$

Paquetes utilizados para el análisis de la información recogida y versión de R utilizada

La versión de R utilizada fue la 4.0.3 y para realizar la regresión logística se utilizaron los siguientes paquetes:

Librería Haven (versión: 2.4.1): Permite a R leer varios formatos de datos de otros softwares estadísticos como lo es Stata, SPSS, SAS, se utilizó para leer la base original de PISA en formato Stata. Creada por Hadley Wickham y Evan Miller.

Librería Tidyverse (versión: 1.3.1): Tidyverse engloba muchos paquetes con múltiples funciones, entre los que se usaron para la investigación están: dplyr; tidyr y ggplot2. Los dos

primeros se utilizaron para limpiar la base original para seleccionar y filtrar las observaciones de interés. Ggplot2 se utilizó para crear gráficos. Creada por Hadley Wickham.

Librería arm (versión: 1.11-2): Este paquete es multi funcional y sirve para hacer análisis de datos usando regresión y modelos jerárquicos multinivel. Se utilizó para graficar los intervalos de confianza de los modelos. Generado por Andrew Galman, Yu-Sung Su, Masanao Yajima, Jennifer Hill, Maria Grazia Pittau, Jouni Kerman, Tian Zheng.

Librería sjPlot (versión: 2.8.8): Este paquete pertenece a un paquete en desarrollo, pertenece a devtools pero para instalarlo se debe hacer por medio de github. Se utilizó para crear plots de los efectos marginales promedio en los tres modelos. Esta librería fue creada por Daniel Lüdecke, Alexander Bartel, Carsten Schwemmer, Chuck Powell, Amir Djalovski y Johannes Titz.

Librería see (0.6.4): Este paquete es un complemento visual para ggplot. Se utilizó para dar una estética más llamativa a los gráficos. Creado por Daniel Lüdecke, Dominique Makowski, Indrajeet Patil, Mattan S, Ben-Shachar, Brenton M. Wiernik, Brenton M. Wiernik y Jeffrey R. Stevens.

Librería ggeffects (1.1.0): Es un complemento para la visualización de datos junto con ggplot para la estética de los gráficos. Creado por Daniel Lüdecke, Frederik Aust, Sam Crawley y Sam Crawley.

Librería Margins (0.3.26): Margins es un paquete original del software Stata y es bastante funcional para obtener resultados de modelos de regresión. Se utilizó en la investigación para

realizar los efectos marginales promedio en los tres modelos. Sus autores son Thomas J. Leeper, Jeffrey Arnold, Vincent Arel-Bundock y Jacob A. Long

Resultados

Al realizar la regresión logística, se obtienen los coeficientes numéricos asociados a cada una de las variables seleccionadas y el valor asociado, con el cual determina el grado de significancia de estas. La interpretación de los resultados en esta sección se concentrará en los coeficientes que resultaron significativos. En la tabla 3, se presentan los coeficientes numéricos para los modelos logísticos.

Tabla 3

Coefficientes numéricos obtenidos en los modelos logísticos propuestos

Variable	Modelo 1	Modelo 2	Modelo 3
Género	0,3970***	0,5045***	0,7386***
Índice económico	0,6131***	0,7071***	0,9871***
Nivel educativo madre 1	0,2939**	0,1935	-0,1832
Nivel educativo madre 2	0,4081***	0,5746***	0,0690
Nivel educativo madre 3	0,2074*	0,3708*	-0,0302
Nivel educativo madre 4	0,2049*	0,4156*	-0,1851
Nivel educativo madre 5	0,1517	0,3599*	0,0020
Estatus migratorio 2	-0,4001	-0,6138	-0,5545
Estatus migratorio 3	-0,7144*	-0,6185	-1,0088

Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

En primer lugar, los coeficientes de índice económico y nivel educativo de la madre poseen signo positivo por lo que podemos establecer que para los tres modelos estudiados la relación entre la variable respuesta y estas variables predictivas es positiva. De manera contraria, el coeficiente de estatus migratorio posee signo negativo, por lo que el tener estatus migratorio 2 o 3 disminuye las probabilidades de estar en el grupo de interés.

Para facilitar la interpretación de los resultados, se traducen términos de razón de probabilidad u *Odds Ratio* en la Tabla 4.

Tabla 4

Razón de probabilidades (Odds ratio) para los coeficientes de los modelos logísticos propuestos

Variable	Modelo 1	Modelo 2	Modelo 3
Género	1,4873	1,6562	2,0929
Índice económico	1,8461	2,028	2,6835
Nivel educativo madre 1	1,3416	1,2135	0,8326
Nivel educativo madre 2	1,5040	1,7765	1,0715
Nivel educativo madre 3	1,2305	1,4489	0,9702
Nivel educativo madre 4	1,2274	1,5153	0,831
Nivel educativo madre 5	1,1638	1,4331	1,002
Estatus migratorio 2	0,6703	0,5413	0,5744
Estatus migratorio 3	0,4895	0,5388	0,3647

La variable de índice económico muestra que ante un incremento de una unidad en la escala de índice económico los *odds* de estar en el nivel de 2 alfabetización se multiplican 1.84 veces (aumenta 84%), manteniendo todo lo demás constante. Este coeficiente toma una magnitud mayor para el nivel 3 de alfabetización en el que la razón de probabilidad se multiplica por 2.02 (sube 102%). Bajo la misma lógica, la razón de probabilidad del nivel 4 de alfabetización aumenta en 168,3% al incrementar el índice económico en una unidad.

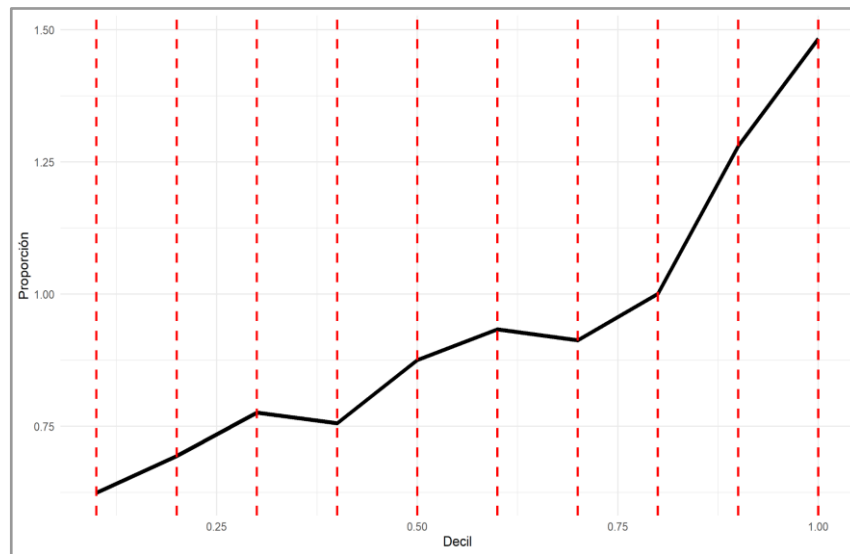
En cuanto a la educación de la madre, esta es significativa (con excepción del nivel 1) para la probabilidad de estar en el nivel de alfabetización 3. El que la madre del estudiante reporte tenencia de educación secundaria completa o vocacional (nivel 2), hace que los *odds* de pertenecer al nivel de alfabetización 2 sean 1.7765 veces más altos en comparación a los individuos cuya madre no cuenta con grados académicos, manteniendo constantes las demás variables. Los demás grados de escolaridad aumentan la razón de probabilidad, pero en menor cuantía al nivel 2.

La variable de género es significativa en los tres modelos, y arroja resultados interesantes. En términos de razón de probabilidad, se encuentra que para el modelo 1 la razón de probabilidad es 1.48 veces (48%) más alta para los hombres que para las mujeres. Para el modelo 2, esta razón de probabilidad es 1.6562 veces la de las mujeres. Por último, en el 5% de calificaciones más altas o en el nivel de alfabetización 4, la razón de probabilidad de los hombres es 2,09 veces la de las mujeres.

Lo anterior confirma la hipótesis inicial de que la variable de género e índice económico tomaría una relevancia mayor conforme se explora la probabilidad de colocarse en la cola derecha de la distribución. No obstante, no se cumple con el nivel educativo de la madre, mientras que, para la variable de estatus migratorio, la magnitud del efecto parece mantenerse en los tres modelos. Asimismo, este hallazgo respecto a la variable de género denota la brecha existente respecto a la distribución de hombres y mujeres en la cola derecha de la distribución de calificaciones, tal como se ilustra en la figura 2.

Figura 2

Gráfico de proporción de hombres respecto a mujeres por decil de calificación en matemática



Fuente: Elaboración propia con datos de PISA (2012).

Al observar el gráfico, se nota que por encima del octavo decil hay alrededor de 1,2 estudiantes masculinos por cada estudiante femenina, y para el 5% de calificaciones más altas, hay aproximadamente 1.8 estudiantes masculinos por cada estudiante mujer. Por el contrario, en el primer decil de calificaciones hay 1,6 mujeres por cada hombre, y en el segundo decil esta proporción cambia a 1,4 mujeres por cada hombre.

Para realizar el segundo objetivo planteado, se llevó a cabo el análisis de los modelos por medio de los efectos marginales promedio, los cuales se presentan en la tabla 5.

Tabla 5*Efectos marginales promedio de los modelos logísticos propuestos*

Variable	Modelo 1	Modelo 2	Modelo 3
Género	0,0851***	0,0644***	0.0309***
Índice económico	0,1305***	0,0898***	0.0420***
Nivel educativo madre 1	0,0623**	0,02543	-0,0075
Nivel educativo madre 2	0,0869***	0,0782**	0,0031
Nivel educativo madre 3	0,0444*	0,0492*	-0,0013
Nivel educativo madre 4	0,0441*	0,0571*	-0,0075
Nivel educativo madre 5	0,0328	0,0486*	0,0001
Estatus migratorio 2	-0,0822	-0,0659	-0,0192
Estatus migratorio 3	-0.1410**	-0,0663*	-0,0295

Códigos de significancia: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1

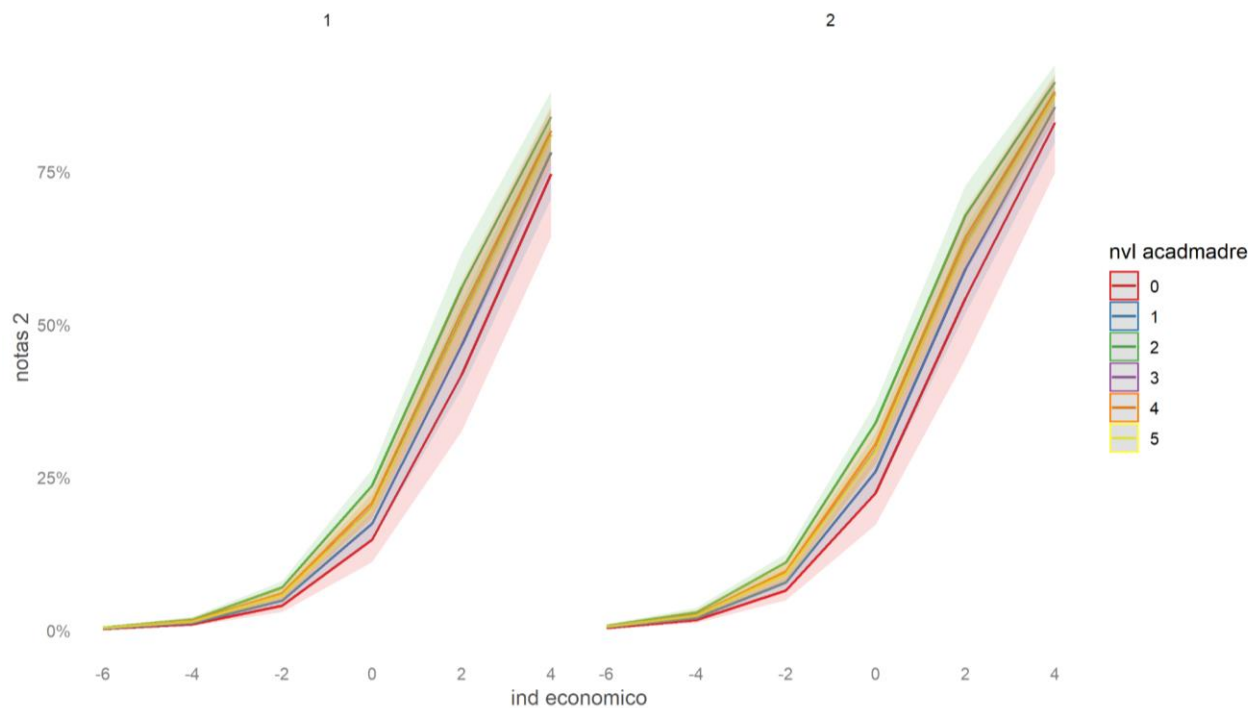
De acuerdo con los efectos marginales promedio, pasar de ser mujer a hombre ocasiona un aumento en la probabilidad de pertenecer al grupo de interés en 8.51 pp, 6.44 p.p y 3.09 p.p para el modelo 1, 2 y 3 respectivamente, manteniendo todo lo demás constante. El ser migrante de primera generación ocasiona una disminución en la probabilidad de pertenecer al nivel de alfabetización 2 en 14.1 p.p y de pertenecer al nivel de alfabetización 3 en 6.6. Para el modelo 3, esta variable no aporta información relevante ya que únicamente se cuenta con 1 persona no residente que forme parte del 5% de notas más alto. A pesar de que por este motivo el coeficiente no sea interesante de analizar, es importante tener en cuenta la falta de representatividad de grupos de trasfondos culturales diversos dentro de las calificaciones más altas en la región latinoamericana.

La figura 3 ilustra los efectos marginales promedio del modelo 2 para las calificaciones obtenidas por nivel académico según género. Los mismos denotan que los hombres obtienen calificaciones más altas para cada uno de los grados de escolaridad, además se encuentra que

conforme incrementa el nivel académico de la madre, también lo hace la calificación obtenida por los sujetos.

Figura 3

Gráfico de efectos marginales promedio del nivel educativo de la madre e índice económico sobre la probabilidad de pertenecer al nivel 3 de alfabetización matemática



Discusión

Esta investigación buscó el impacto de cuatro variables sobre la probabilidad de pertenecer a tres grupos determinados: nivel de alfabetización matemática 2, 3 y 4. En términos de percentiles, esto corresponde a los estudiantes que se colocan por encima del percentil 56, 83 y 95 en la distribución de calificaciones. Siendo estos últimos, los estudiantes de alto rendimiento dentro de la muestra seleccionada de estudiantes latinoamericanos.

Algunas de las limitaciones de este estudio se encuentran en la no significancia de la variable estatus migratorio dentro de los estudiantes de América Latina. Asimismo, para encontrar una explicación más robusta del diferencial de rendimiento, deberían incluirse otras variables como el tipo de colegio al que asiste el individuo, ansiedad ante las matemáticas, interés, autoeficacia y horas de estudio fuera de la institución educativa. Las mismas no pudieron ser incluidas dentro esta investigación por la gran incidencia de valores faltantes para los países de la región.

El estatus socioeconómico también resultó ser un predictor significativo para el nivel de alfabetización matemática de los estudiantes. Según Keegan (2005), este hallazgo va en la línea de muchas investigaciones desarrolladas en la región que han estudiado cómo la inseguridad económica lleva a los estudiantes a poca accesibilidad a recursos, un ambiente más inestable dentro del hogar y, por consiguiente, a desventajas en el ámbito educativo.

Se encontró la existencia de una brecha de género dentro de los estudiantes de alto rendimiento, de modo tal que los hombres tienen más del doble de probabilidad en comparación a las mujeres de estar en el 5% más alto. Este resultado alerta respecto a una desigualdad palpable dentro del sistema educativo latinoamericano. Debido a que las puntuaciones obtenidas en las pruebas estandarizadas definen el ingreso a una universidad, son determinantes del trabajo

de una persona y su éxito futuro. Si las mujeres reciben menos puntajes en los exámenes, entonces también reciben menos oportunidades.

Además, la brecha de género tiene implicaciones más profundas dentro de la elección de carrera profesional que seleccionan las mujeres. De acuerdo con Zhou et al., (2017), existe evidencia de que las mujeres que eligen carreras de tipo STEM, tienen notas altas en las pruebas estandarizadas de matemática. La falta de participación de mujeres en campos STEM amenaza el bienestar de las mujeres, sus familias y la fortaleza de la economía. Si la sub-representación está realmente ligada a la brecha de género en rendimiento matemático de pruebas estandarizadas, se debe comenzar por estudiar los programas de educación vigentes y promover una educación más igualitaria, con estándares de calidad y accesible para todos.

Referencias

- Alderete, A. (2006). Fundamentos del Análisis de Regresión Logística en la Investigación Psicológica. *Evaluar*, (6), 52-67.
- Andrew Gelman and Yu-Sung Su (2020). arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. R package version 1.11-2. <https://CRAN.R-project.org/package=arm>
- Castro, G., Gimenez, G. & Pérez, Domingo. (2017). Educational Inequalities in Latin America, 2012 PISA: Causes of Differences in School Performance between Public and Private Schools (English version). *Revista de Educación*. 376, 32-59.
- Chipman, S., Krantz, D. & Silver, R. (1992). Mathematics Anxiety and Science Careers among Able College Women. *Sage Journal*, 3 (5), 292-296.
- Dustmann, C., Frattini, T. & Lanzara, G. (2014). Educational achievement of second-generation immigrants: an international comparison. *Economic Policy*. 27(69), 143-185.
- Keegan, M. (2005). Social-Demographic, School, Neighborhood, and Parenting Influences on the Academic Achievement of Latino Young Adolescents. *J Youth Adolescence* 34, 163–174.
- Feinstein, L., Duckworth, K. & Sabates, R. (2008). Education and the Family: Passing Success Across the Generations. <https://www.routledge.com/Education-and-the-Family-Passing-Success-Across-the-Generations/Feinstein-Duckworth-Sabates/p/book/9780415396370>
- Freeman, J. (1991). Mentoring gifted pupils: an international view. *Educating Able Children*, 5, 6-12.

- Furnham, A., Reeves, E. & Budhani, S. (2002). Parents Think Their Sons Are Brighter Than Their Daughters: Sex Differences in Parental Self-Estimations and Estimations of Their Children's Multiple Intelligences. *The Journal of Genetic Psychology*, 163(1), 24–39.
- González, A & De la Rica, S. (2012). Gender Gaps in PISA Test Scores: The Impact of Social Norms and the Mother's Transmission of Role Attitudes. IZA Institute for the Study of Labor Discussion Paper Series, 34.
- Gurría, A. (28 de febrero de 2014). Global Forum on Competition 2014 - Remarks by Angel Gurría, OECD Secretary General. <https://www.oecd.org/competition/sg-speech-global-forum-on-competition-2014.htm>
- Hadley Wickham and Evan Miller (2021). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.4.1. <https://CRAN.R-project.org/package=haven>
- Husin, M. (2014). Assessing Mathematical Competence in Second Language: Exploring DIF Evidences from PISA Malaysian Data. *Theses and Dissertations*. 695, 1-44.
- León, J & Youn, M.-J. (2016). El efecto de los procesos escolares en el rendimiento en Matemática y las brechas de rendimiento debido a diferencias socioeconómicas de los estudiantes peruanos. *Revista Peruana De Investigación Educativa*, 8(8), 149–180.
- Liu, L & Wilson, M. (2009). Gender Differences in Large-Scale Math Assessments: PISA Trend 2000 and 2003. *Applied Measurement in Education*, 22(2), 164–184.
- Lüdecke D (2021). *_sjPlot: Data Visualization for Statistics in Social Science_*. R package version 2.8.8, <URL: <https://CRAN.R-project.org/package=sjPlot>>.

Lüdecke, Patil, Ben-Shachar, Wiernik, Waggoner & Makowski (2020). Visualisation Toolbox for 'easystats' and Extra Geoms, Themes and Color Palettes for 'ggplot2'. CRAN.

Available from <https://easystats.github.io/see/>

OECD. (31 de agosto de 2017). PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics, Financial Literacy and Collaborative Problem Solving.

[https://www.oecd.org/pisa/aboutpisa/ebook%20-%20PISA-](https://www.oecd.org/pisa/aboutpisa/ebook%20-%20PISA-D%20Framework_PRELIMINARY%20version_SPANISH.pdf)

[D%20Framework_PRELIMINARY%20version_SPANISH.pdf](https://www.oecd.org/pisa/aboutpisa/ebook%20-%20PISA-D%20Framework_PRELIMINARY%20version_SPANISH.pdf)

Organisation for Economic Co-operation and Development., & Programme for International Student Assessment. (2014). PISA 2012technical report. Paris: OECD.

Räty, H., Vänskä, J., Kasanen, K. & Kärkkäinen, R. (2002). Parents' explanations of their child's performance in mathematics and reading: a replication and extension of Yee and Eccles.

Sex Roles: A Journal of Research, 46 (3–4), 121–128.

Rodríguez, D & Guzmán, R. (2019). Rendimiento académico y factores sociofamiliares de riesgo. Variables personales que moderan su influencia. Perfiles educativos, 41(164), 118-134.

Smith, Jay., Kenneth, S. & Zhanna, S. (2008). Factors Inhibiting Hispanic Parents' School Involvement. ERIC, 29(2), 8-13.

Thomas J. Leeper (2021). margins: Marginal Effects for Model Objects. R package version 0.3.26.

Trenton D. Mize. (2019). Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects. Sociological Science, 6, 81-117.

- Vitola, L. (2015). Regresión Logística: una aplicación en la identificación de variables que inciden en el rendimiento académico, en el área de matemáticas. *Revista Educación y Desarrollo Social*, 9(1), 118-131.
- Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- Wooldridge, J. M. (2010). INTRODUCCIÓN A LA ECONOMETRÍA: UN ENFOQUE MODERNO (4.a ed.). MEXICO: CENGAGE LEARNING.
- Zhou, Y., Fan, X., Wei, X. & Tai, R. (2017). Gender Gap Among High Achievers in Math and Implications for STEM Pipeline. *The Asia-Pacific Edu Res*, 26.

Anexos

Tabla 1.a

Estadísticos descriptivos de las variables utilizados en los modelos de regresión logística

Variable	Nombre asignado	Valor máximo	Valor mínimo	Observaciones perdidas
PV1MATH	notas.1			
	notas.2	703.54	59.28	0
	notas.2			
escs	índice económico	2.506	-4.93	101
fiscd	nvl.acadpadre	0	6	193
innmig	“status.migra”	1	3	300
ST04Q01	género	1	2	0

Fuente: Elaboración propia con datos de PISA 2012