PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS SOCIALES

**Entregable Nº3**

Curso: Estadística para el Análisis Político 2

Clave: POL 304

Horario: 689A

Integrantes:

Romina Ariana Loayza Gaitán (20211323)

Yhara Alexandra Laime Aguilar (20203872)

**Tema de investigación:** Análisis de los factores sociales, económicos e infraestructurales que determinan la práctica de la lectura de niños y adolescentes.

**Justificación del tema:**

La lectura es esencial para el desarrollo integral del ser humano, teniendo un impacto significativo en su desarrollo intelectual, social, espiritual y moral. Además, la literatura enriquece el vocabulario, influencia en la formación del carácter y despierta el interés por el conocimiento. La enseñanza de la lectura y la literatura se considera un objetivo primordial en la educación, con un énfasis en el desarrollo de habilidades y hábitos de lectura desde edades tempranas. El acceso a bibliografías de calidad y la promoción de la lectura desde la infancia son aspectos clave para fomentar el hábito de leer. A pesar de las mejoras en la disponibilidad de recursos, persisten desafíos como la falta de interés por parte de algunos estudiantes, lo que resalta la importancia del trabajo sistemático por parte de maestros, profesores y bibliotecarios para cultivar el amor por la lectura (Domínguez et al., 2015, p.5-6). En este sentido, el análisis de los determinantes socioeconómicos e infraestructurales en la práctica lectora de niños y adolescentes es fundamental para comprender las variables significativas involucradas en esta problemática. En consecuencia, al determinar dichos factores, se podrá trabajar en propuestas que contribuyan a combatir las problemáticas.

**Justificación de las variables indicadas como predictores:**

En primer lugar, los factores externos incluyen elementos del contexto cultural y académico en el que se encuentran los estudiantes, como el entorno social, la situación económica, y las estrategias metodológicas utilizadas en la enseñanza. Se destaca que estos factores externos, también conocidos como factores exógenos, están más allá del control individual de los estudiantes y pueden influir significativamente en su rendimiento en la lectura (Centeno, 2021, p.9).

Dentro de estos factores externos, que se encuentran más allá del control de los estudiantes, también se encuentran las mismas condiciones del hogar. Es en este sentido que, si un estudiante cuenta en el hogar con las condiciones óptimas para poder practicar la lectura, esto desembocará en la práctica del hábito mismo. El acceso a internet, agua y electricidad podría influir en el hábito de lectura. La disponibilidad de internet proporciona a los estudiantes acceso a recursos educativos y literarios. La electricidad es esencial para la iluminación adecuada y para el uso de dispositivos electrónicos que contienen libros digitales. El acceso a agua corriente asegura un entorno de vida saludable y cómodo, lo cual es fundamental para la concentración y el tiempo dedicado a la lectura. Además, la cantidad de personas en la vivienda puede afectar el ambiente de estudio; en hogares con menos personas, es más probable que haya un espacio tranquilo y adecuado para la lectura. Por lo tanto, estos factores socioeconómicos y ambientales son cruciales para fomentar el desarrollo de hábitos de lectura en los estudiantes.

Es así también que, la escuela desempeña un papel fundamental en motivar el interés por la lectura y desarrollar el hábito de leer, con la infraestructura siendo crucial para este fin. Sin embargo, es esencial que los profesores comprendan adecuadamente la importancia de la lectura para lograr resultados efectivos. Así también, la familia juega un papel importante en la formación de los intereses y hábitos de lectura de los estudiantes. Cuando los padres y otros miembros de la familia muestran interés por la lectura, los hijos tienden a imitar ese comportamiento. El clima cultural en el hogar, la infraestructura de las escuelas y la disponibilidad de libros también afectan la formación de los estudiantes (Domínguez et al., 2015, p.7).

Por ende, para este trabajo se tomarán en cuenta las variables: tasa de acceso a agua, tasa de electricidad por cableado, tasa de acceso a internet, tasa de acceso a internet por medio del celular, cantidad de personas por vivienda y el análisis de calidad de infraestructura educativa, calidad del equipo de la institución educativa, calidad de enseñanza, calidad de materiales educativos y calidad de la participación de los padres en la educación. Estas últimas variables relacionadas al aspecto de la calidad, son las percepciones que se tienen en la Encuesta Nacional de Lectura.

**Justificación de la variable de control:**

La pertinencia de la variable de control “Población peruana de 6 a 17 años que se encuentren cursando un grado de estudio” en este trabajo es relevante. Los estudiantes que están cursando un grado de estudio están en contacto directo con los docentes, quienes, como orientadores, despiertan el interés por la lectura en los niños; ya que sugieren y permiten que los alumnos exploren y manipulen la lectura (Sobrino, 2019). Al controlar esta variable, podemos examinar más de cerca el efecto de la enseñanza y la orientación del docente en la práctica lectora de los estudiantes. Esto subraya la relevancia de esta variable de control para el estudio.

**Pregunta de investigación:**

¿Cuáles son los factores que influyen en el desarrollo del hábito de lectura en los niños y adolescentes entre los 6 a 17 años?

**Hipótesis:**

* Primera hipótesis: la práctica lectora en niños y adolescentes es causada por el acceso a Internet y por el acceso a recursos físicos educativos de calidad.
* Segunda hipótesis: la práctica lectora en niños y adolescentes es causada por el acceso a Internet, por el acceso a recursos físicos educativos de calidad, el acceso a Internet desde un celular móvil y que la vivienda cuente con electricidad.
* Tercera hipótesis: la práctica lectora en niños y adolescentes es causada por el acceso a Internet, el acceso a Internet desde un celular móvil y que la vivienda cuente con electricidad.

**Metodología**

La variable central sobre el hábito de lectura es una tasa que está calculada por cada 3 niños, por lo que el valor muestra la cantidad que sí leen por cada tres niños. Respecto a la operacionalización de las variables independientes, se encuentran las tasas de agua, electricidad, acceso a internet y acceso a internet por medio de celular. Todas estas tasas están calculadas por vivienda. El valor de la tasa de agua muestra la cantidad de viviendas por provincia que no tienen acceso a agua los siete días de la semana, la tasa de electricidad muestra la cantidad de viviendas que no tienen acceso a electricidad por medio de cableado, la tasa de internet muestra las viviendas que no tienen acceso a internet en su hogar y la tasa de acceso a internet por medio del celular muestra la cantidad de viviendas que no tienen acceso a internet por medio de celular. En cuanto a la operacionalización de las variables de calidad educativa, las 5 percepciones van de un valor del 0 al 1, donde 1 sería mejor calidad educativa en determinado aspecto mientras que 0 sería peor percepción de calidad educativa.

**Limpieza de la base de datos**

Las variables para este estudio provienen de diversas fuentes, todas ellas proporcionadas por el Instituto Nacional de Estadística e Informática (INEI), en la web de Microdatos. La variable dependiente se extrae exclusivamente de la Encuesta Nacional de Lectura del año 2022. Esta encuesta proporciona una visión integral y actualizada de los hábitos de lectura en el país.

En cuanto a las variables independientes, se utilizan dos conjuntos de datos principales. El primero es la última data mencionada proporcionada por el INEI, que ofrece indicadores socioeconómicos. El segundo es la Encuesta Nacional de Hogares (ENAHO) del año 2022, que presenta los resultados anuales de diversos aspectos de la vida de los hogares en Perú. Ambas fuentes de datos son de acceso público y se pueden encontrar en el portal web oficial del INEI.

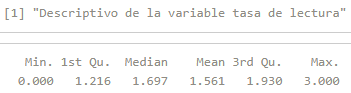
Respecto a las datas obtenidas de la Encuesta Nacional de Lectura, la variable dependiente estaba operacionalizada en valores del 1 al 9, donde 1 significaba ‘Diariamente, 2 ‘Varias veces a la semana’, 3 ‘Una vez a la semana’, 4 ‘Varias veces al mes’, 5 ‘Una vez al mes’, 6 ‘Una vez cada tres meses’, 7 ‘ Por lo menos una vez al año’, 8 ‘No leyó’ y 9 ‘No sabe-No responde’. Para poder obtener los valores que deseamos de esta variable dependiente, se filtró la base de datos por la variable ‘EDAD’ y de esa manera obtener los valores de ese grupo etario. Luego de esto, se modificó la dependiente a valores de 1 y 0, en donde los valores previos de 1 y 2 pasarían a ser 1 y el resto se volvería 0. Esto con el fin de identificar más fácilmente quienes practican el habito de lectura. Posteriormente, se juntaron los valores por ‘UBIGEO’ y sumando la cantidad de valores ‘1’ y la cantidad de valores ’0’ en dos columna diferentes que se llamarían ‘leen\_sí’ y ‘leen\_no’ respectivamente; además de añadir una columna de ‘leen\_total’ para saber la cantidad de niños y adolescentes por distrito. Para proceder con el análisis por provincia se sumaron las filas y se agruparon por provincia, para luego obtener la tasa de niños y adolescentes que sí practican el hábito de la lectura por cada 3 habitantes.

Luego, se obtuvo la variable independiente de número de personas por hogar (PERS\_NRO), en donde los valores iban del 1 al 18, por lo que al juntar por Provincia se obtuvo una media de los valores. Respecto de las variables sobre electricidad y agua, ambas se extraen de la ENAHO. La variable de agua iba del 1 al 2 y significan sí y no respectivamente; mientras que la variable de electricidad va del 0 al 1 y significan no y sí respectivamente. Ambas variables se modificaron a que 1 valga ‘Sí’ y 0 valga ‘No’. Esto para poder obtener la cantidad de hogares que tienen acceso a agua y electricidad por UBIGEO, para luego juntarlo por provincia y obtener las variables: ‘agua\_sí’, ‘agua\_no’, ‘agua\_total’, ‘elec\_sí’, ‘elec\_no’ y ‘elec\_total’. Luego de obtenidas estas variables se podrá sacar la tasa de hogares que no tienen acceso a agua y electricidad por cada 10 hogares. Este mismo procedimiento se siguió con las variables de acceso a internet en el hogar y acceso a internet mediante celular, en donde se obtuvieron los valores de ‘tasa\_internet’ y ‘tasa\_celinter’ por cada 3 hogares.

Por último, las variables respecto a la calidad educativa se extrajeron de la ENAHO, en donde tomaban los valores del 1 al 5, significando ‘Muy malo', ‘Malo’, ‘Bueno’, ‘Muy bueno' y ‘No sabe’ respectivamente. Se modifica la variable para que tome los valores de 0 y 1, donde 3 y 4 se volverían 1; mientras que 1, 2 y 5 se volverían 0. Esto para posteriormente poder armar un índice. Luego, se agruparon los datos por provincia y se modoficaron las variables para poder obtener un promedio de cada una. Una vez obtenidos los valores se procedió al análisis factorial en donde nos dio como resultado dos índices, uno que contiene la percepción de la calidad de infraestructura (cedu\_infra) y la percepción de la calidad de los equipos (cedu\_equipo). El otro índice obtenido está compuesto por la percepción de la calidad de participación de los padres (cedu\_partpadres), percepción de la calidad de enseñanza (cedu\_ensenanza) y la percepción de la calidad de los materiales educativos (cedu\_material).

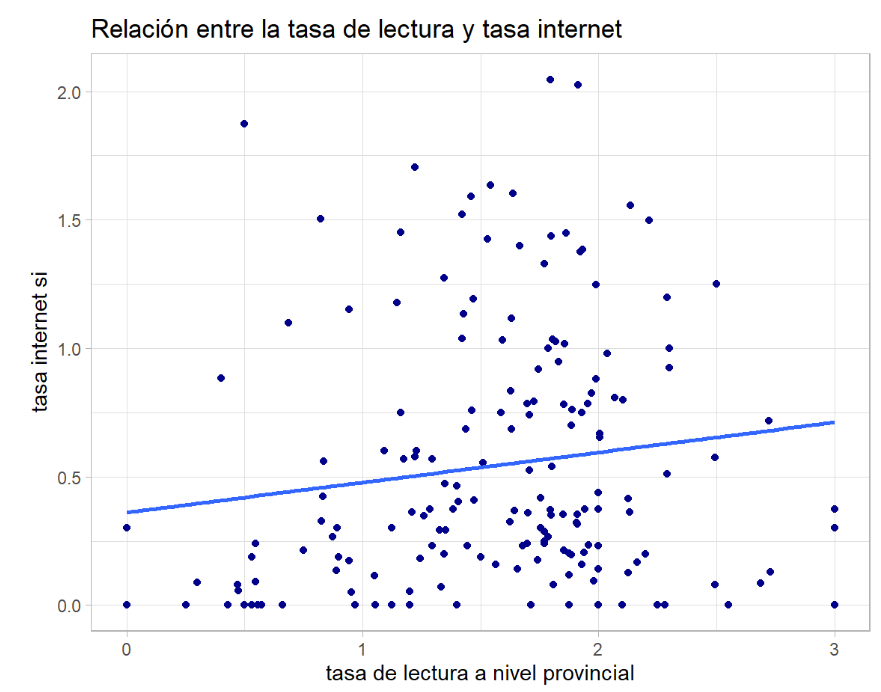
**Hallazgos:**

1. *Variable central*

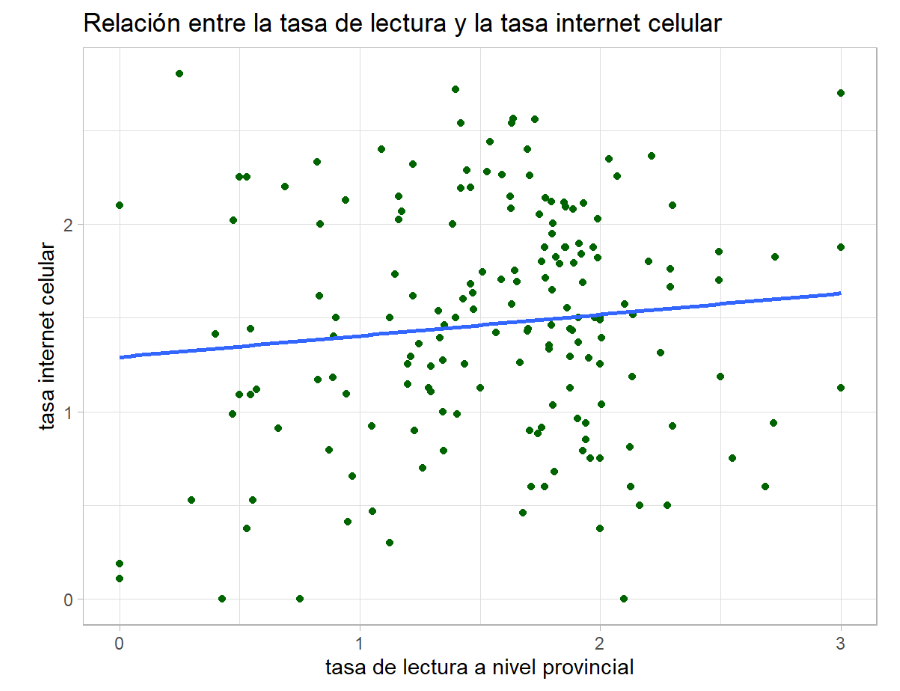


La tabla descriptiva de la variable "tasa de lectura" muestra un valor mínimo de 0.000 y un valor máximo de 3.000 entre los niños que practican el hábito de la lectura. El valor promedio de la tasa de lectura es 1.561. En general, estos datos sugieren que, existe una variabilidad considerable en la frecuencia con que se practica el hábito de lectura.

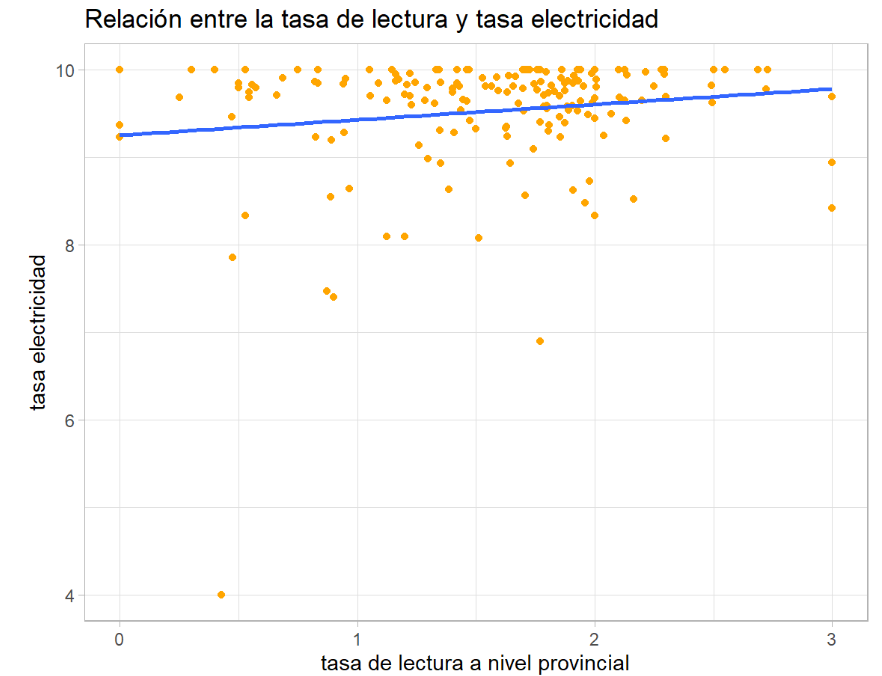
1. *Correlaciones univariadas con la variable central:*



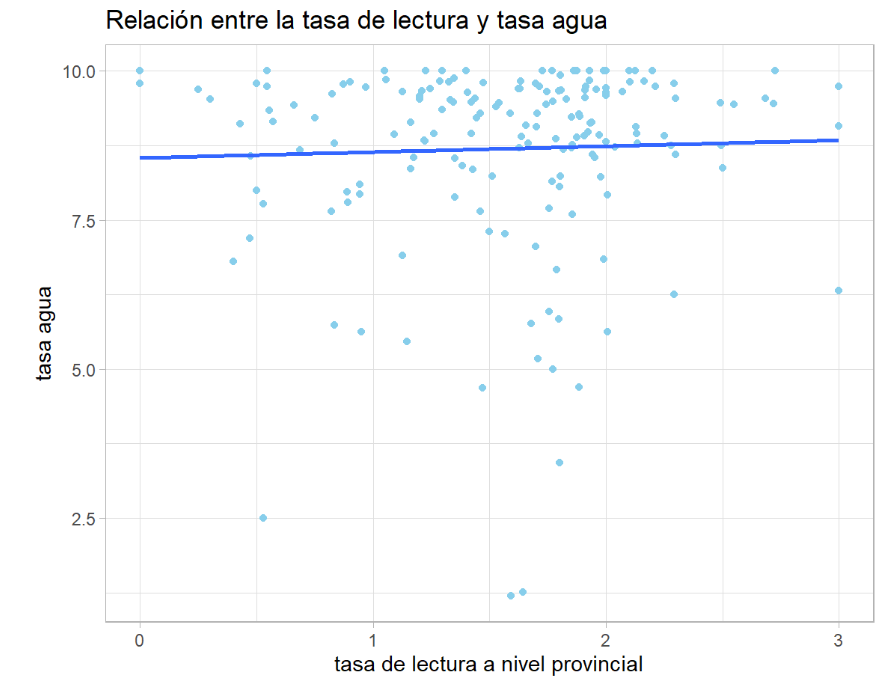
Esta variable representa la tasa de viviendas con acceso a Internet. Tiene una relación positiva y débil con la práctica de lectura, con un coeficiente de Pearson de 0.140.



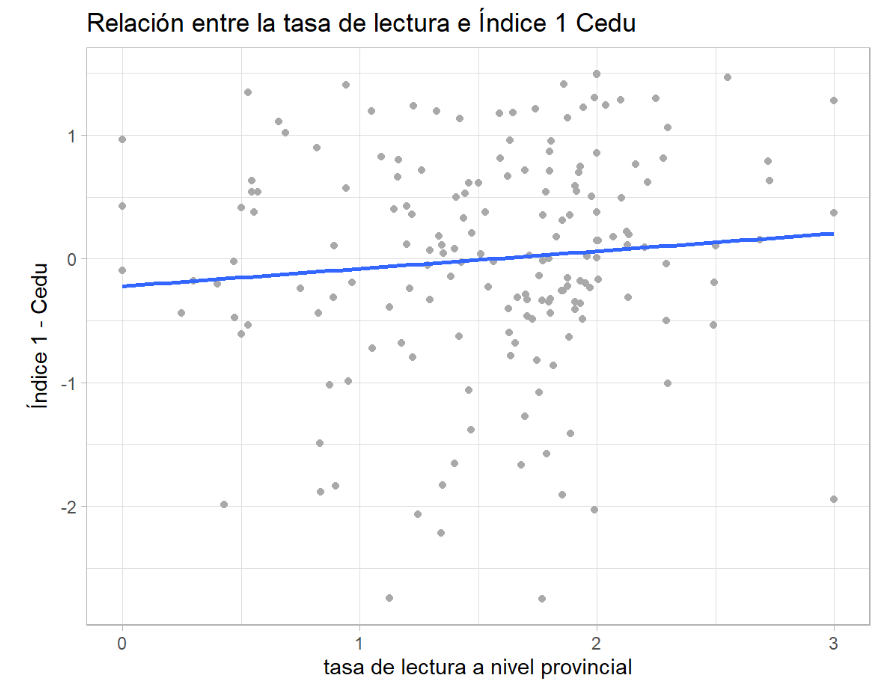
Esta variable indica la tasa de viviendas con acceso a Internet desde un teléfono móvil. Posee una relación positiva y débil, con un coeficiente de Pearson de 0.108.



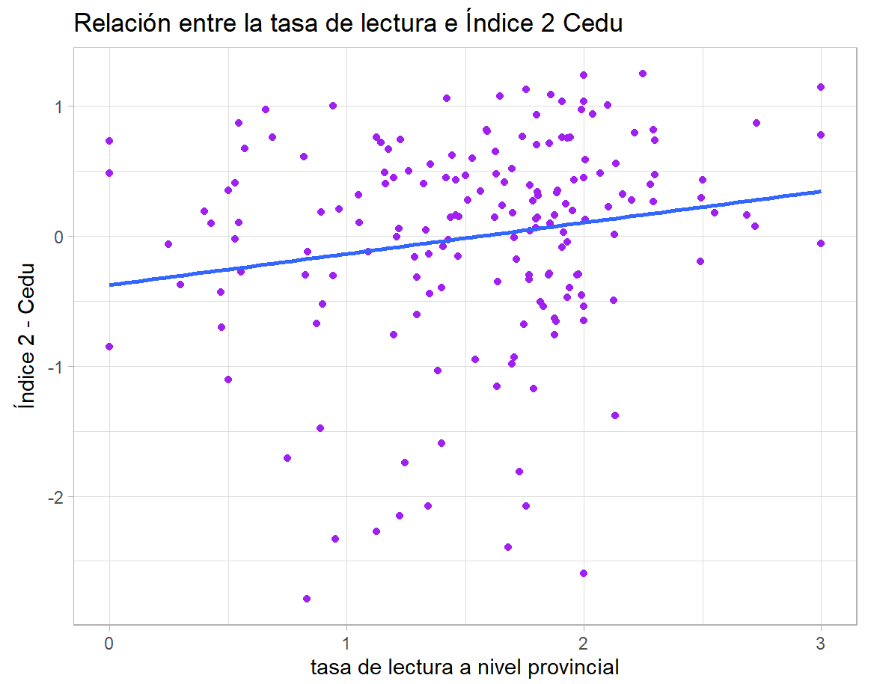
Esta variable muestra la tasa de viviendas con acceso a electricidad. Tiene una relación positiva, aunque débil, con un coeficiente de Pearson de 0.151. A pesar de ser baja, se incluirá en el análisis.



Esta variable se refiere a la tasa de viviendas con acceso a agua potable desde una red pública. La relación es positiva pero muy débil, con un coeficiente de Pearson de 0.037, por lo que no se utilizará en los análisis posteriores.

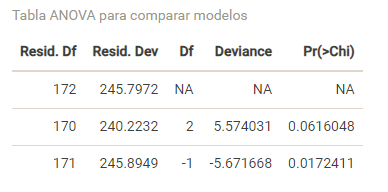
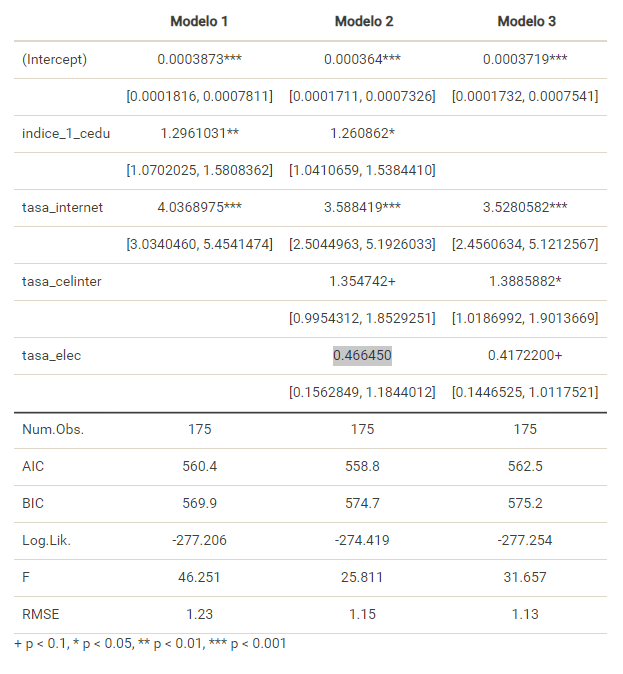


Se observa una relación positiva entre esta variable y la práctica de lectura, con un coeficiente de Pearson de 0.123. Aunque la relación es débil, es más fuerte que la del Índice 2, por lo que se utilizará en el análisis.



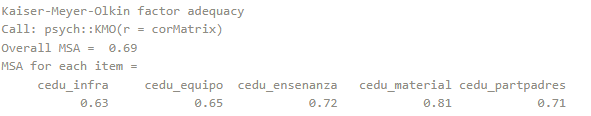
La relación entre esta variable y la práctica de lectura es positiva pero débil, con un coeficiente de Pearson de 0.099.

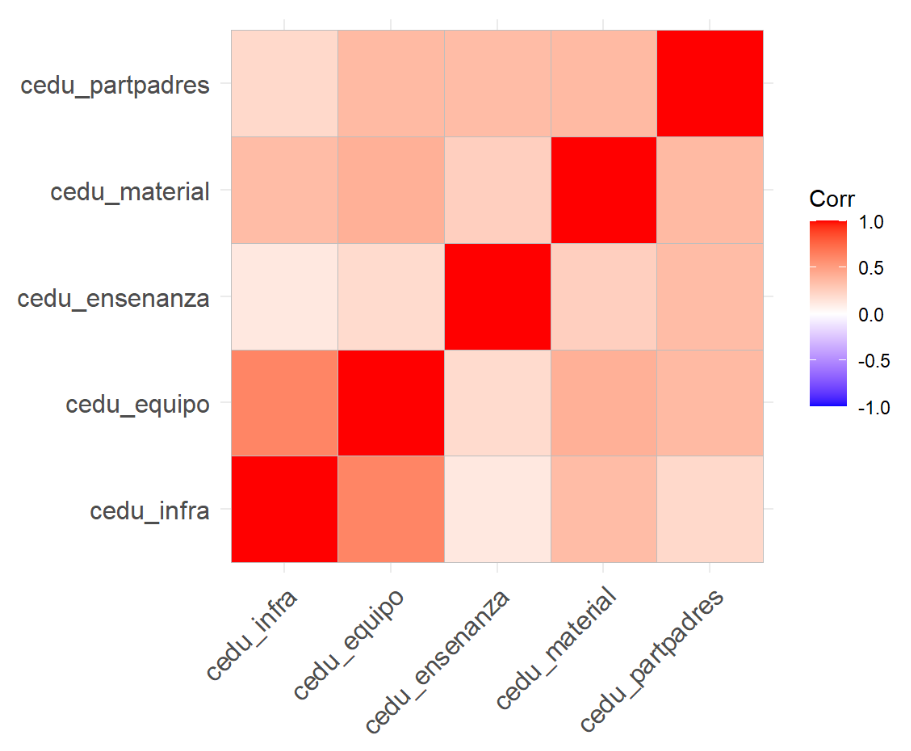
1. *Modelos de regresión: Poisson*

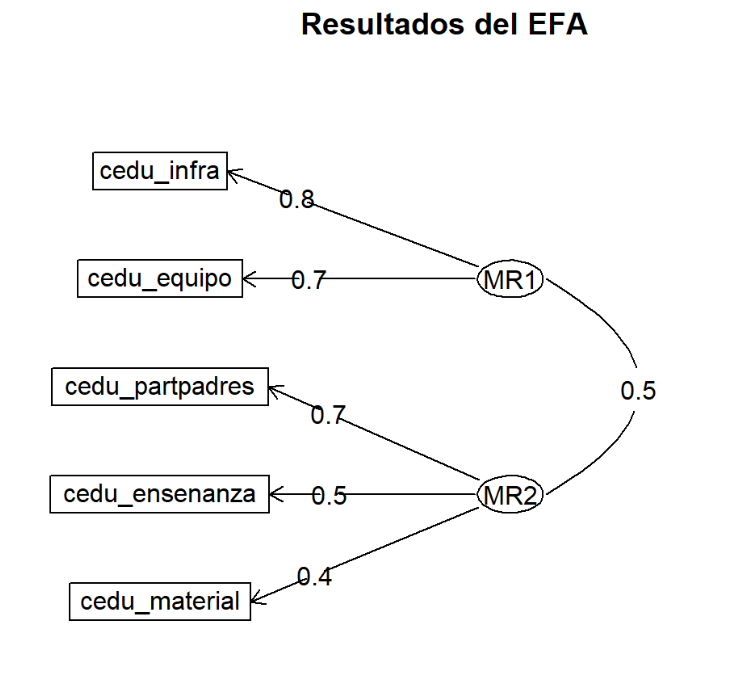


La regresión elegida es la de Poisson, ya que nuestras variables están medidas en tasas, lo que la convierte en la opción indicada. Se inició con el MODELO 1, utilizando las variables “índice\_1\_cedu” y “tasa\_internet”. En este modelo, se identificó la significancia de ambas variables. Por lo tanto, se procedió a agregar las variables restantes: “tasa\_internet\_celular” y “tasa\_electricidad”. Sin embargo, “índice\_1\_cedu” perdió significancia, mientras que las variables agregadas no resultaron significativas. Se decidió quitar la variable “índice\_1\_cedu” para el tercer modelo, y este mejoró.

Luego de proceder a la comparación de los 3 modelos tanto por AIC y por una tabla de ANOVA usando el Chi-Cuadrado, se concluye que el mejor modelo es el segundo, ya que posee un menor AIC y menor Resid. Dev.

1. Análisis Factorial





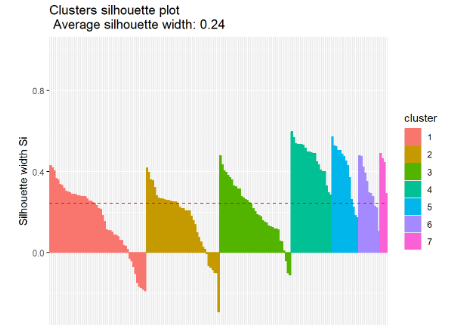
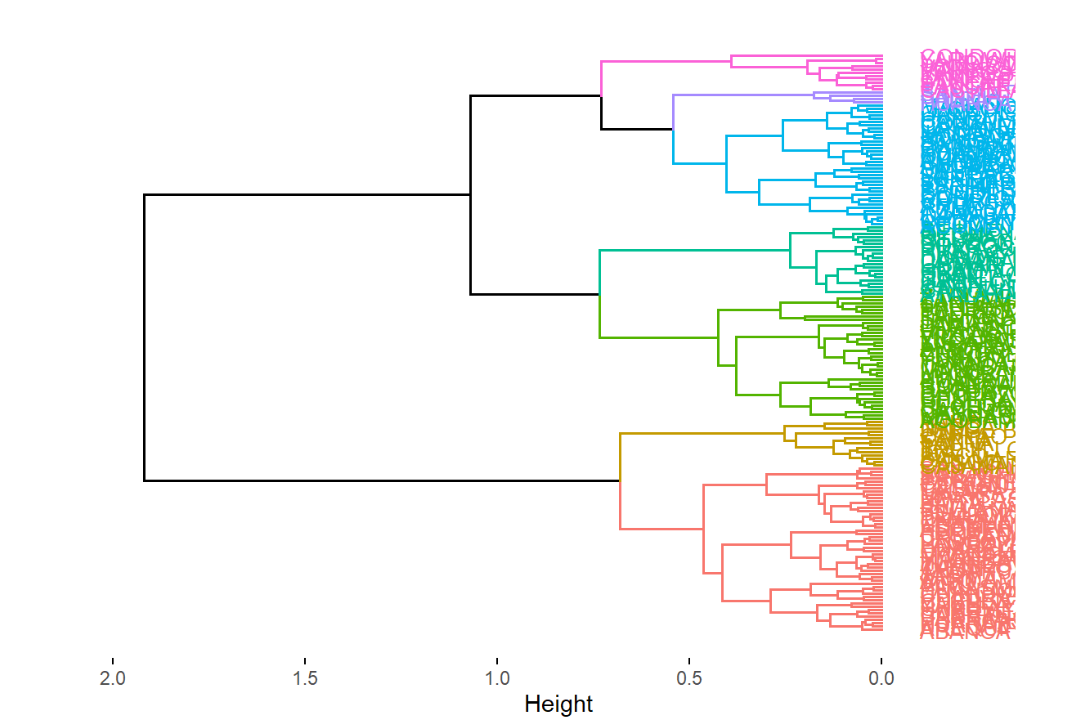
Previo al análisis de nuestras variables independientes, se llevó a cabo un Análisis Factorial Exploratorio basado en las variables CEDU. Los resultados mostraron correlaciones positivas moderadamente adecuadas, con un coeficiente KMO de 0.67. Además, las pruebas para determinar si la matriz era una identidad o singular resultaron negativas, lo que permitió continuar con el análisis factorial.

La función fa.parallel sugirió la existencia de 2 variables latentes. En el gráfico de “Resultado de EFA” se observan claramente dos factores: MR1 y MR2.

El MR1 incluye “cedu\_infra” (calidad educativa de la infraestructura) y “cedu\_equipo” (calidad educativa de los equipos en las instituciones educativas), ambos relacionados con los recursos físicos educativos. Estos formarán el “índice\_1\_cedu”.

Por otro lado, el MR2 engloba “cedu\_partpadres” (participación de los padres), “cedu\_enseñanza” (calidad de la enseñanza) y “cedu\_material” (materiales educativos), factores que tienen un enfoque más individualizado al entorno personal del estudiante. Estos constituirán el “índice\_2\_cedu”.

1. Clusterización:

****

Al realizar el proceso de clusterización y análisis por conglomerados, se observó que el mejor modelo resultó ser AGNES, con un Silhouette de 0.24. Aunque este valor es bajo, es el mejor y nos muestra una clara agrupación. El eje ‘Height’ nos muestra que el “costo” de conglomerar fue de 2.0. Además, el cuarto grupo es el mejor agrupado. Por último, estas son las provincias mal clusterizadas: AREQUIPA, CANDARAVE, CARABAYA, CHUPACA, FERREÑAFE, HUARAZ, HUAYLAS, JULCAN, LAMBAYEQUE, LUCANAS, MARISCAL NIETO, OXAPAMPA, PICOTA, SAN MIGUEL, SAN ROMAN, TARATA, VICTOR FAJARDO, VILCAS HUAMAN y VIRU.

**Conclusiones:**

En base a los resultados obtenidos en la regresión, se concluye que las variables ‘tasa\_internet’, ‘índice\_1\_cedu’ y ‘tasa\_celinter’ son significativas para el análisis de la variable central. Por tanto, por cada unidad que aumenta ‘tasa\_internet’, la cantidad esperada para la cantidad de niños y adolescentes que practican el hábito de lectura aumenta en 2.58 unidades. Asimismo, por cada unidad que aumente el Índice 1, relacionado al ambiente del hogar, se espera que la cantidad de niños y adolescentes que practican el hábito de lectura aumente en 0.26 unidades. Por último, por cada unidad que la tasa de viviendas que cuentan con acceso a internet por medio del celular, la cantidad de niños y adolescentes que practican el hábito de lectura aumenta en 0.35 unidades. Asimismo, se puede evidenciar cómo el ambiente del hogar, que se ve reflejado en el ‘índice\_1\_edu’ tiene influencia respecto a la práctica del hábito de lectura.

**Bibliografía**

Centeno, A. F., & Silva, S. S. (2021). Prácticas de lectura y escritura desde la literacidad. *Diálogos sobre educación. Temas actuales en investigación educativa*, *12*(23).

Domínguez, I. D., Delgado, L. R., Ávila, Y. T., & Ávila, M. M. R. (2015). Importancia de la lectura y la formación del hábito de leer en la formación inicial. *Estudios del desarrollo social: Cuba y América Latina*, *3*(1), 94-102.

Encuesta Nacional de Lectura - ENL (2022). *Características de la vivienda y del hogar, actividades de fomento de la lectura en el hogar y TI*C. Recuperado de la base de datos del Instituto Nacional de Estadística e Informática.

<https://proyectos.inei.gob.pe/microdatos/Detalle_Encuesta.asp?CU=19558&CodEncuesta=782&CodModulo=1722&NombreEncuesta=Encuesta+Nacional+de+Lectura+-+ENL&NombreModulo=CARACTER%C3%8DSTICAS+DE+LA+VIVIENDA+Y+DEL+HOGAR,+ACTIVIDADES+DE+FOMENTO+DE+LA+LECTURA+EN+EL+HOGAR+Y+TIC>

Encuesta Nacional de Lectura - ENL (2022). *Prácticas lectoras de los residentes habituales de 0 a 17 años de edad*. Recuperado de la base de datos del Instituto Nacional de Estadística e Informática.

<https://proyectos.inei.gob.pe/microdatos/Detalle_Encuesta.asp?CU=19558&CodEncuesta=782&CodModulo=1725&NombreEncuesta=Encuesta+Nacional+de+Lectura+-+ENL&NombreModulo=PR%C3%81CTICAS+LECTORAS+DE+LOS+RESIDENTES+HABITUALES+DE+0+A+17+A%C3%91OS+DE+EDAD>

Ortiz-Salazar, M. A., & Peña Castaño, J. M. (2019). La lectura en la infancia y niñez: incidencia en la construcción del sujeto lector. *Sophia*, *15*(2), 111-117.

Sobrino Cortés, B (2019). Estrategias didácticas en docentes para promover la lectura. Revista Vinculando, 17(2). [Estrategias didácticas en docentes para promover la lectura (vinculando.org)](https://vinculando.org/educacion/estrategias-didacticas-en-docentes-para-promover-la-lectura.html)