

Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Monterrey



Modelación Estadística para
la Toma de Decisiones (Gpo 201)

Nombre de la profesora:
Viridiana Itzel Méndez Vásquez

Nombre de la profesora cotitular:
María Guadalupe Tobías Lara

Informe Final

Análisis Holístico sobre la Lectura en México

Emilio Alejandro González Huerta — A01286440
Laurie Camila Hernández Pacheco — A01286569
Jacqueline Peña San Juan — A00842524
Hugo Edel Gamboa Sesma — A00841605
Pablo Alonso Robles Nava — A00573454

Fecha: 25/10/2025

CONTENTS

I. Resumen	2
II. Introducción	3
A. Objetivo	3
B. Justificación	3
III. Metodología	4
A. Variables Seleccionadas	4
B. Fuente de Datos	5
C. Sección Descriptiva	5
D. Sección Inferencial	5
IV. Análisis Descriptivo e Inferencial	6
A. Medidas descriptivas	6
B. Análisis Inferencial	8
1. Intervalos de Confianza	8
2. Prueba Chi-Cuadrada	8
3. Regresión Logística Binomial	8
4. Prueba de Fisher	9
V. Resultados	10
VI. Discusión y Conclusiones	10
VII. Anexos	12
A. Liga del RMD	12
B. Gráficos	12
C. Tablas	12
VIII. Referencias	13

I. RESUMEN

La carencia del hábito lector se reconoce como un problema complejo acentuado por factores estructurales. Poniendo en contexto a la situación de un mexicano promedio, la desigualdad socioeconómica impacta el acceso y la calidad educativa, lo que a su vez se relaciona con menores prácticas de lectura desde la edad temprana, respaldado por (Ibarra et al., 2025) y (Castro, s.f.). La problemática se agrava por la marginación de comunidades indígenas que no hablan español y por la deserción escolar. Asimismo, el uso de redes sociales e internet, aunque ofrece acceso a la información, contribuye a disminuir la capacidad de atención y el pensamiento crítico, afectando el tiempo dedicado a la lectura.

Este proyecto de investigación se enfoca en analizar la encuesta MOLEC (Módulo sobre Lectura) del INEGI en México, implementada desde 2015, con el fin de interpretar los datos superficiales y comprender la carencia del hábito de lectura en la población mayor de 18 años. Dicho lo anterior, el objetivo es identificar la raíz del problema, así como examinar qué factores influyen en el hábito y las barreras para leer. Se busca integrar

el análisis de MOLEC con fuentes bibliográficas sobre las causas profundas del mal hábito lector con el fin de formar una sugerencia o planteamiento de hallazgos para permitir el desarrollo de futuras políticas públicas más sostenibles que fomenten la acción de la lectura en distintos sectores de la sociedad mexicana.

Las variables analizadas en esta investigación provienen de la encuesta MOLEC y se dividen en dos grandes grupos: sociodemográficas y relacionadas con los hábitos de lectura. Entre las primeras se incluyen la edad, el nivel educativo y la condición de actividad u ocupación, las cuales permiten contextualizar los comportamientos según el perfil de los participantes. En cuanto a las variables de lectura, se consideraron la frecuencia con la que las personas leen libros, revistas o medios digitales, los motivos principales de lectura (como trabajo, estudio, cultura general o entretenimiento) y los tipos de contenido que consumen (informativo, práctico o de entretenimiento).

Para abordar esta problemática, se aplicaron técnicas de análisis descriptivo e inferencial utilizando los datos del MOLEC correspondientes a los años 2020, 2022 y 2024. El análisis descriptivo permitió caracterizar las variables mediante medidas de tendencia, dispersión y visualizaciones que mostraron patrones de comportamiento. Posteriormente, se desarrolló un análisis inferencial que incluyó intervalos de confianza, la prueba Chi-cuadrada para evaluar la independencia entre el nivel educativo y los tipos de lectura, y un modelo de regresión logística binomial para estimar la probabilidad de lectura de distintos contenidos según la escolaridad. Este enfoque permitió identificar relaciones significativas entre el nivel educativo y las preferencias lectoras, aportando evidencia empírica sobre cómo la educación influye en el hábito de lectura en México.

El análisis de los datos del MOLEC muestra que las personas con mayor nivel educativo presentan una mayor frecuencia de lectura, destacando los libros ($\approx 47\%$) y los medios digitales ($\approx 42\%$) como los materiales más comunes, mientras que las revistas son las menos leídas ($\approx 30\%$). Los principales motivos de lectura son el entretenimiento y la cultura general, aunque los niveles educativos más bajos leen principalmente por religión y los más altos por estudio o trabajo. Con base en los intervalos de confianza al 96%, la mayoría de la población se concentra entre los niveles educativos 2 y 4, y se determinó que tanto el nivel educativo como el tipo y motivo de lectura no son variables independientes. Los resultados evidencian que a mayor escolaridad aumenta la lectura de contenidos informativos y disminuye la de entretenimiento, sin encontrarse relación significativa con la lectura de libros prácticos. En conjunto, estos hallazgos confirman que la educación influye de manera notable en los hábitos lectores de la población mexicana.

II. INTRODUCCIÓN

La lectura es una de las competencias fundamentales para el desarrollo académico, cultural y social de las personas, así como también constituye un indicador clave del nivel educativo y de participación ciudadana en un país. Sin embargo, en México el hábito lector presenta niveles preocupantemente bajos que reflejan tanto un problema cultural como estructural. A partir del año 2015, el Instituto Nacional de Estadística y Geografía (INEGI) puso en marcha el Módulo sobre Lectura (MOLEC), con el propósito de generar información estadística sobre los hábitos lectores de la población mexicana mayor de 18 años. Esta encuesta busca ofrecer datos confiables que sirvan como base para el diseño de políticas públicas orientadas a fomentar la lectura en distintos sectores sociales.

Pese a estos esfuerzos institucionales, la lectura en México ha mostrado un deterioro sostenido, demostrado en una reducción del número de libros leídos por persona y en la proporción de adultos que reconocen no leer por falta de tiempo, interés o recursos. Esta situación ha alarmado la preocupación de investigadores, educadores y organismos gubernamentales, quienes han señalado la necesidad de comprender el problema más allá de los resultados superficiales de las encuestas.

A. Objetivo

El presente proyecto surge de la necesidad de profundizar en las causas estructurales y socioculturales detrás del bajo hábito lector en México, empleando para ello los datos recolectados por la encuesta MOLEC a lo largo de sus años de aplicación. El análisis de estas bases de datos permitirá identificar patrones y factores asociados al comportamiento lector de los mexicanos con especial atención en los aspectos educativos, ocupacionales y socioeconómicos, los cuales podrían explicar las diferencias observadas entre distintos grupos poblacionales.

Estudios previos en América Latina, como el realizado por la Universidad de Cuenca (Ibarra et al., 2025), han demostrado que el contexto socioeconómico y el entorno escolar tienen un impacto directo en las prácticas de lectura, especialmente en edades tempranas. En instituciones con menores recursos se observaron deficiencias significativas en la comprensión lectora y el interés por la lectura frente a las escuelas con mayores privilegios económicos. Este hallazgo resulta particularmente relevante para el caso mexicano, donde las desigualdades educativas son amplias y las condiciones de acceso a materiales de lectura varían drásticamente entre regiones urbanas y rurales.

En México, factores económicos, culturales y

lingüísticos se entretienen para agravar el problema. En comunidades indígenas o rurales, el acceso limitado a la educación formal y a material en su lengua materna restringe el desarrollo de competencias lectoras desde edades tempranas (Blanco, 2020). Además, la deserción escolar, influenciada tanto por dificultades académicas (SEP, 2016) como por factores estructurales como la inseguridad y la pobreza (Piña, 2021), contribuye a que amplios sectores de la población adulta carezcan de los hábitos y habilidades necesarias para integrar la lectura en su vida.

Otro elemento que ha modificado el panorama es el creciente uso del internet y las redes sociales. Si bien estas herramientas han democratizado el acceso a la información y han puesto al alcance de todos bibliotecas digitales y recursos de aprendizaje abiertos, también han fomentado hábitos de consumo de contenido efímero. La constante exposición a información breve ha reducido los niveles de concentración y ha reemplazado el tiempo que antes se dedicaba a la lectura de contenido literario (por ejemplo), afectando de manera directa la calidad del hábito lector en la población.

Comprender cómo interactúan los factores económicos, sociales, tecnológicos y educativos con la lectura resulta importante para construir un entendimiento holístico del fenómeno. El análisis de la encuesta, en conjunto con bibliografías nacionales e internacionales, permitirá identificar las relaciones entre variables como el nivel educativo, el tipo de ocupación y el acceso a recursos, proporcionando evidencia empírica que pueda orientar estrategias de fomento lector más efectivas.

Con base en lo anterior, la investigación plantea la siguiente pregunta detonadora:

¿Existe relación entre el nivel educativo de una persona y los tipos de lectura que prefiere?

Esta pregunta surge de la observación de que, aunque el nivel educativo suele asociarse con un mayor desarrollo de habilidades lectoras, no necesariamente garantiza un hábito sostenido de lectura fuera del ámbito académico. Incluso en muchos casos, los individuos con educación superior o posgrado tienden a concentrar su lectura en contenido especializado, mientras que las lecturas recreativas o literatura pueden disminuir. Por otro lado, las condiciones laborales también pueden limitar la práctica de la lectura, enfatizando las ocupaciones con jornadas extensas o escaso tiempo libre.

B. Justificación

El estudio de estos tres aspectos (nivel educativo, factores socioeconómicos y ocupación) permitirá com-

prender de manera más profunda las causas del bajo hábito lector en México. A partir de los resultados del análisis estadístico descriptivo e inferencial, se busca identificar patrones que expliquen las diferencias en el comportamiento lector y ofrecer evidencia que sirva de base para políticas públicas orientadas a fortalecer la educación, promoviendo la lectura en todos los sectores sociales.

La justificación de esta investigación se subraya en el impacto académico y social que se busca generar. Aporta un análisis desde un punto empírico de la situación del lector en México sustentada en datos oficiales, lo que permite contrastar percepciones con evidencia cuantitativa en otras encuestas que podría haber. Asimismo, ofrece información útil para plantear las bases para sugerir educadamente políticas o estrategias educativas que fomenten el hábito de lectura desde edades tempranas y en contextos marginales.

III. METODOLOGÍA

El presente estudio se sustenta en un enfoque cuantitativo descriptivo e inferencial basado en el análisis de los datos provenientes de MOLEC del Instituto Nacional de Estadística y Geografía (INEGI), aplicado entre los años 2019 y 2024. Esta encuesta tiene como propósito medir los hábitos lectores de la población mexicana de 18 años o más, así como los motivos, tipos y frecuencia de lectura. A partir de estos datos, se busca dar respuesta a la pregunta rectora de investigación.

La metodología se dividió en una sección de estadística descriptiva y otra de análisis inferencial. Esto se realiza con el propósito de explorar inicialmente la estructura de los datos y, posteriormente, comprobar mediante técnicas estadísticas la presencia de relaciones significativas entre las variables.

A. Variables Seleccionadas

La pregunta central se abordó a través de un conjunto de variables sociodemográficas y de hábitos lectores seleccionadas del cuestionario MOLEC, las cuales se agruparon en tres categorías.

Variables Sociodemográficas

- **Edad (variable cuantitativa discreta):** permite identificar diferencias generacionales en los hábitos de lectura.

Variables de Hábitos de Lectura

- **Frecuencia de consumo (p3.1, p3.2, p3.5):** indican si el encuestado leyó libros, revistas o medios digitales (1 = sí, 2 = no).

- **Motivos de lectura (p5, p11, p25):** variables nominales que reflejan las razones principales para leer (trabajo, estudio, cultura general, gusto o religión).

Variables Agrupadas Derivadas

- **Nivel educativo (nivel):** Esta variable era categórica dividida en 9 distintos niveles, abarcando desde personas sin estudios hasta aquellas con doctorado. Se decidió agrupar los niveles educativos en 5 categorías distintas, ya que existían grupos con menos del 1% de la muestra. A continuación se muestra el agrupamiento de personas por nivel:

- Nivel 1 - Sin educación: Se juntarán a las personas que marcaron que no tienen estudios o solo estudios de preescolar.
- Nivel 2 - Educación Básica y Media Básica: Se juntarán a las personas que marcaron que su nivel máximo de estudios es de primaria o secundaria
- Nivel 3 - Educación Media Superior: Se incluirán únicamente a aquellos que hayan seleccionado su nivel máximo de estudios como de bachillerato
- Nivel 4 - Educación Superior: Se incluyen a aquellas personas que hayan seleccionado escuela normal, escuela técnica o licenciatura.
- Nivel 5 - Posgrado: Se juntarán a aquellas personas que hayan seleccionado haber cursado una maestría o doctorado.

- **p6_info, p6_prac y p6_entr:** se construyeron estas variables a partir de la categorización de libros en informativos, prácticos y de entretenimiento, esto con el objetivo de identificar patrones en una cantidad más reducida de categorías de manera que fuera comparable con las revistas. A continuación se muestra su agrupación:

- Informativo: Cultura general y académico
- Práctico: Manuales y autoayuda
- Entretenimiento: Literatura

- **p12_info, p12_prac y p12_entr:** construidas a partir de la categorización de revistas en informativas, prácticas y de entretenimiento, con el fin de analizar patrones de preferencia de forma comparativa con los libros. A continuación se muestra su agrupación:

- Informativa: Política, científica y cultura general
- Práctica: Didáctica y salud
- Entretenimiento: Deportes, religiosas y entretenimiento

Estas variables fueron empleadas tanto para caracterizar la muestra descriptivamente, como para establecer las relaciones estadísticas mediante la inferencia que permitieran responder la pregunta de investigación.

B. Fuente de Datos

Para este análisis se está considerando 5 años de encuestas MOLEC, levantadas anualmente entre 2019 y 2024. Cada base contiene información sobre los hábitos de lectura de la población mexicana, complementada con datos sociodemográficos que permiten caracterizar de manera más detallada los patrones de consumo.

Cada base anual (M19–M24) presenta aproximadamente 2000 observaciones (personas encuestadas) y 108 variables (preguntas de la encuesta). El diseño de la encuesta se mantiene constante en los seis años considerados, con las mismas preguntas y categorías de respuesta, lo cual asegura la comparabilidad temporal de los datos. Para efectos de simplicidad se trabajó con la muestra combinada de los tres años a fin de incrementar el tamaño de los datos y la capacidad de crear inferencias.

La calidad de los datos después de la selección de variables relevantes resultó en no tener ningún valor vacío o NaN, lo que es un resultado increíble para el análisis, puesto que no se requirió una limpieza o eliminación de observaciones para dejar las bases listas. No obstante, se llevó a cabo una recategorización de variables así como la creación de nuevas variables para reorganizar información. Estas variables modificadas fueron mencionadas en la sección de **Variables Agrupadas Derivadas** en la metodología.

C. Sección Descriptiva

El análisis descriptivo tiene como propósito explorar las características generales de la población y los patrones de lectura. Para esto se calculan medidas de tendencia central y dispersión (media, mediana, cuartiles, desviación estándar, coeficiente de variación) para las variables cuantitativas (si aplican), así como frecuencias absolutas y relativas para las variables categóricas. Adicionalmente, se obtiene medidas de forma como el sesgo y la curtosis mediante la librería *moments* en R con el fin de identificar la simetría y concentración de los datos en relación con una distribución normal.

En esta etapa también se desarrollará diversas gráficas, sean estas histogramas, diagramas de caja y bigote, y gráficos de barras y pastel para describir la proporción de lectura por tipo de material y las diferencias en los motivos de lectura según el nivel educativo y la ocupación.

D. Sección Inferencial

Para comprobar la existencia de una relación significativa entre el nivel educativo y los tipos de materiales o contenidos leídos, se implementarán varias técnicas inferenciales como

Estimación de IC para proporciones.

Intervalos de confianza al 96% ($\alpha = 0.04$) para proporciones poblacionales de nivel educativo y preferencias de lectura. Este nivel de significancia, específico del estudio, equilibra sensibilidad y especificidad estadística, controlando simultáneamente errores Tipo I y II. El tamaño muestral ($n = 5998$) satisface los supuestos de normalidad por el Teorema del Límite Central, validando la metodología.

Prueba de independencia Chi-cuadrada (χ^2)

Se aplica la prueba χ^2 de independencia para buscar con un nivel de confianza de 96% que pueda existir o no alguna relación entre el nivel educativo y el tipo de texto leído.

Regresión logística binomial

Se implementará un modelo de regresión logística binomial para cuantificar la fuerza y dirección de la relación entre el nivel educativo y la probabilidad de leer distintos tipos de contenidos (informativos, prácticos y de entretenimiento). Se estiman los odds ratios (OR) de cada modelo, donde un valor superior a 1 indica que a mayor nivel educativo, mayor probabilidad de lectura del tipo de contenido en cuestión.

Prueba de independencia de Fisher

Finalmente, para evaluar la dependencia entre el nivel educativo y los motivos de lectura de la muestra, se construyeron tablas de contingencia para los distintos materiales de lectura (libros, revistas y medios digitales). Dado que algunas categorías presentan frecuencias bajas o incluso nulas, se aplicó la prueba exacta de Fisher, ya que la prueba de Chi-cuadrada no puede aplicarse cuando no se cumplen los supuestos de muestras grandes y frecuencias esperadas suficientemente altas en todas las categorías. Debido al elevado número de categorías, la prueba se realizó utilizando simulación de Monte Carlo para estimar los valores de p de manera confiable. Esta prueba permite determinar si existe evidencia estadística suficiente para rechazar la hipótesis nula de independencia entre las variables categóricas.

IV. ANÁLISIS DESCRIPTIVO E INFERENCIAL

A. Medidas descriptivas

Considerando la naturaleza de cada una de las variables, se deben de realizar distintos análisis, así como tratar a cada tipo de dato de una manera distinta. De primera instancia, es importante considerar las variables ‘edad’ y ‘nivel’. No se obtienen estadísticos del nivel porque es una variable categórico mapeada a numérica.

Edad

##	2020	2022	2024	Acum Tot
## Mínimo	18.00	18.00	18.00	18.00
## Q1	30.00	31.00	32.00	31.00
## Mediana	44.00	43.00	44.00	44.00
## Media	44.74	45.08	45.68	45.17
## Q3	58.00	57.00	59.00	58.00
## Máximo	93.00	97.00	94.00	97.00
## Rango Medio	37.50	39.50	38.00	39.50
## Desv Est	17.01	16.42	16.76	16.74
## Coef Var	2.63	2.74	2.73	2.70
## Sesgo	0.34	0.35	0.30	0.33
## Curtosis	2.23	2.28	2.17	2.22

Con este análisis cuantitativo, se puede empezar a identificar características de la población encuestada: la edad promedio gira en torno a 45 años y la mediana se mantiene en 44, indicando que la mayoría de la población se encuentra entre estas edades. Además, el 50% de los datos se encuentran entre 31 y 58 años. La edad presenta una dispersión moderada, en promedio de $\sigma = 16.72$, indicando que es una muestra heterogénea entre distintas generaciones.

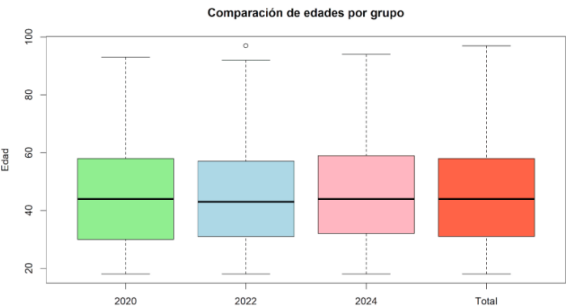


FIG. 1: Comparación de Edad por Boxplots

Además, se incluyen histogramas para visualizar la información de mejor manera.

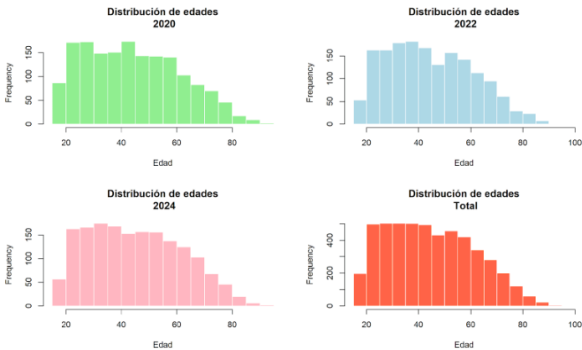


FIG. 2: Comparación de Edad por Histogramas

Con respecto al sesgo, se identifica uno de 0.33 para todo el conjunto de datos, indicando que la cola de datos se encuentra más hacia la izquierda, es decir, hay más personas jóvenes que personas mayores.

Finalmente, la curtosis del total de datos al ser $2.23 < 3.00$, indica que hay una distribución más plana que la distribución normal, por lo que las edades se encuentran menos concentradas en el rango de la media.

Tipo de textos leídos

En la base de datos, existen una gran cantidad de variables que solamente deben de tomar dos valores: 1 o 2, donde 1 representa afirmación y 2 representa negación. Este tipo de variables consisten en aquellas de qué tipo de contenido leen los encuestados. A continuación, se muestran gráficos de barras referentes a la pregunta 3, que buscan conocer si el encuestado ha leído libros, revistas o páginas en internet.

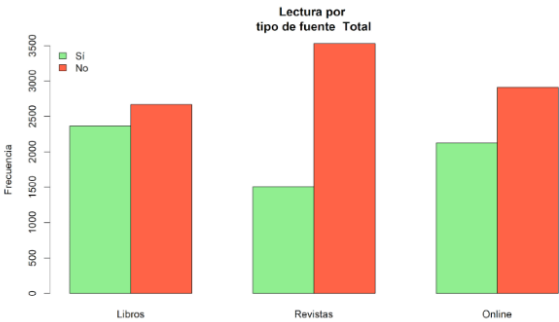


FIG. 3: Comparación de Edad por Histogramas

En general, la gente tiende a leer más libros físicos que páginas en línea o revistas. De la misma manera, es interesante ver el decremento gradual de la cantidad de gente que sí lee revistas: en el año 2020 se identifica una frecuencia de más de 600 para esta selección, mientras que para 2024 se reduce hasta apenas 400.

Motivos de lectura de libros

Dentro de las opciones, se tiene la pregunta 5 que consiste en el motivo principal por el que el encuestado leyó libros, así como también el tipo de contenido dentro de los libros leídos.

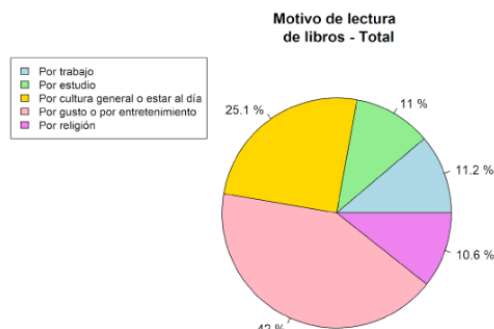


FIG. 4: Comparación de Edad por Histogramas

Con base en los gráficos, se ve cómo la mayoría lee por gusto o por entretenimiento. Este porcentaje tuvo un aumento en 2022, donde el 44% de los encuestados leían por gusto, a comparación del 2020 y 2024, donde el 40.8% y 41% leían por entretenimiento respectivamente.

Motivos de lectura de revistas

También se cuenta con la información relacionada al tipo de contenido en revistas. En total, en la encuesta se dieron 5 opciones distintas para seleccionar. Para revistas, también se muestra esta tendencia de lectura por entretenimiento, seguido de cultura general.

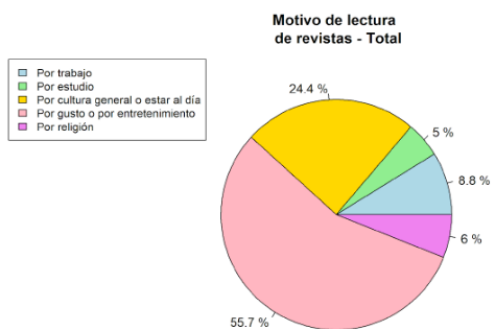


FIG. 5: Comparación de Edad por Histogramas

Relación entre nivel educativo y motivos de lectura

Los motivos de lectura, en todos los casos, al igual que en las gráficas anteriores están codificados de la siguiente forma: (1) Por trabajo, (2) Por estudio, (3) Por cultura general, (4) Por gusto y (5) Por religión.

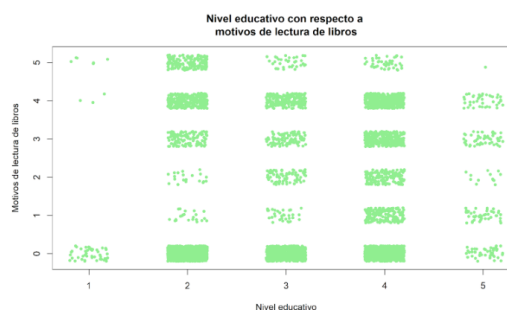


FIG. 6: Jitter Plot de nivel educativo contra motivo de lectura de libros

Se puede ver cómo la gente con menor nivel de educación leen principalmente por religión, y se ven tres puntos en el motivo de entretenimiento. En niveles educativos más avanzados (aquellos que completaron al menos la primaria) se ve una mayor tendencia a la lectura por entretenimiento pero también por religión. Sin embargo, aquellas personas con nivel educativo medio superior o más tienden a leer más por trabajo y mucho menos por religión.

No obstante, el principal motivo de lectura en cada uno de los niveles es el entretenimiento.

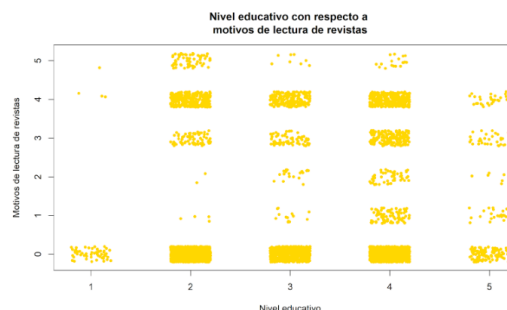


FIG. 7: Jitter Plot de nivel educativo contra motivo de lectura de revistas

En este gráfico, se puede ver nuevamente cómo los motivos de lectura difieren entre niveles educativos. La religión predomina en el nivel educativo básico, así como también el entretenimiento y cultura general. Mientras tanto, se visualiza cómo la cantidad de personas que leen por trabajo o estudio aumenta a la par del nivel educativo. Esto resulta interesante y prueba el punto de distintos motivos de lectura con respecto a la educación de cada individuo.

Tipo de contenido leído por nivel educativo

Por extensión, todos los gráficos de esta sección de encuentran en el apartado de anexos (gráficos 8-12). En este caso se observa un aumento en la proporción de textos de índole religiosa que leen los encuestados

conforme aumentan los niveles de estudios. Similarmente se observa un descenso significativo de la proporción de la gente que lee textos literarios, respaldando la hipótesis inicial de que al tener mayores niveles de estudios, los encuestados tienden a tener menor tiempo de leer textos por motivos lúdicos, como usualmente se leen los textos literarios.

En cuanto al resto de tipos de textos, se observa poco cambio entre niveles educativos. Cabe resaltar que la lectura de textos culturales alcanza su pico en el nivel tres, y que los textos académicos, contrario a lo que se pudiera asumir, observan una tendencia a la baja.

B. Análisis Inferencial

1. Intervalos de Confianza

A continuación, se determinará que proporción de la muestra representa cada nivel de estudios:

```
## Nivel 1 representa: 1.131626 % de la muestra
## Nivel 2 representa: 35.79512 % de la muestra
## Nivel 3 representa: 21.32222 % de la muestra
## Nivel 4 representa: 37.74072 % de la muestra
## Nivel 5 representa: 1.131626 % de la muestra
```

También, se pueden obtener proporciones para la lectura por tipo de materiales:

```
## El 46.99226 % de los encuestados afirma leer
  libros, mientras que el 53.00774 afirma no
  leerlos.
## El 29.89875 % de los encuestados afirma
  leer revistas, mientras que el 70.10125
  afirma no leerlas.
## El 42.18781 % de los encuestados afirma
  leer materiales digitales, mientras que el
  57.81219 afirma no leerlos.
```

A partir de esto es posible obtener algunos IC de la proporción real que representa cada grupo de la población, asumiendo normalidad y considerando que se cumple el Teorema del Límite Central (TLC) por el tamaño de la muestra, con el valor de alfa sugerido de 0.04:

TABLE I: Intervalos de confianza de una proporción para los niveles

Nivel	Límite Inf	Límite Inf (Real)	Límite Sup
Nivel 1	-1.746%	0.000%	24.009%
Nivel 2	33.476%	33.476%	38.114%
Nivel 3	18.755%	18.755%	23.889%
Nivel 4	35.457%	35.457%	40.024%
Nivel 5	-0.397%	0.000%	2.660%

Como la proporción no puede ser menor a 0, se va a limitar el intervalo de confianza real a ser mínimo 0.000%.

Esto reafirma lo que se había detectado en la etapa anterior. Del total, la población está mayoritariamente

representada en el nivel dos, tres y cuatro, con los niveles uno y cinco representando porciones marginales de esta.

```
## La proporción real de la población que lee
  libros está, con un 96% de confianza, entre
  45.548 % y 48.43651 %

## La proporción real de la población que lee
  revistas está, con un 96% de confianza,
  entre 28.57395 % y 31.22355 %

## La proporción real de la población que lee
  materiales digitales está, con un 96% de
  confianza, entre 40.7587 % y 43.61692 %
```

De igual forma se conserva el patrón detectado en la etapa previa: La cantidad de la población que lee cualquier tipo de texto no llega a superar nunca el 50%, lo cual reafirma el problema prevalente de lectura en México, independientemente del nivel de estudios.

2. Prueba Chi-Cuadrada

Dado que las variables “nivel educativo” y “tipo de material leído” son categóricas, se utilizó la prueba de independencia Chi-cuadrada para identificar si existe relación entre ambas.

A continuación, se realizará una prueba de hipótesis utilizando el test de χ^2 para identificar si existe una diferencia de proporción entre el nivel de estudios con base en las categorías que se designaron con anterioridad y el tipo de texto que leen las personas: Libros, Revistas y Medios digitales. Las hipótesis se establecen como sigue:

- H_0 : El nivel educativo y el tipo de material de lectura son independientes (no existe asociación)
- H_a : El nivel educativo y el tipo de material de lectura están asociados (existe relación)

Usando la tabla de contingencia IV donde se analiza nivel educativo contra tipo de texto leído por los encuestados, resulta ser que con esta prueba que el valor $p = 4.922 \times 10^{-8}$, lo que es significativamente menor a $\alpha = 0.04$. Se rechaza la hipótesis nula H_0 ya que existe evidencia estadística para inferir que entre las proporciones entre los niveles educativos y los tipos de texto (libros, revistas y medios digitales) alguna es distinta.

3. Regresión Logística Binomial

La regresión logística binomial es un modelo que busca explicar la relación entre una variable dependiente binaria y una variable independiente que puede ser categórica. Este modelo permite cuantificar la relación existente entre las variables independientes y las posibles dependientes. Para poder usar el modelo de regresión logística binomial, es necesario que existan únicamente dos valores para la variable dependiente: característica que se

cumple con las variables a usar. Esta regresión consiste en obtener una función lineal de las distintas variables independientes que permiten clasificar las observaciones en uno de los dos subconjuntos dentro de los posibles valores de la variable dependiente (Berlanga, 2014). Esto tiene la siguiente forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k = \log \frac{p}{1-p} \quad (1)$$

Donde β_0 es la intersección del modelo, β_k son los coeficientes estimados a partir de los datos, X_k son las variables independientes y p es la probabilidad de que ocurra el evento de interés. El objetivo de cada uno de los coeficientes β_k es maximizar la verosimilitud V del modelo, mientras se minimiza la distancia L (Fernández, 2004), definida como:

$$L = -2 \ln(V) \quad (2)$$

Se busca calcular el *Odds Ratio* (OR), que representa cuántas veces es más probable el evento cuando la variable independiente aumenta en una unidad, manteniendo constantes las demás variables y se calcula de la siguiente manera:

$$OR = e^{\beta_1} \quad (3)$$

A continuación, se elaborarán las tablas de contingencia (encontrada en anexos como V y VI) para los tipos de contenidos dentro de la lectura, tanto para libros como revistas. Se omiten medios digitales puesto que en ninguna encuesta MOLEC se encuentra información recabada relacionada al contenido digital.

El uso de la regresión logística binomial para este proyecto será identificar relación alguna entre el tipo de contenido que leen las personas con respecto a su nivel educativo. En este caso, se contará con seis modelos distintos. La finalidad de la regresión logística binomial es hacer uso de los coeficientes y ORs para identificar qué tan probable es que lea una persona cierto tipo de contenido con base en su nivel educativo. Para cada modelo, se realiza una prueba de hipótesis utilizando el parámetro del coeficiente β_1 :

- $H_0 : \beta_1 = 0$. Esto indica que no existe relación entre el nivel y la probabilidad de leer el tipo de material que se está evaluando.
- $H_a : \beta_1 \neq 0$. Esto indica que existe relación alguna entre el nivel y la probabilidad de leer el tipo de material que se está evaluando.

Esta es una prueba de hipótesis bilateral, y para rechazar o no H_0 se utilizará el valor p del modelo logístico binomial. Se recuerda que el valor de $\alpha = 0.04$, por lo que se utilizará la comparativa del valor p con respecto a α para estas pruebas de hipótesis y esta comparación será la decisión final para rechazar o no la hipótesis nula.

A continuación, se muestran tablas con los coeficientes β_1 , el Odds ratio y el valor p de cada modelo:

TABLE II: Tabla de coeficientes β_1 , Odds Ratio y valores p para cada modelo. Autoría propia.

Tipo Contenido	Coeficiente	Odds Ratio	Valor p
Libros informativos	0.8244	2.2804	4.9603×10^{-94}
Libros prácticos	0.0545	1.0560	0.1537
Libros entretenimiento	0.6387	1.8940	4.9802×10^{-56}
Revistas informativas	0.6227	1.8639	1.6501×10^{-46}
Revistas prácticas	0.2961	1.3446	4.2135×10^{-10}
Revistas entretenimiento	-0.1886	0.8281	1.0795×10^{-05}

Se puede ver cómo el coeficiente β_1 y el OR de los libros informativos son los más elevados, así como su valor p es el más bajo, con un orden de magnitud de 10^{-94} . Además de rechazar H_0 , se indica que existe una relación estrecha entre el nivel educativo y la lectura de libros informativos puesto que el coeficiente es positivo y el OR es mayor a 1. Esto también se refleja en las revistas informativa, dando una clara visión de que la frecuencia de lectura del contenido informativo como papers y publicaciones políticas aumenta conforme aumenta el nivel.

Por otro lado, se puede ver que en las revistas de entretenimiento, el coeficiente β_1 es negativo, así como su OR es inferior a 1. Si bien se rechaza H_0 , existe una relación es negativa con base en β_1 y el OR, indicando que entre menor es el nivel educativo, aumenta la lectura de revistas de entretenimiento.

Finalmente, se ve que en el modelo de libros prácticos el valor p es mayor a 0.04, indicando que no se rechaza H_0 . Con una certeza estadística del 96%, las personas leen libros prácticos independientemente de su nivel educativo.

4. Prueba de Fisher

Como última prueba de hipótesis para el reporte, se optó por identificar si existía una relación entre el nivel de estudios y las motivaciones para la lectura de libros, revistas y medios digitales. El planteamiento de las hipótesis se muestra a continuación:

- H_0 : El nivel educativo y el motivo de lectura de un texto en específico son variables independientes.
- H_a : El nivel educativo y el motivo de lectura de un texto en específico son variables dependientes.

En el apéndice se encuentran las tablas de contingencia de cada nivel con respecto al motivo de lectura de libros (tabla VII), revistas (tabla VIII) y medios digitales (tabla IX).

El motivo de elección de la prueba de Fisher radica en que, aunque la prueba χ^2 suele ser más adecuada cuando el tamaño muestral es grande, la prueba de Fisher ofrece mayor precisión al analizar tablas de contingencia con frecuencias pequeñas en variables categóricas. Esto se evidencia, por ejemplo, en la categoría de personas que eligen la religión como motivo principal de lectura, cuya frecuencia alcanza valores tan bajos como cinco observaciones, tal como se muestra en la Tabla IX.

Asimismo, se observa la presencia de frecuencias nulas (0) en la intersección de los grupos con nivel educativo de posgrado y motivo de lectura religioso, lo que imposibilita el uso de la prueba χ^2 , ya que esta requiere que todas las frecuencias esperadas sean mayores a cero. En este contexto, el modelo de Fisher resulta óptimo ya que permite evaluar la independencia entre variables incluso en tablas con frecuencias bajas. Sin embargo, dado que la tabla de contingencia es relativamente grande para una prueba de Fisher exacta, se recurrió a la simulación de Monte Carlo para estimar el valor p . Este método permite obtener una aproximación precisa, especialmente en presencia de frecuencias nulas.

Finalmente, se presenta una tabla con los resultados obtenidos para cada tipo de material de lectura y los valores p derivados de la prueba de Fisher, empleando un total de un millón de simulaciones de Monte Carlo:

TABLE III: Pruebas de Fisher: Relación entre Nivel Educativo y Motivaciones de Lectura

Tipo de Material	Valor p
Libros	3.0×10^{-6}
Revistas	2×10^{-5}
Medios Digitales	1.23×10^{-4}

Se puede ver que los tres tipos de material tienen valores p menores a α , que fue definido como 0.04 desde el inicio de la sección de estadística inferencial. Los resultados muestran con un intervalo de confianza del 96% que existe evidencia estadística suficiente para sugerir dependencia entre el nivel educativo y el motivo de lectura de cada tipo de material.

V. RESULTADOS

El análisis descriptivo permitió identificar que la mayoría de los encuestados leen por entretenimiento o cultura general, y que la lectura de revistas presenta una tendencia decreciente, mientras que el consumo de libros y medios digitales se mantiene estable. Además permitió identificar que más del 97% de la muestra se ubica en las categorías intermedias de nivel educativo.

Del análisis descriptivo se descubrió también que para todos los tipos de textos la cantidad de la muestra que

afirma leerlo es inferior al 50%. Esto revela un patrón preocupante de lectura en este país.

Este análisis también nos reveló los motivos principales de lectura. En general tanto para libros como revistas y medios digitales el motivo principal de lectura reportado es gusto y cultura, con trabajo, estudio y religión representando motivos menores. En relación con el nivel educativo, no hay muchas diferencias en cuanto a los motivos, excepto por el nivel 4 que muestra un nivel más alto en cuanto a lectura por estudio, posiblemente porque al ser estudiantes de nivel universitario requieren hacerlo con mayor frecuencia.

En contraste, el contenido leído muestra un patrón interesante, que la lectura de textos religiosos aumenta conforme al nivel educativo. Por otra parte la lectura de textos académicos o informativos muestra un patrón similar, y la de textos de entretenimiento decrece respecto al nivel.

En el análisis inferencial, de primera instancia mediante la prueba χ^2 se identificó dependencia entre el nivel educativo y el tipo de material de lectura (libros, revistas y medios digitales). Con un valor $p = 4.922 \times 10^{-8}$, se rechazó la H_0 , abriendo puertas para el análisis de dependencia entre el tipo de contenidos de los materiales de lectura y el nivel educativo. Para ese análisis, se hizo uso de la regresión logística binomial.

Los resultados de esta segunda prueba evidenciaron que los contenidos informativos son los más asociados a un mayor nivel educativo (OR = 2.28 para libros y OR = 1.86 para revistas), mientras que las revistas de entretenimiento mostraron una relación inversa (OR = 0.83), indicando una menor tendencia a su consumo conforme aumenta la escolaridad. Para libros prácticos no se identificó relación con respecto al nivel educativo, puesto que el OR era muy cercano a 1, así como su valor p era mayor que α . Mediante la regresión logística binomial, se ha obtenido evidencia estadística suficiente para sugerir una relación entre la lectura de contenido de carácter informativo con respecto al nivel educativo de las personas.

Finalmente, con una tercera prueba de hipótesis mediante la prueba de Fisher se identificó dependencia también entre el nivel educativo y los motivos de lectura de libros, revistas y también de medios digitales, que tenían valores p de 3.0×10^{-6} , 2×10^{-5} , y 1.23×10^{-4} , respectivamente. Con respecto a α , se determinó que se rechazaba H_0 .

VI. DISCUSIÓN Y CONCLUSIONES

Los puntos mencionados en los resultados son cruciales ya que se parte del análisis descriptivo para poder

avanzar más allá dentro del análisis inferencial. Respecto al nivel educativo es necesario mencionar que a pesar de que los extremos representan una proporción pequeña de la muestra, y en consecuencia, de la población, fue de valor incluirlas en nuestro análisis ya que el enfoque de la investigación requería conocer el comportamiento de los casos extremos.

Tras realizar el análisis descriptivo se detectan patrones que sugerirían que existe una dependencia entre el tipo de textos leídos y el nivel educativo. Del mismo modo, sugiere que el nivel educativo podría tener un efecto en los motivos de lectura. Además, es necesario mencionar el porcentaje de la muestra que afirmó leer cualquier texto, ya que lo preocupantemente bajo que fue delata problemas serios respecto a lo que se lee en este país.

La pregunta rectora original del reporte sí pudo ser respondida con éxito debido a que el análisis estadístico logró establecer y cuantificar la naturaleza de la relación (dependencia o independencia) entre el nivel educativo máximo alcanzado (la variable principal) y los tipos de lectura preferidos (la variable de resultado), lo cual era el objetivo del estudio. Se estableció con un nivel de confianza del 96% que el nivel educativo y el tipo de lectura no son variables independientes, reforzado por la prueba Fisher que indica que el nivel educativo y el motivo de lectura tampoco son independientes. La mayoría de la población se concentra entre los niveles 2 y 4, con base en los intervalos de confianza para estimar las proporciones poblacionales.

Los patrones específicos de relación encontrados son: entre mayor es el nivel educativo, se tiende a leer más revistas y libros informativos, mientras que hay una menor tendencia en lectura de revistas de entretenimiento. Asimismo, se descubrió que no hay relación (son independientes) entre el nivel educativo y la lectura de libros prácticos como manuales.

A pesar de que hubo éxito en los resultados, no se puede afirmar con certeza que el análisis es completamente certero. Entre las principales limitaciones de la investigación se encuentra el hecho de que el estudio se basa exclusivamente en la encuesta MOLEC, lo que restringe el análisis a las variables y preguntas establecidas por dicha fuente, sin considerar otros factores cualitativos o contextuales. Aunque se utilizaron datos de varios años (2019–2024), no se realizó un análisis de evolución temporal que permitiera observar tendencias o cambios en los hábitos lectores a lo largo del tiempo. Asimismo, se asumieron supuestos de normalidad y homogeneidad para la aplicación de intervalos de confianza y pruebas Chi-cuadrada, los cuales podrían no cumplirse completamente en todos los casos. Una posible área de mejora consiste en complementar el enfoque cuantitativo con métodos cualitativos como entrevistas o encuestas

abiertas que aporten una comprensión más profunda sobre los factores culturales y sociales que influyen en el hábito lector en México.

Las recomendaciones se fundamentan en el bajo hábito lector general (menos del 50%) y la concentración poblacional. Los programas pueden enfocarse en los niveles educativos intermedios, ya que representan la mayor parte de la población lectora potencial. Una sugerencia es integrar la lectura informativa y crítica desde etapas tempranas, dado que esta correlaciona fuertemente con los niveles educativos altos. Además se debe promover la lectura por placer en educación superior dada la tendencia a abandonar la lectura recreativa por otras (se dificulta la formación de un hábito lector sostenido).

Extendiendo la investigación para futuros hallazgos, podemos plantearnos preguntas no abarcadas por MOLEC como: ¿Cuáles son los factores culturales y contextuales (no medidos por MOLEC) que explican el bajo hábito lector? ó ¿Cómo ha evolucionado la lectura en medios digitales (2019-2024) en los diferentes niveles educativos? (pues había poca cobertura en este medio de lectura y las tendencias tecnológicas apuntan para allá).

VII. ANEXOS

A. Liga del RMD

Repositorio en GitHub: MA1001B-Estadística

B. Gráficos

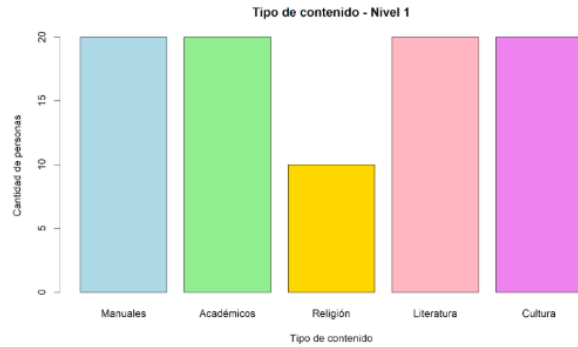


FIG. 8: Tipo de contenido leído por nivel N1

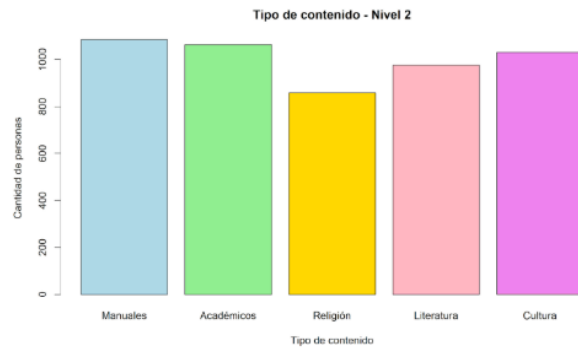


FIG. 9: Tipo de contenido leído por nivel N2

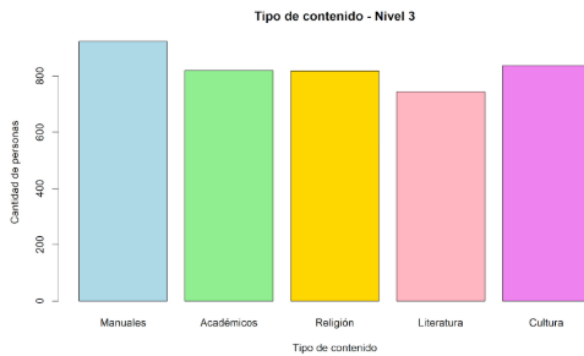


FIG. 10: Tipo de contenido leído por nivel N3

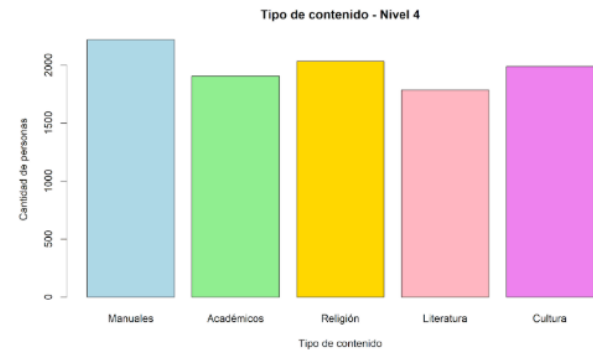


FIG. 11: Tipo de contenido leído por nivel N4

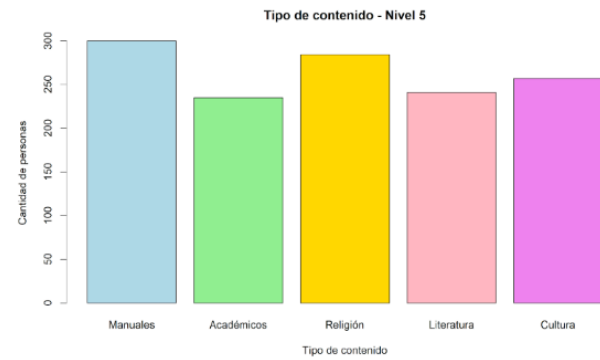


FIG. 12: Tipo de contenido leído por nivel N5

C. Tablas

TABLE IV: Tabla de contingencias de nivel educativo y tipos de texto. Autoría propia.

Tipo Texto	N1	N2	N3	N4	N5	Total
Libros	10	567	477	1156	157	2367
Revistas	4	436	315	662	89	1506
Digitales	7	416	513	1050	139	2125
Total	21	1419	1305	2868	385	5998

TABLE V: Tabla de contingencias de nivel educativo y tipos de textos de libros. Autoría propia

Tipo Contenido	N1	N2	N3	N4	N5	Total
Informativos	0	164	228	652	114	1158
Prácticos	10	316	165	358	42	891
Entretenimiento	0	157	210	526	73	966
Total	10	637	603	1536	229	3015

TABLE VI: Tabla de contingencias de nivel educativo y tipos de textos de revistas. Autoría propia.

Tipo Contenido	N1	N2	N3	N4	N5	Total
Informativos	0	136	167	427	71	801
Prácticos	1	155	106	251	42	555
Entretenimiento	3	289	162	228	8	690
Total	4	580	435	906	121	2046

TABLE VII: Tabla de contingencias de nivel educativo y motivos de lectura de libros. Autoría propia

Motivo	N2	N3	N4	N5	Total
Trabajo	1	11	58	21	91
Estudio	2	17	58	7	84
Cultura general	21	25	85	15	146
Entretenimiento	26	39	97	6	168
Religión	6	4	6	0	16
Total	56	96	304	49	505

TABLE VIII: Tabla de contingencia entre nivel educativo y motivos de lectura de revistas. Autoría propia.

Motivo	N2	N3	N4	N5	Total
Trabajo	1	6	42	16	65
Estudio	1	10	33	6	50
Cultura general	15	23	94	22	154
Entretenimiento	35	55	129	17	236
Religión	4	2	6	0	12
Total	56	96	304	61	517

TABLE IX: Tabla de contingencia entre nivel educativo y motivos de lectura de medios digitales. Autoría propia.

Motivo	N2	N3	N4	N5	Total
Trabajo	3	10	63	23	99
Estudio	4	14	42	7	67
Cultura general	22	36	110	23	191
Entretenimiento	26	35	86	8	155
Religión	1	1	3	0	5
Total	56	96	304	61	517

VIII. REFERENCIAS

Berlanga, V., & Vilà Baños, R. (2014). **Cómo obtener un modelo de regresión logística binaria con SPSS**. Revista d'Innovació i Recerca en Educació, 7(2).

Blanco Bosco. (2020). **Discriminación étnico-racial y oportunidades educativas en México**. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-01732020000300139&lng=es&tlng=es

Castro Castillo. (s.f.). **Factores socioeconómicos que influyen en la deserción escolar de los estudiantes de un sistema de preparatorias en Yucatán**. Disponible en: https://www.comie.org.mx/congreso/memoriaelectronica/v11/docs/area_10/1784.pdf

Fernández, V. P., & Fernández, R. S. M. (2004). **Regresión logística multinomial**. Cuadernos de la Sociedad Española de Ciencias Forestales, 18, 323–327.

Ibarra, E., Espinoza, J., & Moscoso, S. (2025). **Influencia del entorno socioeconómico en el hábito lector de estudiantes de educación básica media**. Recuperado de: https://www.researchgate.net/publication/388116303_Influencia_del_entorno_socioeconomico_en_el_habito_lector_de_estudiantes_de_educacion_basica_media

Instituto Nacional de Estadística y Geografía [INEGI]. (2023). **Módulo sobre lectura (MOLEC) 2023**. Recuperado de: <https://www.inegi.org.mx/programas/molec/>

Piña Osuna. (2021). **Deserción escolar y participación en actividades de tráfico de drogas en Sonora, México**. Disponible en: <https://doi.org/10.15174/au.2021.3193>

Secretaría de Educación Pública [SEP]. (2016). **Informe de factores que motivan el abandono escolar**.