

# Determinantes Socioeconómicos y Demográficos de la Educación Formal en México: Un Enfoque Econométrico Basado en la ENOE

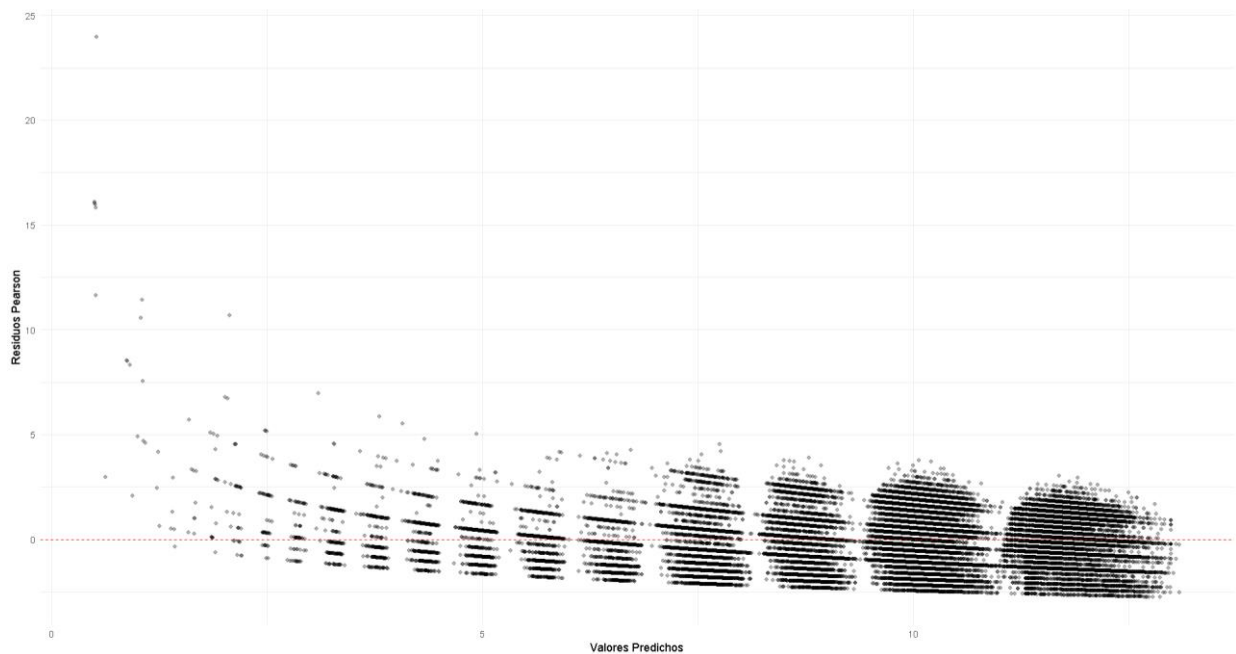
Kevin Isaac Monreal Nuñez |

347884

| [ki.monrealnunez@ugto.mx](mailto:ki.monrealnunez@ugto.mx)

Econometría de la Salud

---



## **Resumen**

Este estudio analiza cómo los factores sociales y económicos, así como las tendencias poblacionales, afectan el número de años que las personas en México pasan en la escuela. Utiliza información de la Encuesta Nacional de Ocupación y Empleo. Se utilizó un tipo de modelo llamado Binomial Negativa para tratar la dispersión excesiva de los datos. Los resultados muestran que la edad, el estado civil, el número de hijos menores de cinco años y el tamaño del hogar afectan significativamente la educación, reflejando desigualdades estructurales. Las limitaciones del análisis muestran que necesitamos incluir poblaciones marginadas y más factores para asegurarse de que Los resultados pueden utilizarse en políticas públicas.

**Palabras clave:** Educación formal, factores socioeconómicos, Encuesta Nacional de Ocupación y Empleo (ENOE), sobredispersión, modelo de binomial negativa, tamaño del hogar, estado civil, política pública, desigualdades educativas, análisis econométrico

## Introducción

La educación es uno de los pilares más importantes para el desarrollo tanto individual como social. Juega un papel clave en la movilidad económica y en la apertura de oportunidades. Sin embargo, el nivel educativo que las personas pueden alcanzar está influenciado por varios factores socioeconómicos y demográficos. En este sentido, esta investigación se enfoca en analizar cómo variables como el género, el número de hijos, el estado civil, la edad y el tamaño del hogar afectan los años de educación formal en México, usando datos de la Encuesta Nacional de Ocupación y Empleo (ENOE).

Estudios previos han demostrado que el acceso y la permanencia en el sistema educativo dependen de muchos factores. Por ejemplo, se ha encontrado que las mujeres con hijos enfrentan mayores barreras para continuar sus estudios debido a las demandas del cuidado infantil. No obstante, en algunos casos, la maternidad puede motivar a las madres a esforzarse más en su educación para asegurar un mejor futuro para sus hijos (Lactancia materna y deserción escolar). Además, la edad y el estado civil también juegan un papel significativo: la juventud suele estar asociada con niveles más altos de escolaridad, mientras que el matrimonio o el divorcio pueden causar interrupciones significativas en el proceso educativo (Susana et al.).

Por otro lado, la estructura del hogar también es crucial. Los hogares con más integrantes suelen tener menos recursos disponibles por persona, lo que puede traducirse en menos oportunidades educativas (Plitt, 2023). En el caso de México, donde las desigualdades socioeconómicas son bastante marcadas, es esencial entender cómo estos factores interactúan para poder diseñar políticas públicas que reduzcan las brechas educativas.

Este trabajo utiliza un enfoque econométrico para explorar estas dinámicas, empezando con la selección de modelos estadísticos que se ajusten a las características de los datos, como los modelos de Poisson y binomial negativa, para analizar variables dependientes discretas. A través de este análisis, buscamos aportar evidencia empírica que no solo explique los determinantes de los años de educación formal, sino que también ayude a identificar áreas clave para la intervención de políticas públicas.

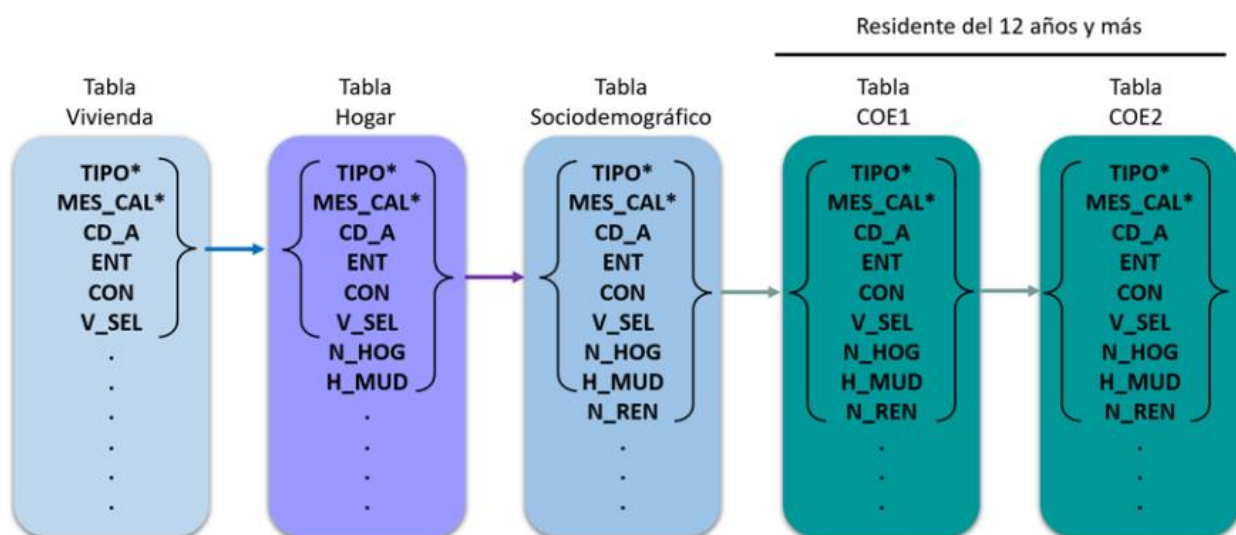
## Objetivo Principal

Analizar cómo factores socioeconómicos y demográficos, como la edad, el estado civil, el número de hijos menores de cinco años y el tamaño del hogar, influyen en los años de educación formal alcanzados en México, utilizando un enfoque econométrico basado en datos de la Encuesta Nacional de Ocupación y Empleo (ENOE).

## Desarrollo y Argumentación

Lo primero que empecé a explicar es como funciona la base de datos de la encuesta nacional de ocupación y empleabilidad (ENOE) esta tiene un total de 5 partes en la cual se divide, donde vienen datos de vivienda, del hogar, sociodemográficos y dos documentos de empleo.

*Figura 1: Relación de las 5 tablas*



*Nota: esta imagen explica la relación que hay entre las 5 tablas de datos que presenta la ENOE.*

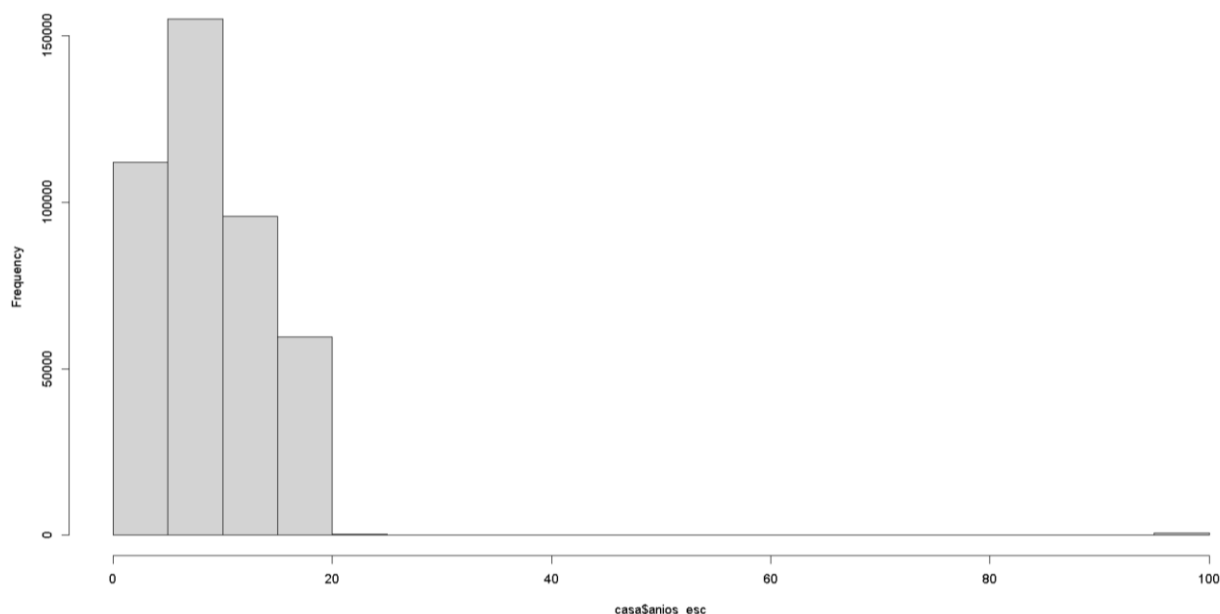
Sabiendo como funciona la ENOE lo siguiente fue analizar un problema simple con datos reales, para el problema decidí explicar los “años de educación formal” en base a cinco factores, el primero es el sexo del individuo, básicamente si es hombre o mujer, el segundo factor son la cantidad de hijos igual o menores a 5 años, el tercer factor es el estado conyugal, el cuarto factor es el tamaño del hogar, entiéndase como la cantidad de personas que viven en el hogar y el último pero no menos importante es la edad.

Algo que se tiene que explicar es que la ENOE entrevista a un **número representativo** de personas cada mes, y no a toda la población, lo que permite obtener una visión general del mercado laboral sin necesidad de encuestar a cada individuo. Al final de cada uno de los 5 documentos hay una columna que se llama “FAC\_...” puede terminar en mensual o trimestral, lo que significa es factor, por ejemplo, supongamos que una fila del documento es la respuesta a toda la encuesta y en la columna final del “factor trimestral (FAC\_tri)” viene un número que

representa a la cantidad de personas que se identifican en esa situación contestando de esa misma manera y así para cada una de las filas de la encuesta.

Lo siguiente en hacer fue visualizar los datos para ver cómo se repartían en un histograma y poder visualizar detalles que entre grandes datos

*Grafica 1: Histograma de años de educación*

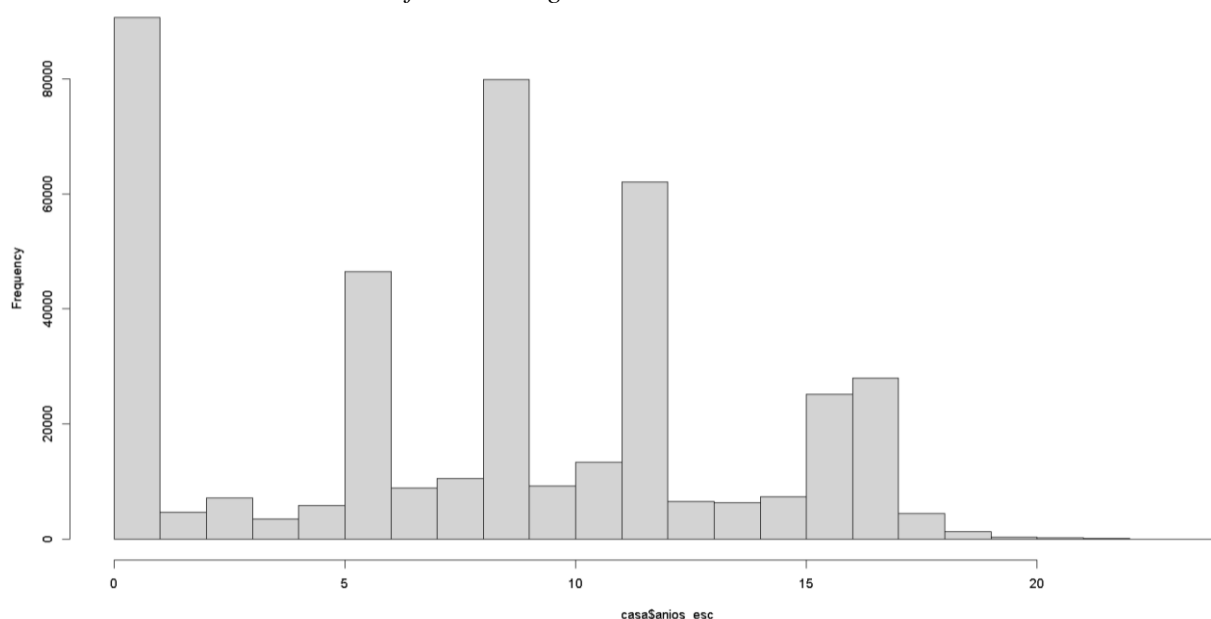


*Nota: esta grafica muestra la distribución de los años de escolaridad formal de las personas. El número final 99 son las personas que no conocen el número de años que estudiaron*

Lo principal fue limpiar la muestra, así que filtre los datos donde las personas no conocieran sus años de escolaridad, así que para los modelos los quite, además, para el sexo genere una nueva variable la cual seria “*genero*”, esta variable estaría haciendo la comparación respecto al hombre, ya que la base de datos original de la ENOE tiene al hombre como 1 y la mujer como 2, entonces por comodidad solo volví a la variable de mujer como 0, esto para facilitar la interpretación de los resultados.

Después de hacer el filtrado de datos, ignorando a los que tienen marcado 99 significando que no saben el total de años de educación formal, nos queda un histograma mejor definido en el cual podemos ver marcado cuando una persona termina la primaria, la secundaria, el bachillerato y la universidad, de tal modo que nos queda una gráfica con tendencia negativa, donde muy pocas personas terminan ciertos niveles de educación posteriores a secundaria.

*Grafica 2: histograma de años de educación*



*Nota: esta es la misma grafica que la grafica 1, pero con la correccion de las personas sin conocimiento de su total de años estudiados*

Para iniciar con el modelo, primero inicié descartando modelos que creo no son útiles para mi investigación, estaba curioso por explorar otras alternativas y ver qué más podía encontrar. El modelo Tobit me llamó la atención, pero al analizar mis datos, noté que la variable de años de educación no tenía muchos ceros o valores negativos. Por eso decidí descartar el modelo Tobit, ya que su principal utilidad es lidiar con variables censuradas, y en mi caso, esa variable no parecía estar censurada.

También consideré el modelo de Box-Cox, pero ya estaba trabajando con variables categóricas, y ese modelo se enfoca principalmente en transformar variables continuas. Por otro lado, el modelo de elección binaria no me pareció adecuado, ya que mi variable dependiente no se expresaba en términos de “sí” o “no”, sino que tenía un rango continuo de valores.

Por esto seleccioné el modelo de Poisson ya que se adapta a mis datos que son enteros no negativos, aunque más adelante veremos que no fue mi modelo final.

Habiendo seleccionado el modelo, lo siguiente fue hacer una primera regresión para saber cómo se ven las estimaciones y poder darles un significado.

Tabla 1: Primer regresión de Poisson

Variable	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	2.7133325	0.0070394	385.452	< 2e-16
genero	-0.2477484	0.0021453	-115.487	< 2e-16
eda	-0.0025111	3.28E-05	-76.561	< 2e-16
e_con	-0.0017269	0.0003081	-5.604	2.09E-08
hij5c	-0.1300489	0.0009623	-135.145	< 2e-16
n_hij	-0.0232053	0.0064427	-6.706	2.00E-11

Para interpretar de manera correcta estos datos, tenemos que aplicar exponencial al valor de estimado y restarle 1 para tener el valor real, pero seria mas eficiente viendo cuales son los valores marginales, con eso podríamos ver mejor las estimaciones, y para eso tenemos los siguientes resultados marginales:

- Genero (genero) = -2.492: Los hombres en promedio, tienen 2.4 años menos de educación formal que las mujeres
- Edad (eda) = -0.02525: En promedio, con cada año de vida hay 0.2% menos años de educación
- Estado Conyugal (e\_con) = -0.01737: En promedio, cada que un individuo pasa por un cambio en el estado conyugal hay un atraso de 0.1% en sus años de educación
- Numero de hijos menores a 5 años (hij5c) = -1.308: Por cada hijo, en promedio hay un atraso de 13% en los años de educación en mujeres (ya que esta variable solo aplica a mujeres).
- Numero de integrantes por hogar (n\_hij) = -0.2345: En promedio por cada integrante del hogar, hay un 23% menos en los años de educación que se puede alcanzar.

Mientras desarrollaba el problema, hice una prueba de Breusch-Pagan para ver la teoría, me dio el siguiente resultado:

data: m3 BP = 17745, df = 6, p-value < 2.2e-16

Este resultado es un valor muy bajo para la prueba p, por ende, rechazamos la hipótesis nula de homocedasticidad, lo que quiere decir que en el modelo hay errores de heterocedasticidad y para abordarlos lo primero que hice fue hacer una matriz de covarianzas robustas, lo cual debe tomar en cuenta que hay heterocedasticidad en el modelo.

*Tabla 2: Matriz de covarianzas robustas*

Columna1	(Intercept)	e_con	hij5c	n_hij	genero	eda
(Intercept)	8.49E-05	-1.09E-06	-2.94E-06	-7.15E-05	-6.50E-06	-5.31E-08
e_con	-1.09E-06	1.59E-07	1.89E-07	1.22E-07	4.57E-08	-2.46E-09
hij5c	-2.94E-06	1.89E-07	3.00E-06	-9.03E-08	1.95E-06	-9.61E-08
n_hij	-7.15E-05	1.22E-07	-9.03E-08	7.02E-05	-4.89E-08	1.31E-08
genero	-6.50E-06	4.57E-08	1.95E-06	-4.89E-08	1.19E-05	5.39E-08
Eda	-5.31E-08	-2.46E-09	-9.61E-08	1.31E-08	5.39E-08	6.57E-09

La matriz de covarianza corregida permite obtener errores estándar robustos y, por lo tanto, inferencias más precisas incluso cuando se viola la suposición de homocedasticidad. Esto es crucial para garantizar que los intervalos de confianza y los valores p de los coeficientes del modelo sean fiables.

Ahora corrí un modelo de prueba z tomando las varianzas corregidas y poder estimar un modelo que lo tenga en cuenta esto.

*Tabla 3: Prueba z con las covarianzas corregidas*

Variable	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	2.7075	0.0092146	293.8315	< 2.2e-16
e_con	-0.002924	0.00039862	-7.3353	2.21E-13
hij5c	-0.14647	0.0017325	-84.5433	< 2.2e-16



n_hij	-0.04319	0.0083801	-5.1538	2.55E-07
genero	-0.20888	0.0034429	-60.6702	< 2.2e-16
Eda	-0.0014055	8.10E-05	-17.3443	< 2.2e-16

Este vendría siendo el modelo final para la investigación, pero aun no es todo, había pasado por alto la mejor prueba que hay para determinar si el modelo se puede hacer por Poisson, así es, la prueba de si la varianza es igual a la media. Calculando la media y la varianza de años de educación formal salió lo siguiente:

- Media = 8.2657
- Varianza = 30.7312

Podemos ver como la varianza es alrededor de 3.5 veces mas grande que la media, por lo tanto, hay una **sobredispersión** de los datos, esto nos dice que muy probablemente el modelo de Poisson no es el mejor para estos datos y esto nos lleva al que para mi fue el modelo final y el que mejor explicaría los datos.

### Modelo final: Binomial Negativa

El modelo de la binomial negativa es una extensión del modelo de Poisson que permite que la varianza sea mayor que la media de mi variable dependiente, por lo tanto, podría contra el problema de sobredispersión. Así es como tenemos la primera regresión con este modelo.

*Tabla 4: Regresión por binomial negativa*

		Estimate	Std. Error	z value	Pr(> z )
<b>1</b>	<b>genero</b>	0.266000	0.003126	85.082	2e-16 ***
<b>2</b>	<b>eda</b>	0.002841	0.000047	60.388	2e-16 ***
<b>3</b>	<b>e_con</b>	0.001626	0.000445	3.655	0.000257 ***
<b>4</b>	<b>hij5c</b>	0.138700	0.001371	101.158	2e-16 ***

		Estimate	Std. Error	z value	Pr(> z )
<b>5</b>	<b>n_hij</b>	0.024680	0.009202	4.855	1.2e-06 ***

Ahora teniendo estos coeficientes sería tan sencillo como sacar exponencial o marginales al modelo y así procedí a hacerlo, pero había un error, calcular las marginales resultaba en error, investigue las causas y esto se debe principalmente a problemas de colinealidad entre mis variables, había una que daba errores así que procedí a hacer una prueba de colinealidad.

• Error en `vif.default(m3_nb)`: there are aliased coefficients in the model

Este error me decía que no se podía calcular porque una de mis variables estaba altamente alineada con otra, la solución estaba en eliminar el género, parece que esta variable ya esta siendo explicada por una de las otras, lo mas probable, por la variable `hij5c` que solo aplicaba para mujeres. Ya habiendo eliminado la variable género, volví a correr la prueba y esta vez hubo un mejor resultado:

Tabla 5: Resultados de la prueba de colinealidad

<b>eda</b>	<b>e_con</b>	<b>hij5c</b>	<b>n_hij</b>
<b>1.815582</b>	1.136705	7.956447	7.538601

Con esta prueba, mientras los coeficientes sean menores a 10, indica que no hay problemas de colinealidad entre mis variables y así se puede seguir con las regresiones de manera segura y así poder hacer una mejor estimación, hay dos valores que se acercan a tener linealidad que seria los hijos menores a 5 años y la cantidad de personas por hogar, ya que en la cantidad de integrantes del hogar se podría decir que ya van incluidos los hijos.

Continuando con lo que tenemos, la regresión si la variable “genero” da lo siguiente:

Tabla 6: regresión por binomial negativa corregida

Variable	Estimate	Std. Error	z value	Pr(> z )
Intercepto	2.789	0.01397	199.541	< 2e-16 ***
eda	-0.001816	0.00007478	-24.278	< 2e-16 ***
e_con	-0.01125	0.0006263	-17.967	< 2e-16 ***
hij5c	-0.1591	0.001651	-96.362	< 2e-16 ***
n_hij	-0.202546	0.01302	-3.491	0.000481 ***

Como vemos en nuestra tabla, se encuentran todos los datos con un valor p muy bajo, lo que indica que son estadísticamente relevantes para nuestra investigación. Con esto podemos aplicar las marginales para poder hacer una mejor interpretación:

*Tabla 7: Resultados marginales finales de la binomial negativa*

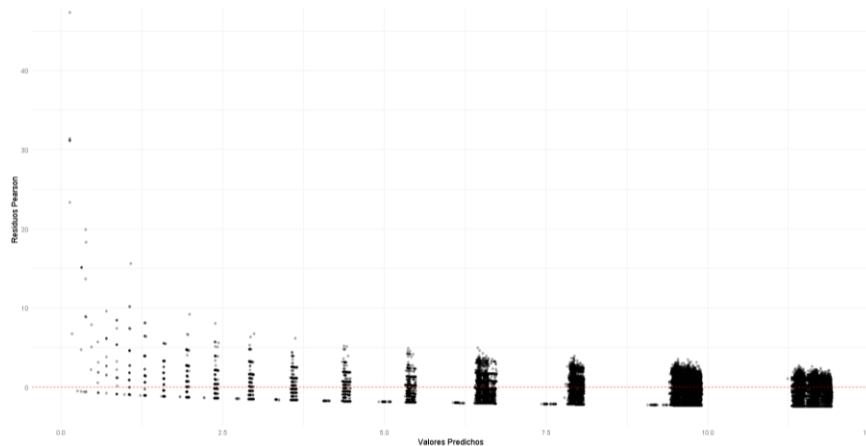
eda	e_con	hij5c	n_hij
-0.001778	-0.03924	2.211	-2.023

Esto lo podemos interpretar de la siguiente manera:

- **Edad:** En promedio por cada año de vida, hay una disminución de 0.1% en años de educación, para nuestra muestra seria que, mientras mas avanzamos en las edades de nuestra muestra, será mas probable ver que mientras avanzamos por la edad los adultos mayores tengan menos años de estudio, que los actuales jóvenes.
- **Estado conyugal:** En promedio, cada vez que las personas cambian de estado conyugal, hay un 3% de atraso en los años de educación, respecto a alguien que se mantiene en su estado conyugal. Esto se puede ver reforzado ya que el proceso de divorcio puede ser emocionalmente agotador. Esto puede llevar a una disminución en la concentración y el rendimiento académico, especialmente si el exesposo está en un proceso de formación o educación continua.
- **Hijos menores a 5 años por mujer:** En promedio por cada hijo las mujeres tienen 2.21 años de educación mas que las mujeres que no tienen hijos. Esto es algo complejo, ya que si bien puede pasar que en algunas investigaciones sugieren que las mujeres que son madres pueden experimentar un aumento en la motivación para completar su educación, ya que desean proporcionar un mejor futuro para sus hijos. Esta motivación puede traducirse en un mejor rendimiento académico, pero también puede que sea un error de la regresión y aunque este es un fenómeno muy complejo y multifacético que cambia la manera en que se puede interpretar.
- **Personas por hogar:** Por cada integrante más en el hogar hay una disminución de 2 años en la educación que puede alcanzar una persona con un integrante menos en el hogar. En general, los hogares con muchos hijos pueden enfrentar limitaciones en recursos económicos y tiempo. Esto puede traducirse en menos inversión en educación, como tutorías, actividades extracurriculares o materiales escolares.

Ahora veamos las gráficas de los valores predichos vs los valores de Pearson que es la mejor manera para ver los valores en una binomial negativa:

*Grafica 3: Residuos de Pearson vs Valores predichos*

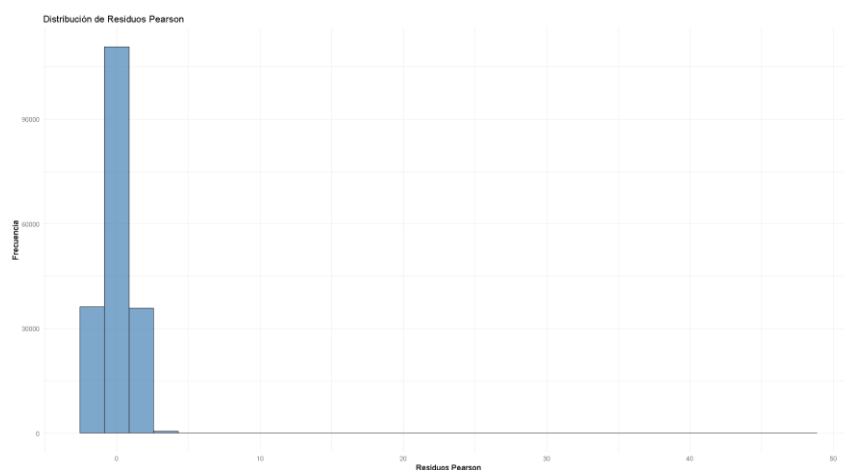


Vemos como evidentemente hay una especie de tendencia, y estos parecen estar alrededor del cero conforme avanza la muestra, hay cierta evidencia de **heterocedasticidad** (varianza no constante), especialmente en valores predichos más pequeños:

En valores predichos cercanos a 0 y 2, los residuos tienen una mayor dispersión (en valores de hasta 40). En valores predichos más grandes (cercanos a 10), los residuos se concentran y tienen menor dispersión.

Esto podría indicar que el modelo tiene problemas para explicar completamente los valores en el rango inferior (valores predichos pequeños).

*Grafica 4: distribución de los residuos*



Concentración alrededor de cero: La mayoría de los residuos de Pearson se agrupan cerca de cero, lo cual es positivo. Esto sugiere que el modelo generalmente se ajusta bien a los datos observados.

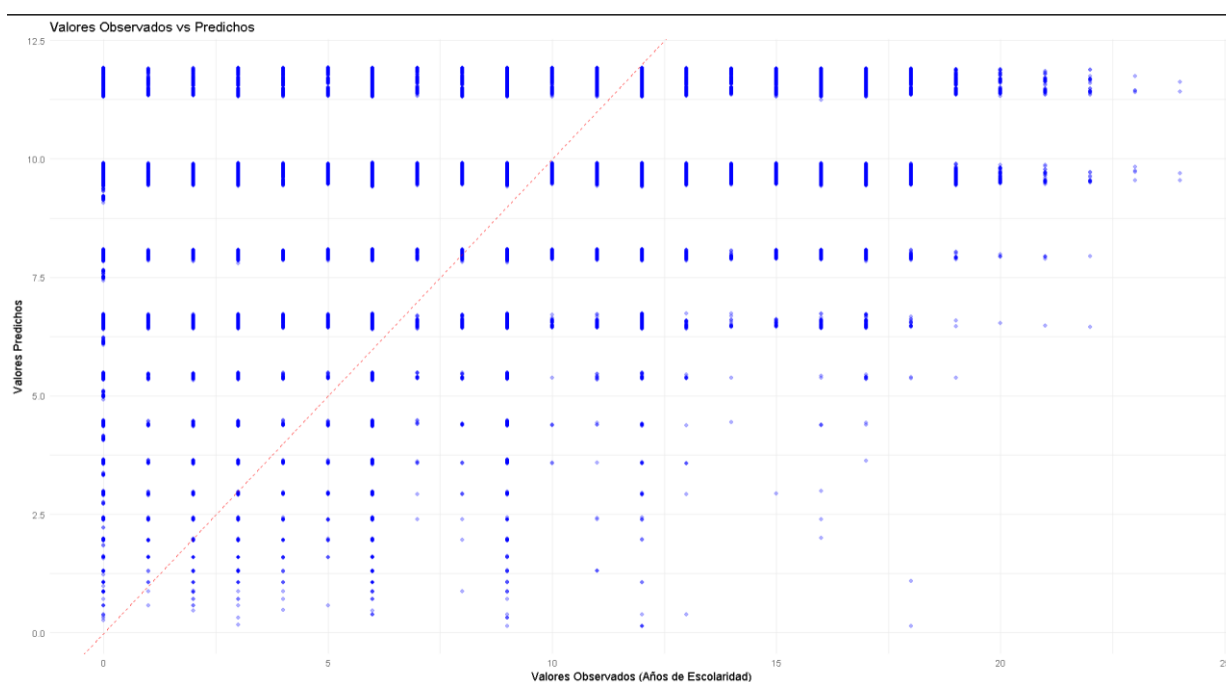
Asimetría: Aunque la concentración es fuerte alrededor de cero, también hay una cola hacia la derecha (valores positivos altos). Esto podría indicar la presencia de valores atípicos o puntos que no se ajustan adecuadamente al modelo.

Escala de los residuos: Los valores altos de los residuos (alrededor de 20-40) son menos comunes, pero su existencia sugiere que algunos puntos extremos están afectando significativamente el ajuste del modelo. Estos podrían ser casos de datos atípicos o errores no capturados por el modelo.

La mayoría de los residuos se encuentran cerca de cero, lo que sugiere que el modelo binomial negativo está capturando adecuadamente la relación en la mayoría de las observaciones. Sin embargo, la presencia de residuos grandes indica que hay ciertas observaciones en los datos que no están siendo explicadas de manera adecuada por el modelo. Esto podría ocurrir por:

- Observaciones atípicas.
- Problemas en la especificación del modelo (falta de una variable relevante).
- Heterocedasticidad residual no modelada.

*Grafica 5: Valores predichos vs Valores reales*



Esta grafica parece tener un efecto rendija porque la variable “años de escolaridad” es una variable discreta que al ser modelada por un modelo como el de binomial negativo que tiende a graficar variables continuas y sumando que la variable tiene cierto rango de enteros, es por esto que se produce esta peculiar gráfica, aunque si la vemos bien para la mayoría de valores se hace una buena estimación dado que los puntos parecen estar alrededor de los observados tan solo con

las variables de edad, cantidad de hijos menores a 5 años, estado conyugal y número de integrantes del hogar.

Ahora viendo la muestra llegue a otro análisis: **¿Cómo es que los ceros pueden afectar el análisis?**

- **Efecto desproporcionado en los resultados:** Los ceros pueden sesgar los estimadores porque tienden a bajar los valores promedio. Esto puede hacer que las relaciones entre las variables explicativas y los años de estudio no representen correctamente a las personas que tienen al menos algún nivel educativo.
- **Diferencias en los procesos que generan ceros y valores positivos:** Si las personas con cero años de escolaridad están influenciadas por factores estructurales (como la pobreza extrema o la marginación), es posible que no compartan el mismo proceso generador de datos que aquellas con más de cero años de escolaridad. Por lo tanto, incluir ambos grupos en el mismo modelo puede no ser adecuado.

Además, considero que los ceros corresponden a un grupo específico con características extremas (por ejemplo, pobreza extrema o falta total de acceso a la educación), y no son lo suficientemente comparables con el resto de la población, por esto tiene sentido excluirllos del análisis principal.

Así, los factores como número de hijos o tamaño del hogar solo se analizarán en el contexto de las personas que sí tienen algún nivel educativo por lo menos un año.

De igual modo comparando los modelos, las graficas de predicho vs observada, tienen un peor ajuste que las graficas donde incluimos a la población con cero años de educación formal.

A manera de recomendación, otra metodología que podría enriquecer este análisis es el uso de modelos de regresión por cuantiles. Esta técnica nos permite explorar cómo los factores socioeconómicos y demográficos afectan no solo el promedio de los años de escolaridad, sino también su distribución en diferentes niveles, como el percentil 25, la mediana o el percentil 75. Esto es especialmente útil en contextos donde las desigualdades educativas pueden variar significativamente entre quienes tienen menos y más años de escolaridad.

Por ejemplo, mientras que los modelos como el de Binomial Negativa se enfocan en la media de la variable dependiente, los modelos de cuantiles pueden mostrar si ciertos factores, como el tamaño del hogar o el estado civil, tienen un impacto más fuerte en los grupos con niveles de escolaridad más extremos. Esto nos ayudaría a identificar subpoblaciones específicas que podrían beneficiarse de políticas públicas más focalizadas. Además, la regresión por cuantiles es robusta frente a valores atípicos, lo que podría resolver algunos problemas como los residuos grandes y la heterogeneidad en la muestra. Incorporar este enfoque en futuras investigaciones permitiría un análisis más profundo de las desigualdades educativas en México.

## **Análisis del Modelo: Fortalezas y Limitaciones**

Este análisis tiene varias fortalezas, como la selección adecuada de modelos para manejar datos que son discretos y no negativos, y la implementación de técnicas para corregir problemas de heterocedasticidad y colinealidad. Sin embargo, también encontré que algunas limitaciones importantes que debemos tener en cuenta al interpretar los resultados.

Primero, detecté sobredispersión en los datos, lo que llevó a usar el modelo de Binomial Negativa en lugar del modelo de Poisson. Aunque este modelo es mejor para manejar la sobredispersión, todavía es posible que haya otras especificaciones más adecuadas. Además, tuve que excluir ciertas variables como "género" debido a colinealidad, lo que sugiere que podría haber factores no modelados que están afectando los resultados.

También note que algunos residuos grandes y dispersión asimétrica en los valores predichos indican que no todos los puntos extremos o valores atípicos fueron explicados adecuadamente. Esto podría deberse a la falta de variables explicativas relevantes en el modelo o problemas en la calidad de los datos de la ENOE, como errores de medición o sesgos en las respuestas.

Finalmente, al excluir a las personas con cero años de escolaridad, el modelo se centra en una población específica y no capta plenamente las dinámicas que afectan a aquellos en condiciones extremas de marginación. Esto limita la generalización de nuestros hallazgos y su aplicabilidad para formular políticas públicas dirigidas a esta población.

## Conclusión

En mi trabajo, me propuse explorar cómo los factores socioeconómicos y demográficos influyen en los años de educación formal en México. Utilicé los datos de la Encuesta Nacional de Ocupación y Empleo (ENOE) y un enfoque econométrico para analizar variables clave como la edad, el estado civil, el número de hijos menores de cinco años y el tamaño del hogar, evaluando su impacto en los años de escolaridad alcanzados. Tras identificar sobredispersión en los datos, seleccioné el modelo de Binomial Negativa como el más adecuado.

Los resultados mostraron patrones claros y estadísticamente significativos. Descubrí que cada año de vida adicional reduce en promedio los años de educación en un 0.1%, reflejando cómo las generaciones mayores tuvieron menos acceso a la educación formal. Además, los cambios en el estado civil, como el divorcio, reducen la escolaridad en un 3%, probablemente debido a las interrupciones emocionales y económicas que estos eventos provocan. Asimismo, identifiqué que en hogares más grandes, la escolaridad disminuye en promedio dos años por cada integrante adicional, lo cual evidencia la presión que los recursos limitados ejercen en familias numerosas.

Una observación interesante fue que las mujeres con hijos menores de cinco años tienen en promedio más años de escolaridad que aquellas sin hijos pequeños. Esto podría estar relacionado con una motivación sociocultural compleja para garantizar un mejor futuro para sus hijos. No obstante, esta interpretación debe tomarse con cautela, ya que podría estar influenciada por errores en el modelo o características específicas de la muestra.

El estudio también destacó algunas limitaciones importantes. La presencia de valores atípicos y residuos grandes sugiere que el modelo no captura completamente todas las dinámicas subyacentes, indicando la necesidad de incluir variables adicionales como ingresos del hogar, acceso a infraestructura educativa o calidad de los servicios educativos. Además, al excluir a las personas con cero años de escolaridad, el análisis no aborda las barreras estructurales que estas poblaciones enfrentan, limitando la aplicabilidad de los hallazgos a políticas públicas que busquen reducir las desigualdades más extremas.

Para finalizar, mi investigación ofrece una visión valiosa sobre los factores que determinan los años de educación formal en México, destacando la necesidad de políticas públicas que consideren el impacto del tamaño del hogar y el estado civil en el acceso a la educación. Sin embargo, para lograr un análisis más completo, futuros estudios deberán incluir a las poblaciones marginadas y explorar variables que capturen mejor las desigualdades estructurales. Solo mediante un enfoque multidimensional podremos diseñar estrategias efectivas para cerrar las brechas educativas y fomentar el desarrollo social y económico del país.



## Referencias

Susana, N. M. C., Caridad, P. C., & Marta, C. P. (n.d.). *Consecuencias del divorcio-separación en niños de edad escolar y actitudes asumidas por los padres.*

[http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-21252017000300003](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21252017000300003)

Instituto Nacional de Estadística y Geografía (Ed.). (2023). *Conociendo la base de datos.*

[https://www.inegi.org.mx/contenidos/programas/enoe/15ymas/doc/con\\_basedatos\\_proy2010.pdf](https://www.inegi.org.mx/contenidos/programas/enoe/15ymas/doc/con_basedatos_proy2010.pdf)

De Estadística Y, I. N. (n.d.). *Encuesta Nacional de Ocupación y Empleo (ENOE), población de 15 años y más de edad.* <https://www.inegi.org.mx/programas/enoe/15ymas/#microdatos>

*Lactancia materna y deserción escolar en estudiantes madres de una facultad de salud* / *Revista Repertorio de Medicina y Cirugía.* (n.d.).

<https://revistas.fucsalud.edu.co/index.php/repertorio/article/view/945/1171>

Fiex, O. (2024, March 26). *Diversidad familiar: los diferentes tipos de familia.* Observatorio FIEX. <https://observatoriofiex.es/diversidad-familiar-los-diferentes-tipos-de-familia/>

Plitt, L. (2023, September 29). *Hijos únicos: qué dice la ciencia realmente sobre el impacto que tiene en los niños no tener hermanos.* BBC News Mundo.

<https://www.bbc.com/mundo/articles/cd121ey7xmmo>