

# **Escolaridad en Mexico: Una exploración usando redes bayesianas**

Santiago Mora, Gabriel Reynoso, Guillermo Villegas

Desde el siglo XX la educación ha sido causa de múltiples políticas públicas de parte del gobierno mexicano, sin embargo los años de escolaridad en México siguen siendo menores que los del promedio de la OCDE. El siguiente artículo tiene como finalidad utilizar información recopilada en la ENAPE de 2021 y explorar las relaciones entre la permanencia educativa y otros factores, esto a través del uso de Diagramas Acíclicos Dirigidos y redes bayesianas. Los resultados revelan una mayor probabilidad de alcanzar una especialidad con un antecedente de educación privada, diferencias entre hombres y mujeres en sus probabilidades de escolaridad y el impacto que tiene el apoyo en casa. Este análisis puede ser útil en la creación de políticas tanto para el sistema educativo así como en el apoyo que se tiene que dar a los padres.

## **Introducción**

En 1948 con la Declaración Universal de los Derechos Humanos la educación se vuelve un derecho para todos, para asegurar esto México a recurrido a múltiples estrategias desde la descentralización la Secretaria de Educación Pública hasta numerosos apoyos sociales para que los estudiantes no tengan que interrumpir sus estudios por falta de ingresos (Ricárdez González and Robles Ortiz 2024). Aunque la meta del país son doce grados de escolaridad, el número promedio de grados escolares concluidos de una población, en 2023 solo se habían alcanzado 9.7 (“Matrícula y Escolaridad En México” March 23, 2024). Para mejorar esta situación es necesario implementar políticas públicas no solo dentro del sistema educativo, también dentro de los hogares mexicanos.

La Encuesta Nacional Sobre Acceso y Permanencia en la Educación (ENAPE) realizada en 2021 por el Instituto Nacional de Estadística y Geografía, tiene como objetivo generar información acerca del acceso a la educación, las razones de no inscripción o conclusión de los estudios, además de algunas condiciones educativas de la población de 0 a 29 años (INEGI 2021). A continuación utilizaremos dicha información para explorar las relaciones entre variables que determinan la escolaridad alcanzada por la población mexicana y responder preguntas como las consecuencias de una educación pública y una privada, las diferencias entre sexos en relación

con la escolaridad y el impacto que tiene el apoyo por parte de personas externas a la institución educativa.

## Métodos

Se utilizó la información recopilada en la ENAPE como base para el análisis, de las bases de datos se extrajeron las variables de interés para responder las preguntas propuestas anteriormente. Posteriormente se realizó una investigación para explorar relaciones de dependencia entre dichas variables, para las representaciones de estas relaciones se utilizaron Diagramas Aciclicos Dirigidos (DAG). Con las relaciones probabilísticas de los nodos es posible crear redes bayesianas- Estas fueron desarrollados en la década de 1970 para modelar creencias y se aplican en problemas basados en incertidumbre (Torres Sainz, Pérez Vallejo, and Trinchet Varela 2024). Para conocer la significancia, si una suposición es compatible con lo observado, de las relaciones de dependencia entre variables se realizaron Pruebas de Hipótesis, estas serán:

$$H_0 : \text{Hijo} \perp\!\!\!\perp \text{Padre} \mid \{\text{Demás padres}\}$$

$$H_a : \text{El hijo es dependiente del padre}$$

Con un intervalo de confianza del 95%, si en la prueba se rechaza la hipótesis nula podemos mantener el arco que se está evaluando. Para conocer que tan bien reflejan las DAGs la estructura de las variables podemos obtener el puntaje de la red o *Network Score*. En este caso se utilizó un criterio basado en la verosimilitud llamado Criterio de Información Bayesiano (BIC).

También se utilizó un algoritmo *Hill Climbing*. Este es un algoritmo que busca la combinación de nodos y arcos que obtiene el mejor puntaje *BIC* en el momento de crear el arco (Russell and Norig 1995). Teóricamente este nos proporcionaría un mejor modelo que los que se propusieron anteriormente. Por último para responder las preguntas presentadas se hicieron Inferencias aproximadas con el método de *Logic sampling*.

## Aplicación

### Variables

De la base de datos se extrajeron las siguientes variables

- S (Sexo): **m** = masculino, **f** = femenino
- A (Edad): **infant** = 0-12 años, **teen** = 13-17 años, **yadult** = 18-24 años, **25adult** = 25 años, **adult** = 26+ años
- P (Tipo de educación): **public** = pública, **private** = privada, **otro** = otro

- E (Escolaridad alcanzada): **no** = Sin escolaridad, **pre** = Preescolar, **elem** = Primaria, **jhig** = Secundaria, **high** = Media superior, **uni** = Superior, **post** = Posgrado
- H (Apoyo en tareas escolares): **no** = no recibe ayuda, **dad** = recibe ayuda del padre, **other** = otro

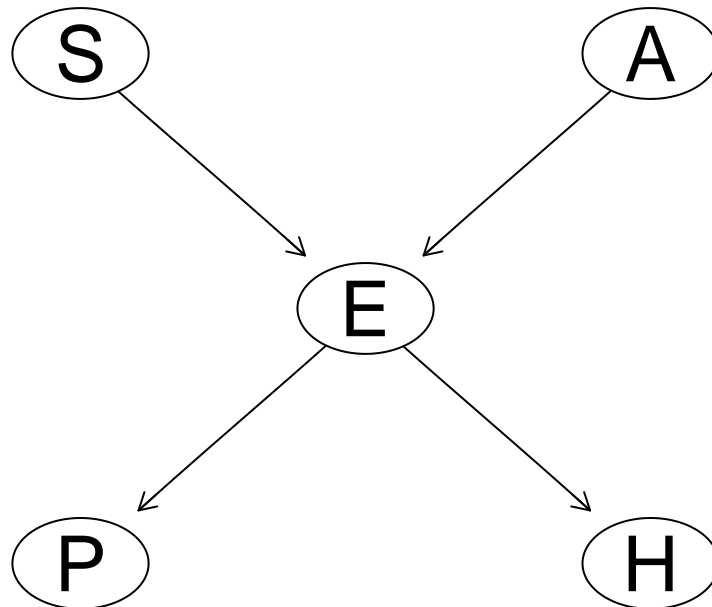
## Diagramas acíclicos dirigidos

De acuerdo a Parker y Pederzini (2000) el sexo y la edad están relacionados con los años de permanencia escolar y, por lo tanto, el nivel de escolaridad que alcanzan. Esto llevó a la decisión de tomar estas dos como nodos padres en los DAGs que se propusieron.

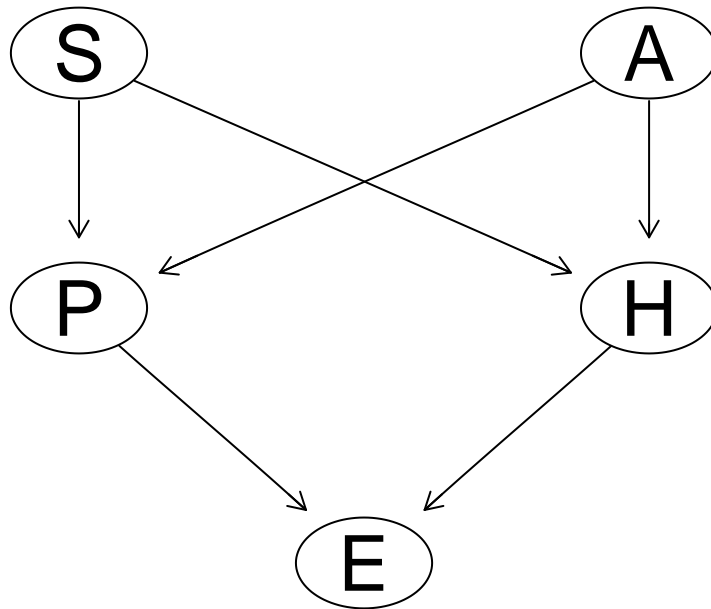
	S	A	P	E	H
1 m	teen	public	elem	other	
2 f	teen	public	uni	other	
3 f	infant	public	elem	other	
4 m	adult	other	high	other	
5 m	adult	private	uni	other	
6 f	yadult	other	uni	other	

### DAG 1

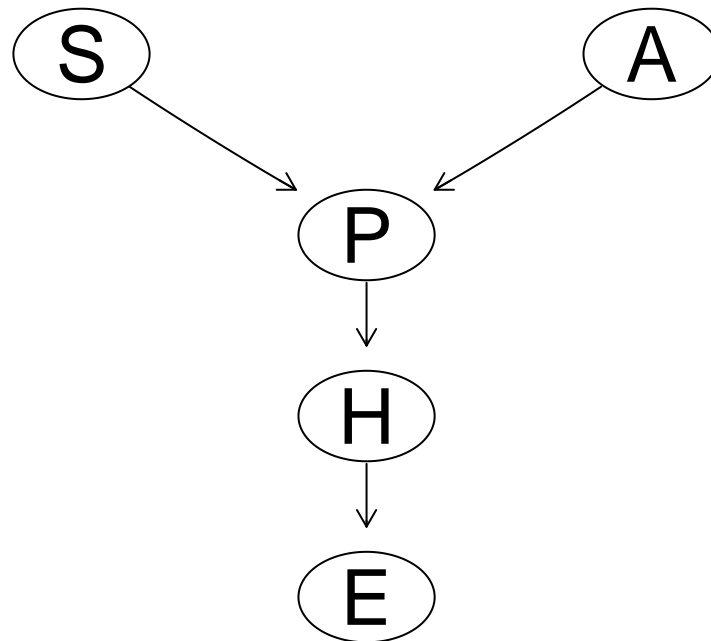
Loading required namespace: Rgraphviz



DAG 2



DAG 3

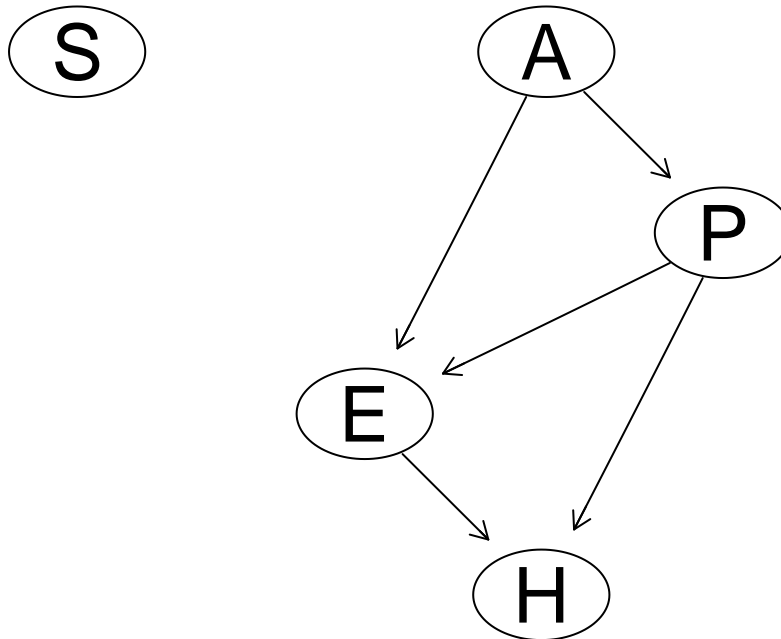


## Puntajes de los DAGs

Diagrama	Puntaje
DAG 1	$-1.3764734 \times 10^5$
DAG 2	$-1.4347211 \times 10^5$
DAG 3	$-1.4760867 \times 10^5$

## Mejor estructura del DAG

Se utilizó el *Hill Climbing* (*HC*) para encontrar una estructura que maximice el *BIC* score. Resultó en el siguiente diagrama:



El *BIC* de esta DAG es  $-1.2996411 \times 10^5$  el cual es mejor que aquellos que presentamos anteriormente, sin embargo el algoritmo *HC* deja fuera el nodo sexo que de acuerdo a algunos estudios (Parker and V. 2000) si tiene impacto en la educación. En base a esto y los puntajes anteriores se decidió usar la DAG 1, para el resto de la exploración.

## Prueba de hipótesis de dependencia entre elementos de DAG

	from	to	strength
1	S	E	0.03100749
2	A	E	0.00000000

3     E   P 0.00000000  
 4     E   H 0.00000000

Rechazamos la hipótesis nula para todas las relaciones de independencia en favor de la hipótesis alternativa. Esto significa que los arcos que hemos establecido previamente si tienen relaciones de dependencia. Es decir, el nivel de educación tiene relación de dependencia con la edad y el sexo, y el tipo de educación y la ayuda recibida tienen relación de dependencia con la educación, dados el sexo y la edad.

## Resultados

**¿Es más probable que un hombre haya hecho la especialidad dado que institución anterior en la que él estuvo inscrito fuera pública o privada?**

$$\mathbb{P}(E = post|P = public, S = m) = 0.0018$$

$$\mathbb{P}(E = post|P = private, S = m) = 0.0147$$

Es más probable que un hombre con educación privada haya hecho la especialidad que uno con educación pública.

**¿Qué tan probable es que una persona a la que nadie ayuda en sus tareas obtenga un nivel de educación igual o superior a bachillerato?**

$$\mathbb{P}(E = [high, uni, post]|H = no) = 0.3563$$

**Probabilidad de que un hombre de 25 años alguna vez haya asistido a la escuela y la Probabilidad de que una mujer de 25 años alguna vez haya asistido a la escuela**

$$\mathbb{P}(S = m, E \neq no) = 0.5143$$

$$\mathbb{P}(S = f, E \neq no) = 0.4847$$

Es más probable que una mujer de 25 años alguna vez haya asistido a la escuela que un hombre de 25 años.

**¿Cuál es la probabilidad que una persona con educación secundaria, lo apoye su papá en sus tareas, proyectos escolares o en resolver sus dudas?**

$$\mathbb{P}(H = dad|E = high) = 0.0187$$

## Conclusiones

Estudiar las relaciones entre variables, tanto con información de profesionales como estadísticamente, permitió formar una red bayesiana con una fuerte dependencia entre variables. De este modelo se pudieron obtener puntos importantes: Aunque es más probable que una mujer de 25 años haya asistido alguna vez a la escuela que un hombre de la misma edad la diferencia entre probabilidades no es tan grande. El antecedente de la educación privada si incrementa la probabilidad de alcanzar una especialidad a comparación de un antecedente de educación pública. Y alguna información sobre el apoyo en las tareas, a nivel secundaria es poco probable el padre ayude en estas, y no es tan probable que una persona que no reciba ayuda en tareas alcance el nivel bachillerato. Esto puede ayudar a orientar las políticas públicas en materia de educación. Aunque útil, este modelo se podría beneficiar de una encuesta más pertinente a los temas que se buscan conocer. También podría ser útil expandir en los factores que influyen las variables de interés.

## Referencias

- INEGI. 2021. “Encuesta Nacional Sobre Acceso y Permanencia En La Educación (ENAPE) 2021.”
- “Matrícula y Escolaridad En México.” March 23, 2024. *CE Noticias Financieras: Spanish*, March 23, 2024.
- Parker, Susan W., and Carla Pederzini V. 2000. “Género y Educación En México.” *Estudios Demográficos y Urbanos* 15 (1 (43)): 97–122. <http://www.jstor.org/stable/40315023>.
- Ricárdez González, Omar, and David Robles Ortiz. 2024. “Las Características de La Educación y Su Efecto En La Pobreza: Un Análisis Para México.” *Análisis Económico* 39 (101): 165–79. <https://doi.org/10.24275/uam/azc/dcsh/ae/2024v39n101/Ricardez>.
- Russell, Stuart J., and Peter Norig. 1995. *Artificial Intelligence: A Modern Approach*. New Jersey: Pearson Higher Education.
- Torres Sainz, Raúl, Lidia María Pérez Vallejo, and Carlos Alberto Trinchet Varela. 2024. “Transformación de Árboles de Fallas En Redes Bayesianas:: Optimización Del Mantenimiento En Sistemas de Ingeniería.” *Desarrollo Sustentable, Negocios, Emprendimiento y Educación* 6 (52): 69–80. <https://doi.org/10.51896/rilcods.v6i52.445>.