

# Redes Bayesianas para el transporte

Jessica Dong

Ebony Valadez

Emiliano Guzmán

Rafael Gutiérrez

2025-08-31

## Resumen

Este estudio utiliza la Encuesta Origen-Destino del INEGI realizada en 2017 para analizar los patrones de movilidad de los ciudadanos en la Ciudad de México. Mediante el uso de la teoría de redes bayesianas, se construye un modelo probabilístico a partir de los datos de la encuesta para analizar las relaciones entre las variables sociodemográficas y de transporte. Se proponen y evalúan diferentes estructuras de Redes Acíclicas Dirigidas (DAGs) para representar estas dependencias. Los resultados del análisis permiten responder a preguntas clave sobre la movilidad. Se busca resaltar la importancia de los modelos probabilísticos para comprender la complejidad de los modelos de transporte en las ciudades mexicanas, con el objetivo de brindar información útil para la planeación de políticas públicas relacionadas con el transporte.

## Tabla de contenidos

<b>1</b>	<b>Introducción</b>	<b>1</b>
<b>2</b>	<b>Metodología</b>	<b>2</b>
2.1	Construcción de la base de datos . . . . .	2
2.2	Redes bayesianas propuestas . . . . .	2
2.3	Selección de la DAG . . . . .	7
<b>3</b>	<b>Aplicación</b>	<b>9</b>
3.1	Uso de auto/moto por sexo y edad (entre semana) . . . . .	9
3.2	Probabilidad de residir en Aguascalientes dado taxi . . . . .	10
3.3	Duración >1 h por tamaño de localidad y modo . . . . .	10
3.4	Autos/camionetas y escolaridad . . . . .	11
<b>4</b>	<b>Conclusiones</b>	<b>12</b>
<b>5</b>	<b>Referencias</b>	<b>12</b>

## 1 Introducción

La movilidad urbana es un aspecto fundamental para el desarrollo socioeconómico de las ciudades modernas, ya que esta influye de manera significativa en la calidad de vida de sus habitantes, el medio ambiente y la economía de cada una. En la Zona Metropolitana del Valle de México, esto presenta una gran complejidad

debido a su tamaño, densidad de población y red de transporte. Comprender los patrones de viaje y los factores que influyen en la toma de decisiones de los ciudadanos es crucial para el diseño de políticas públicas más efectivas y sostenibles.

La Encuesta Origen-Destino en Hogares (EOD) 2017, realizada por el Instituto Nacional de Estadística y Geografía (INEGI) en colaboración con el Instituto de Ingeniería de la UNAM, fue la fuente de datos más importante para este proyecto. A partir de esta encuesta, se obtuvo información que muestra los patrones de movilidad cotidiana de los habitantes de la Ciudad de México y las características de los viajes que estos realizan. Mediante estos datos, es posible analizar las relaciones entre variables sociodemográficas (como género, edad e ingresos) y las decisiones de transporte (como el medio utilizado, la duración y el motivo del viaje) [1].

Este artículo explora estas relaciones de dependencia utilizando la teoría de las **redes bayesianas**. Este enfoque permite modelar la estructura de dependencia probabilística entre múltiples variables, lo que es bastante útil para analizar fenómenos complejos como la movilidad urbana. El objetivo de este estudio es construir y comparar diferentes modelos de Redes Acíclicas Dirigidas (DAGs) que representen las relaciones entre las variables de la encuesta mencionada; para posteriormente, responder preguntas de investigación específicas acerca de la movilidad en la ZMVM.

## 2 Metodología

### 2.1 Construcción de la base de datos

Se integraron cuatro archivos de la EOD-2017 para la ZMVM (ttransporte, tviaje, thogar y tsdem) mediante uniones internas usando las llaves id\_hog, id\_soc e id\_via, filtrando previamente columnas duplicadas para evitar redundancias. A partir de la tabla unificada se seleccionaron las variables relevantes para las consultas (sexo, edad, p5\_14, estrato, p5\_3, ent, tloc, p5\_16\_1\_1, p2\_1\_1 y niv) y se renombraron a descriptores claros: transporte, dia\_viaje, localidad, escolaridad, tiempo\_viaje\_horas, autos\_camionetas y entidad\_origen (manteniendo sexo, edad y estrato). Como control de calidad, se identificaron códigos no informativos; en particular, el valor 99 en tiempo\_viaje\_horas (no sabe/no especificado según INEGI) representó ~1.11% de los registros, por lo que se recodificó a NA y se eliminaron esas filas para no sesgar las estimaciones. La variable entidad\_origen quedó restringida a CDMX, Hidalgo y Edomex por diseño muestral, lo que implica que cualquier consulta que involucre otra entidad arroja probabilidad muestral cero. Con esta base de datos se procedió al aprendizaje de la estructura (DAG), la estimación de las CPTs y la realización de las consultas probabilísticas del estudio.

### 2.2 Redes bayesianas propuestas

	S	T	EST	D	O	L	V	ESC	A	H
1	1	2	3	1	9	1	0	8	2	0
2	1	5	3	1	9	1	0	8	2	0
3	1	14	3	1	9	1	0	8	2	0
4	1	14	3	1	9	1	0	8	2	0
5	1	2	3	1	9	1	0	8	2	0
6	1	5	3	1	9	1	0	8	2	0

Una vez llamada la base de datos, se crea un diccionario y se cambian los valores de las variables a datos de tipo factor, así como al significado correspondiente en texto para un manejo de los datos más sencillo.

	S	T	EST	D	O	L	V	ESC	
"factor"	"factor"	"factor"	"factor"	"factor"	"factor"	"factor"	"factor"	"factor"	
	A	H							
"factor"	"factor"								

	S	T	EST	D	O	L	V	ESC	A
1	Hombre	Colectivo	MedioAlto	Semana	CDMX	Loc_100k+	Autos_0	Licenciatura	Adulto
2	Hombre	Metro	MedioAlto	Semana	CDMX	Loc_100k+	Autos_0	Licenciatura	Adulto
3	Hombre	Caminar	MedioAlto	Semana	CDMX	Loc_100k+	Autos_0	Licenciatura	Adulto
4	Hombre	Caminar	MedioAlto	Semana	CDMX	Loc_100k+	Autos_0	Licenciatura	Adulto
5	Hombre	Colectivo	MedioAlto	Semana	CDMX	Loc_100k+	Autos_0	Licenciatura	Adulto
6	Hombre	Metro	MedioAlto	Semana	CDMX	Loc_100k+	Autos_0	Licenciatura	Adulto
	H								
1	Menos1h								
2	Menos1h								
3	Menos1h								
4	Menos1h								
5	Menos1h								
6	Menos1h								

Posteriormente, se propusieron 3 DAGs diferentes con base en las relaciones entre las variables que se consideraron de mayor impacto, denotando nodos con sus respectivos arcos, así como las relaciones de dependencia. Asimismo, para el manejo de la DAG se asignó una clave corta a cada variable:

- **A** = edad\_cat
- **D** = dia\_viaje
- **EST** = estrato
- **V** = autos\_camionetas
- **L** = localidad
- **H** = tiempo\_cat
- **O** = Entidad
- **ESC** = escolaridad
- **S** = sexo
- **T** = transporte

La DAG 1 se formula con base en que la edad determina la educación y ésta se relaciona con el estrato sociodemográfico de la persona y el tipo de transporte que utiliza, así como arcos en función de cuestiones geográficas y el tiempo de traslado, así como quizá el sexo interfiera con el medio de transporte.

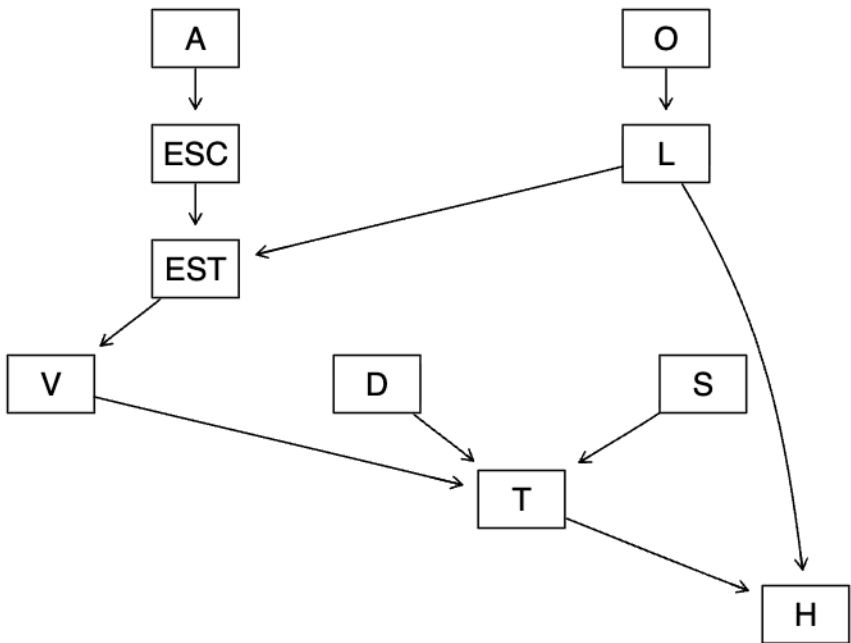
Random/Generated Bayesian network

```

model:
  [A] [D] [O] [S] [L|O] [ESC|A] [EST|L:ESC] [V|EST] [T|D:V:S] [H|L:T]
nodes:
  10
arcs:
  10
  undirected arcs:
    0
  directed arcs:
    10
average markov blanket size:
  3.00

```

average neighbourhood size: 2.00  
average branching factor: 1.00  
generation algorithm: Empty



	from	to	strength
1	A	ESC	0.000000e+00
2	ESC	EST	0.000000e+00
3	EST	V	0.000000e+00
4	V	T	0.000000e+00
5	T	H	0.000000e+00
6	S	T	0.000000e+00
7	D	T	0.000000e+00
8	O	L	0.000000e+00
9	L	EST	0.000000e+00
10	L	H	3.083883e-150

Se observa cómo las relaciones de dependencia son bastante débiles, siendo la mayor la relación entre la localidad de la persona con respecto a las horas que tarda en el transporte; sin embargo, aún sigue siendo muy baja.

La segunda DAG propuesta se centra principalmente en el entorno geográfico, es decir, que del lugar de origen o la localidad se relacionan el estrato, el tiempo del viaje, la cantidad de vehículos, así como otras relaciones entre el sexo y la edad para determinar la escolaridad y el medio de transporte.

Random/Generated Bayesian network

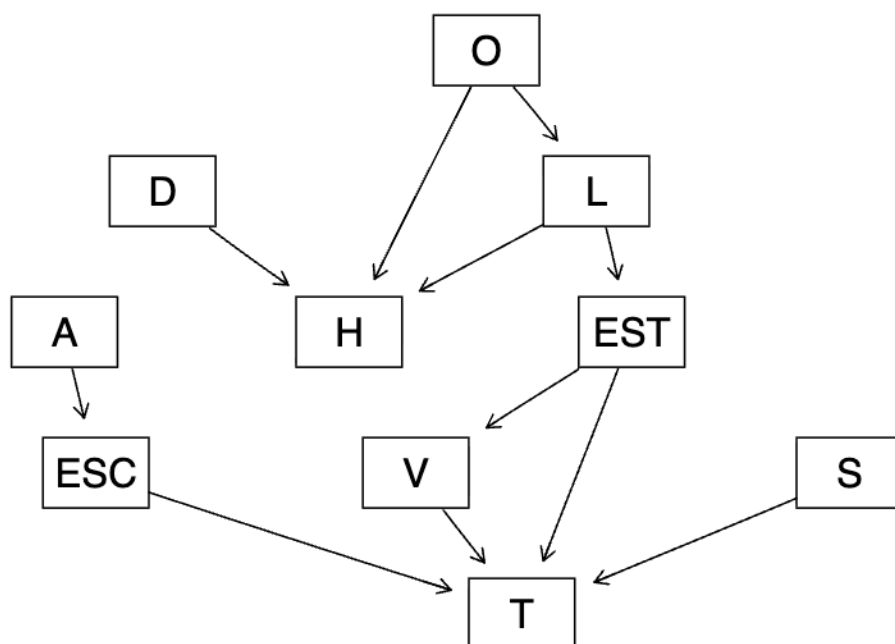
model:  
[A] [D] [O] [S] [L|O] [ESC|A] [EST|L] [H|D:L:O] [V|EST] [T|EST:V:ESC:S]  
nodes: 10

```

arcs:                                11
  undirected arcs:                    0
  directed arcs:                      11
average markov blanket size:          3.60
average neighbourhood size:            2.20
average branching factor:              1.10

generation algorithm:                  Empty

```



	from	to	strength
1	O	L	0.000000e+00
2	L	EST	0.000000e+00
3	EST	T	0.000000e+00
4	O	H	0.000000e+00
5	L	H	2.481191e-195
6	EST	V	0.000000e+00
7	V	T	0.000000e+00
8	S	T	0.000000e+00
9	A	ESC	0.000000e+00
10	ESC	T	0.000000e+00
11	D	H	2.270633e-35

Nuevamente, al analizar las relaciones de dependencia entre las variables, se observa que la dependencia más fuerte va desde el nodo **L** al nodo **H**.

Por último, la tercera DAG se basa principalmente en factores personales como los determinantes de mayor impacto. Teniendo como nodos padres e independientes la edad y el sexo, y que a partir de éstos se condicionen otros aspectos relacionados con la movilidad de las personas, su escolaridad y el estrato en el que se encuentran.

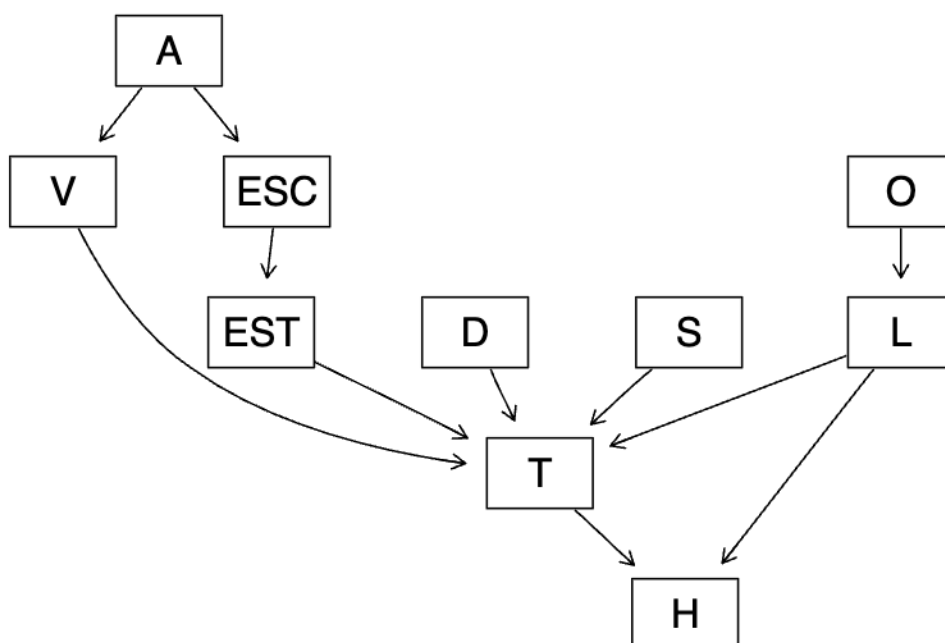
# Random/Generated Bayesian network

```

model:
  [A] [D] [O] [S] [V|A] [L|O] [ESC|A] [EST|ESC] [T|D:EST:V:L:S] [H|L:T]
nodes:
  10
arcs:
  11
  undirected arcs:
    0
  directed arcs:
    11
average markov blanket size:
  4.20
average neighbourhood size:
  2.20
average branching factor:
  1.10

generation algorithm:
  Empty

```



	from	to	strength
1	A	ESC	0.000000e+00
2	ESC	EST	0.000000e+00
3	EST	T	0.000000e+00
4	T	H	0.000000e+00
5	S	T	0.000000e+00
6	A	V	4.784198e-248
7	V	T	0.000000e+00
8	O	L	0.000000e+00
9	L	T	4.880149e-218
10	D	T	0.000000e+00
11	L	H	3.083883e-150

En la tercera DAG destacan las relaciones de dependencia de **A** a **V**, relacionando la edad de la persona con la cantidad de vehículos en su hogar, así como de **L** a **T**, asociando la localidad con el tipo de transporte que utiliza.

## 2.3 Selección de la DAG

Una vez que ya se plantearon las 3 DAGs propuestas, se comparan sus network scores para determinar cuál de las 3 es la que mejor refleja la estructura de dependencia de los datos, para esto se usan las pruebas **BIC** (Bayesian Information Criterion) y **AIC** (Akaike Information Criterion).

El puntaje de la prueba BIC en la primera DAG es:

```
[1] -7863114
```

El puntaje de la prueba AIC en la primera DAG es:

```
[1] -7857363
```

El puntaje de la prueba BIC en la segunda DAG es:

```
[1] -8032815
```

El puntaje de la prueba AIC en la segunda DAG es:

```
[1] -7944145
```

El puntaje de la prueba BIC en la tercera DAG es:

```
[1] -7983492
```

El puntaje de la prueba AIC en la tercera DAG es:

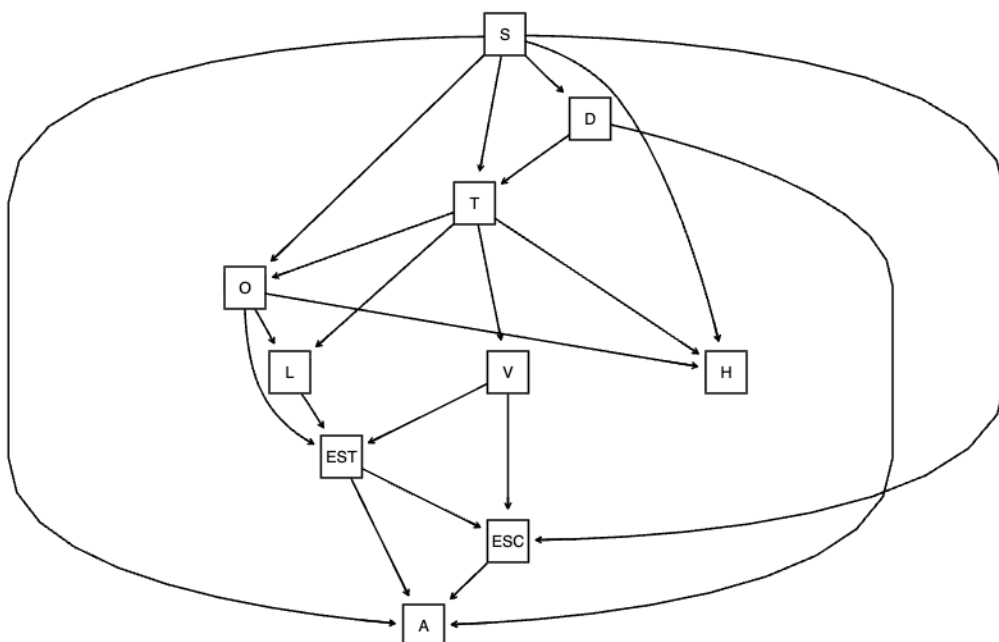
```
[1] -7918357
```

Dado que la primera DAG es la que tiene un mayor puntaje (dentro de la librería bnlearn aquella con mejor ajuste es la que tiene mayor puntaje) en ambas pruebas, entonces es la red que se ajusta mejor a los datos.

De igual forma, se emplea el algoritmo **hill-climbing** para obtener la mejor estructura de la DAG en función de la base de datos.

```
[1] "[S][D|S][T|S:D][O|S:T][V|T][L|T:O][H|S:T:O][EST|O:L:V][ESC|S:EST:V][A|S:EST:D:ESC]"
```

La cual se ve de la siguiente manera, con los arcos creando las relaciones de dependencia de sus determinados nodos.



	from	to	strength
1	ESC	A	0.000000e+00
2	T	V	0.000000e+00
3	T	H	0.000000e+00
4	O	L	0.000000e+00
5	EST	ESC	0.000000e+00
6	L	EST	0.000000e+00
7	V	EST	0.000000e+00
8	V	ESC	0.000000e+00
9	O	EST	0.000000e+00
10	EST	A	0.000000e+00
11	T	O	0.000000e+00
12	S	T	0.000000e+00
13	T	L	0.000000e+00
14	D	T	0.000000e+00
15	S	ESC	0.000000e+00
16	S	A	0.000000e+00
17	O	H	0.000000e+00
18	D	A	0.000000e+00
19	S	O	1.405692e-262
20	S	H	1.017032e-295
21	S	D	2.112147e-52

En esta nueva DAG generada por el algoritmo hill-climbing la fuerza entre los nodos varía, siendo en la mayoría de 0; sin embargo, hay algunas otras relaciones más fuertes como el sexo con si se transportan entre semana o no, o el tiempo que tardan, aunque otras lógicamente no hacen sentido, tal como el sexo con la entidad de origen.

El puntaje en la prueba BIC de la nueva DAG es:

[1] -7781807



El puntaje en la prueba AIC de la nueva DAG es:

[1] -7769003

Se observa que la nueva DAG determinada por hill-climbing tiene un mejor ajuste de su red en la relación de los datos y sus respectivas variables, ya que tiene un puntaje mayor. Por tal motivo, se llevarán a cabo las queries con base en **best\_dag**.

### 3 Aplicación

Trabajamos con la EOD-2017 (ZMVM; ~890 000 registros). Las variables consideradas fueron: S (Sexo), A (Edad categorizada), EST (Estrato), T (Transporte), D (Día de viaje), O (Entidad de origen), L (Tamaño de localidad), V (Autos disponibles), ESC (Escolaridad) y H (Duración del viaje). Tras la limpieza de códigos, recodificamos a factores con etiquetas y discretizamos cuando fue necesario (p. ej., H en *Menos1h* vs *1homas*).

Modelamos con redes bayesianas discretas. Primero propusimos un DAG sustentado en hipótesis sustantivas; al mismo tiempo, aprendimos una estructura desde los datos mediante hill climbing. Para ambas estructuras ajustamos las tablas de probabilidad condicional (CPTs) con suavizado bayesiano.

La selección del modelo se basó en AIC y BIC, buscando el mejor equilibrio entre ajuste y complejidad. Además, evaluamos dependencias locales con `arc.strength` (información mutua) y pruebas de independencia para identificar arcos débiles o prescindibles.

Sobre el modelo seleccionado realizamos las consultas probabilísticas con `cpquery` mediante likelihood weighting, reportando los valores puntuales y señalando las limitaciones muestrales cuando aplicaba.

#### 3.1 Uso de auto/moto por sexo y edad (entre semana)

Se busca comprender si los hombres de edad media y alta, y de estrato social medio bajo tienden a usar automóvil o moto entre semana en comparación con las mujeres con las mismas características.

El evento de interés en esta consulta es si la persona de transporta en automóvil o moto (**T == “Moto”** o **T == “Auto”**); dado que sea hombre (**S == “Hombre”**), se transporte entre semana (**D == “Semana”**), se encuentre en un estrato sociodemográfico medio bajo (**EST == “MedioBajo”**) y sea de edad media o alta (**A == “MedianaEdad”** o **A == “AdultoMayor”**). La probabilidad de que suceda es:

$$P(T \in \{Auto, Moto\} \mid S = \text{Hombre}, D = \text{Semana}, EST = \text{MedioBajo}, A \in \{\text{MedianaEdad}, \text{AdultoMayor}\}) \\ = \frac{\#\{T \in \{Auto, Moto\}, S = \text{Hombre}, D = \text{Semana}, EST = \text{MedioBajo}, A \in \{\text{MedianaEdad}, \text{AdultoMayor}\}\}}{\#\{S = \text{Hombre}, D = \text{Semana}, EST = \text{MedioBajo}, A \in \{\text{MedianaEdad}, \text{AdultoMayor}\}\}}.$$

{Hombre}, D=Semana, EST=MedioBajo, A {MedianaEdad,AdultoMayor}}}. \end{align\*}

[1] 0.1089333

Para analizar este mismo evento, pero ahora con mujeres, únicamente (**S == “Mujer”**) y ahora la probabilidad es:

$$P(T \in \{Auto, Moto\} \mid S = \text{Mujer}, D = \text{Semana}, EST = \text{MedioBajo}, \&A \in \{\text{MedianaEdad}, \text{AdultoMayor}\}) = \\ \frac{\#\{T \in \{Auto, Moto\}, S = \text{Mujer}, D = \text{Semana}, EST = \text{MedioBajo}, A \in \{\text{MedianaEdad}, \text{AdultoMayor}\}\}}{\#\{S = \text{Mujer}, D = \text{Semana}, EST = \text{MedioBajo}, A \in \{\text{MedianaEdad}, \text{AdultoMayor}\}\}}$$

[1] 0.06952873

En conclusión, los hombres de esas características usan más la moto o automovil entre semana que las mujeres con las mismas características.

### 3.2 Probabilidad de residir en Aguascalientes dado taxi

¿Cuál es la probabilidad de que una persona de estrato MedioBajo viva en Aguascalientes dado que se mueve en taxi?

En esta consulta, el evento de interés  $O = \text{Aguascalientes}$  está fuera del soporte muestral de la base utilizada. La EOD-2017 analizada corresponde a la Zona Metropolitana del Valle de México (ZMVM); en este marco, la variable  $O$  (Entidad de origen) solo toma los valores CDMX, Hidalgo y Edomex. Es decir que el dominio observado de  $(O)$  es  $\Omega_O = \{\text{CDMX}, \text{Hidalgo}, \text{Edomex}\}$  por lo que *Aguascalientes* no pertenece al dominio. Esta restricción se confirma con un conteo simple: la frecuencia de registros con  $O = \text{Aguascalientes}$  es 0.

Bajo esta situación, la probabilidad condicional solicitada se evalúa como:

$$P(O = \text{Aguascalientes} \mid EST = \text{MedioBajo}, T = \text{Taxi}) = \frac{\#\{O = \text{Aguascalientes}, EST = \text{MedioBajo}, T = \text{Taxi}\}}{\#\{EST = \text{MedioBajo}, T = \text{Taxi}\}}.$$

Dado que el numerador es 0 (no existen casos con  $O = \text{Aguascalientes}$  en la muestra) y el denominador es positivo (sí hay observaciones con  $EST = \text{MedioBajo}$  y  $T = \text{Taxi}$ ), se obtiene:

$$P(O = \text{Aguascalientes} \mid EST = \text{MedioBajo}, T = \text{Taxi}) = 0$$

Es importante mencionar que se trata de una imposibilidad muestral por ausencia de casos en la muestra, no necesariamente de una imposibilidad poblacional.

### 3.3 Duración >1 h por tamaño de localidad y modo

Nos interesa saber si es más probable que las personas de localidades < 2,500 habitantes realicen viajes de más de 1 hora en automóvil, colectivo/micro, metro, autobús, trolebús, tren ligero, tren suburbano o transporte de personal, en comparación con las personas de localidades > 100,000 habitantes que cumplan con las mismas características.

Para responder lo anterior, se toma en cuenta el evento de interés  $H = \text{"1homas"}$  (tiempo de viaje mayor o igual a una hora), condicionado a dos tipos de evidencia:

1. **Localidad ( $L$ ):** se encuentran las siguientes categorías

$$\Omega_L = \{\text{Loc\_100k+}, \text{Loc\_15k\_99k}, \text{Loc\_2.5k\_14k}, \text{Loc\_menor2.5k}\}$$

2. **Medio de transporte ( $T$ ):** se toman en cuenta solo los medios de interés  
 $\{\text{Auto}, \text{Colectivo}, \text{Metro}, \text{Autobus}, \text{Trolebus}, \text{TrenLigero}, \text{Suburbano}, \text{Personal}\}$

Entonces, la probabilidad condicional es la siguiente:

$$P(H = \text{1homas} \mid L = \text{Loc\_menor2.5k}, T \in \{\dots\}) = \frac{\#\{H = \text{1homas}, L = \text{Loc\_menor2.5k}, T \in \{\dots\}\}}{\#\{L = \text{Loc\_menor2.5k}, T \in \{\dots\}\}}$$

y de la misma manera para  $L = \text{Loc\_100k+}$

[1] 0.1788564

[1] 0.2022569

Los resultados obtenidos con  $n=10^6$  simulaciones fueron:

- $P(H = 1\text{homas} \mid L = \text{Loc\_menor2.5k}, T \in \{\dots\}) \approx 0.18$
- $P(H = 1\text{homas} \mid L = \text{Loc\_100k+}, T \in \{\dots\}) \approx 0.20$

Es decir, bajo el modelo de redes bayesianas, la probabilidad de realizar viajes de una hora o más es ligeramente **mayor en localidades grandes (100k+) que en las localidades muy pequeñas (<2.5k)**.

Así, se observa que la probabilidad de que las personas de una localidad chica hagan más de una hora de viaje en los medios de transporte de interés es menor a la probabilidad de las personas de localidades con más habitantes. Entonces, las personas en localidades grandes tienen ligeramente más probabilidad de hacer viajes largos, que duren más de una hora, que las de localidades muy pequeñas.

### 3.4 Autos/camionetas y escolaridad

¿Cuál es la probabilidad de que una persona tenga al menos un auto o camioneta dado que su nivel de escolaridad es primaria o preescolar?

El evento de interés es que una persona tenga al menos un auto o camioneta, es decir:

$$V \neq 0$$

La evidencia corresponde a que el nivel de escolaridad de la persona sea primaria o preescolar, es decir:

$$\{ESC = \text{Preescolar}\} \cup \{ESC = \text{Primaria}\}$$

De esta forma, la probabilidad condicional puede escribirse como:

$$\begin{aligned} &P(V \neq 0 \mid \{ESC = \text{Preescolar}\} \cup \{ESC = \text{Primaria}\}) \\ &= \frac{\#\{V \neq 0, \{ESC = \text{Preescolar}\} \cup \{ESC = \text{Primaria}\}\}}{\#\{\{ESC = \text{Preescolar}\} \cup \{ESC = \text{Primaria}\}\}} \end{aligned}$$

En este caso, el denominador representa a todas las personas cuya escolaridad es preescolar o primaria, mientras que el numerador cuenta únicamente a aquellas que, además, tienen al menos un auto o camioneta.

[1] 0.2926482

El resultado fue:

$$P(V \neq 0 \mid \{ESC = \text{Preescolar}\} \cup \{ESC = \text{Primaria}\}) \approx 0.292$$

Los resultados muestran que la probabilidad de que una persona con nivel de escolaridad preescolar o primaria tenga al menos un auto o camioneta es cercana al 29%. Esto sugiere que, dentro de los hogares con menor nivel educativo, la disponibilidad de vehículo privado es limitada, lo cual puede tener implicaciones directas en sus patrones de movilidad cotidiana y en el acceso equitativo a oportunidades dentro de la Zona Metropolitana del Valle de México.

## 4 Conclusiones

Este estudio mostró que las **redes bayesianas** constituyen una herramienta poderosa para analizar fenómenos complejos como la movilidad urbana, al permitir representar de forma explícita las dependencias probabilísticas entre variables sociodemográficas y de transporte. A partir de la Encuesta Origen-Destino 2017 del INEGI, se construyeron y evaluaron diferentes estructuras de DAG, identificando que el modelo aprendido mediante **hill-climbing** ofreció el mejor ajuste según los criterios AIC y BIC.

Las consultas realizadas ejemplifican cómo este enfoque permite responder preguntas específicas de interés en movilidad. En particular, se encontró que los **hombres de edad media o alta y estrato medio-bajo** tienen una probabilidad de aproximadamente **10.9%** de usar automóvil o moto entre semana, mientras que para **mujeres con las mismas características** la probabilidad es menor, de **7.0%**.

Por otro lado, se mostró que la **probabilidad de residir en Aguascalientes dado el uso de taxi** es **0%**, lo que confirma la limitación muestral de la base, que solo incluye observaciones para CDMX, Hidalgo y Estado de México. Finalmente, respecto a la **duración de los viajes mayores a una hora**, se observó que en localidades pequeñas (< 2,500 habitantes) la probabilidad es de **17.9%**, mientras que en localidades grandes (> 100,000 habitantes) es ligeramente mayor, alcanzando un **20.2%**.

Asimismo, se estimó que la probabilidad de que una persona con **escolaridad primaria o preescolar** tenga al menos un vehículo (auto o camioneta) es de aproximadamente **29.2%**, evidenciando una menor disponibilidad de transporte privado en estos grupos.

En conjunto, estos hallazgos reflejan la utilidad de los modelos probabilísticos para comprender los patrones de movilidad cotidiana y la influencia de factores sociodemográficos. Tales modelos no solo permiten explorar dependencias estadísticas, sino también aportar evidencia valiosa para el diseño de políticas públicas de transporte más inclusivas y sostenibles en la Zona Metropolitana del Valle de México.

## 5 Referencias

Instituto Nacional de Estadística y Geografía. (2017). Encuesta Origen-Destino en Hogares de la Zona Metropolitana del Valle de México (EOD) 2017. INEGI. <https://www.inegi.org.mx/programas/eod/2017/>

Instituto Nacional de Estadística y Geografía. (2018). Presentación de resultados de la Encuesta Origen-Destino en Hogares de la Zona Metropolitana del Valle de México (EOD) 2017. INEGI. [https://www.inegi.org.mx/contenidos/programas/eod/2017/doc/resultados\\_eod\\_2017.pdf](https://www.inegi.org.mx/contenidos/programas/eod/2017/doc/resultados_eod_2017.pdf)