

### **III. Metodología**

Los objetivos de este trabajo son identificar en qué consiste la problemática del sistema de transporte, identificar cuáles son las variables que influyen en el comportamiento al momento de elegir un modo de transporte y saber cuáles influyen más en este comportamiento. A continuación se menciona la manera en la que se probaron nuestras hipótesis y lograr nuestro objetivos.

#### **1. Muestra**

La muestra consiste de 2200 observaciones que corresponden a individuos de las entidades del DF y los municipios del Estado de México que conforman la ZMCM. Esta muestra fue tomada de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)<sup>1</sup> para el año 2000 que elabora el INEGI. De esta encuesta, se utilizaron los capítulos de Población, Hogares, Ingresos y Gastos para obtener las variables necesarias para el modelo. La muestra consta de 1135 mujeres y 1065 hombres, 51.59% y 48.41% respectivamente.

#### **2. Modelos**

Como pudimos ver en la sección anterior, los autores citados son ejemplo de diversos estudios referentes a los modelos de decisión de modo de transporte. Observamos también que coinciden en algunas variables utilizadas en sus modelos. Con base en esto se construirá un modelo similar utilizando datos para la ZMCM.

La ENIGH es una encuesta que nos ayuda a obtener una muestra aleatoria y que proporciona datos que pueden ser utilizados para la construcción de un modelo de demanda de transporte. De ésta, obtuvimos la siguiente información: gasto en los distintos modos de transporte, gasto en gasolina, ingreso trimestral, edad, sexo, estado civil, educación formal, trabajo que realiza, estrato de la zona habitacional, ubicación geográfica y posesión de

---

<sup>1</sup> La información de cómo se lleva a cabo la recolección de los datos en la ENIGH se puede encontrar en el mismo documento. Ver ENIGH, 2000.

automóvil. Esta información, aunque se recolecta entrevistando distintos hogares, es de manera individual.

Cabe señalar que, como vimos anteriormente, la mayoría de los modelos utilizan variables de tiempos y de costos del transporte, sin embargo, en este caso no se incluye esa información dado que no se encuentra en la ENIGH. Por esto, este trabajo no explica todo lo que ocurre en cuanto a la decisión del modo de transporte, pero sí da una aproximación a la explicación de cómo es que las características socioeconómicas mencionadas determinan el modo de transporte a utilizar.

Otras variables que omitimos en este trabajo fueron los precios de los servicios de transporte, y es que tratándose de un estudio para un solo año, los precios no muestran ninguna variación. Aun cuando se contemplaran distintos años en este trabajo, los precios del transporte público no han cambiado significativamente en los últimos años<sup>2</sup>, por lo que incluir estas variables no nos daría información adicional.

Para este estudio, trabajamos con dos tipos de modelos, de manera de poder comparar si las mismas variables tienen el mismo efecto en los dos.

## 2.1 Modelos lineales

Los primeros modelos consisten en regresiones lineales utilizando el método de mínimos cuadrados para probar que tan significantes son las variables socioeconómicas con cada uno de los distintos medios de transporte urbano y conocer cómo responde la demanda por transporte a cada una de éstas. Los medios de transporte son: Metro, autobús, trolebús y tranvías, colectivos o peseros, y taxi. También aplicamos el mismo modelo al uso del automóvil y a un modelo que contemple a todos los medios de transporte público agregados en una sola variable.

Estas regresiones se asemejan a los modelos de generación de viajes descritos anteriormente. Cabe mencionar que no es de nuestro interés predecir el número de viajes que habrá en un momento determinado sino saber cuáles son las variables que nos permiten explicar el comportamiento de las decisiones de utilizar un modo de transporte en particular.

---

<sup>2</sup> Para ver la evolución de los precios en el transporte público ver anexo B.

El modelo general es el siguiente:

$$y_j = \mathbf{b}_0 + \mathbf{b}_1 \text{ingtri} + \mathbf{b}_2 \text{edad} + \mathbf{b}_3 \text{ed\_formal} + \mathbf{d}_1 \text{sexo} + \mathbf{d}_2 \text{trabajo} + \mathbf{d}_3 \text{vehiculos} + \mathbf{d}_4 \text{estrato} + u$$

En la tabla 14 se muestra la descripción de cada una de las variables.

**Tabla 14. Descripción de las variables**

Variable	Descripción	Tipo
$y_j$ (Gasto en Metro, autobús, trolebús, colectivo, taxi o gasolina) <sup>a</sup>	Gasto realizado trimestralmente en el modo j. (en pesos)	Variable dependiente continua
<i>ingtri</i>	Ingreso trimestral. (en pesos)	Continua
<i>edad</i>	Edad (en años)	Continua
<i>ed_formal</i> <sup>b</sup>	Años de estudio en escuela formal, donde 0 corresponde a ninguna educación hasta 16 que corresponde a postgrado.	Continua
<i>sexo</i>	Género, donde 1 corresponde a hombre y 0 a mujer.	Dummy o binaria
<i>trabajo</i>	Si realiza algún trabajo, donde 1 corresponde a sí y 0 a no.	Dummy o binaria
<i>vehiculos</i> <sup>c</sup>	Posesión de vehículos, donde 1 corresponde a tener uno o más vehículos en el hogar y 0 a ninguno.	Dummy o binaria
<i>estrato</i> <sup>d</sup>	Densidad poblacional, donde 1 corresponde a una localidad de alta densidad y 0 baja densidad.	Dummy o binaria
<i>u</i>	Término de error	Continua

<sup>a</sup> Estas variables no representan directamente el número de viajes que realiza una persona, en vez, se utiliza el gasto en cada uno de los modos como una proxy. En el caso del automóvil, utilizamos el gasto en gasolina para medir el uso de este medio.

<sup>b</sup> Esta variable se podría tratar como categórica pero dada su distribución aproximadamente normal, por simplicidad, se trata como variable continua. Ver anexo C para su distribución.

<sup>c</sup> Esta variable sólo contempla que exista al menos un automóvil en el hogar, sin que sea precisamente esta persona la que haga uso de él. Sólo se contemplan automóviles y camionetas.

<sup>d</sup> En este caso, siguiendo la descripción del INEGI, una localidad de alta densidad se refiere a aquella de 100,000 habitantes o más y las de menos de 100,00 habitantes se reunieron en una sola categoría de baja densidad. Esto se hizo para simplificar el modelo ya que es más sencillo trabajar con una variable binaria, además existían muy pocas observaciones que representaran otro tipo de densidad.

donde  $y_j$  puede tener los siguientes significados dependiendo del modelo:

$gmetro$  = Gasto trimestral en Metro

$gbus$  = Gasto trimestral en autobús

$gtrole$  = Gasto trimestral en trolebús o tranvía

$gpes$  = Gasto trimestral en colectivo o pesero

$gtaxi$  = Gasto trimestral en taxi

$ggas$  = Gasto trimestral en gasolina

El segundo modelo que utilizamos es similar pero en vez de utilizar cada uno de los gastos por modo utilizamos el gasto total en transporte urbano. Es decir, generamos una variable que es la sumatoria de los cinco modos de transporte urbano:

$$gto\_total = \mathbf{b}_0 + \mathbf{b}_1 \text{ingtri} + \mathbf{b}_2 \text{edad} + \mathbf{b}_3 \text{ed\_formal} + \mathbf{d}_1 \text{sexo} + \mathbf{d}_2 \text{trabajo} + \mathbf{d}_3 \text{vehiculos} + \mathbf{d}_4 \text{estrato} + u$$

donde:

$$gto\_total = gmetro + gbus + gtrole + gpes + gtaxi$$

Lo que esperamos obtener es que haya una relación negativa entre el ingreso y los gastos correspondientes a cada medio de transporte, ya que suponemos que estos bienes son inferiores y que al aumentar el ingreso disminuye su consumo; mientras que para el automóvil, esperamos una relación positiva, indicando cierto grado de sustitución. Las variables de edad y de educación esperamos que se comporten de igual manera ya que existe cierta correlación con el ingreso, es decir, al aumentar la edad y la educación esperamos que la gente tenga un mayor ingreso. Con el género, se piensa que existe cierta discriminación en cuanto a lo que gana una mujer por lo que es posible que gane menos que un hombre y por lo tanto esto influya en su comportamiento de elección de medio de transporte. Con la variable *trabajo* esperamos medir si los viajes que se hacen en la ZMCM son en su mayoría viajes al trabajo. Con la posesión de automóviles esperamos que haya un

efecto negativo en lo que se refiere a las regresiones de los medios de transporte público, es decir que el tener un automóvil disminuya el gasto en los medios de transporte; y para la regresión que toma en cuenta el gasto en automóvil, esperamos un efecto contrario, es decir, positivo. Por último, para lo que se refiere a la densidad poblacional, esperamos que para los medios de alta capacidad esta variable tenga un efecto positivo y para los de baja capacidad tenga un efecto negativo.

La tabla 15 presenta la estadística descriptiva para las variables de gasto en transporte público que se utilizan en las distintas regresiones.

**Tabla 15. Descripción estadística del gasto trimestral en transporte público**

Variable	Media	Desviación Estándar	Min	Max
gmetro	34.39	115.83	0	1208.36
gbus	143.57	511.12	0	5754.11
gtrole	1.81	22.45	0	387.87
gpes	581.95	1003.85	0	7544.27
gtaxi	97.24	507.45	0	7416.40
gto_total	858.95	1082.75	0	7544.27

De esta información vemos que el medio en el que se gasta más en promedio es en colectivos. En comparación, el modo de transporte que reporta un menor gasto es el trolebús.

Para comparar estas cifras, de la misma muestra obtenemos lo que los individuos gastan en gasolina, a modo de también utilizarlo como una variable proxy del uso del automóvil. En la tabla 16 se muestra esta información.

**Tabla 16. Descripción estadística del gasto trimestral en gasolina**

Variable	Media	Desviación Estándar	Min	Max
ggas	633.62	1468.37	0	11803.20

De aquí, obtenemos que el promedio de gasto en gasolina es de \$633 que, comparándolo con los otros medios de transporte, es, en promedio, mayor. Sin embargo, existirán individuos que no sólo utilicen un medio de transporte público, por lo que si comparamos

el promedio de gasto en gasolina con el promedio de gasto total en transporte urbano, el uso de automóvil es relativamente más barato.

En la siguiente tabla se presenta la estadística descriptiva para las variables independientes que son continuas.

**Tabla 17. Estadística descriptiva de las variables independientes continuas**

Variable	Media	Desviación Estándar	Min	Max
ingtri	12403.72	16310.27	150	187800
edad	29.91	16.96	5	90
ed_formal	9.46	3.93	1	16

Vemos que el ingreso per cápita promedio es de \$12,403, pero lo más interesante es ver la diferencia entre el mínimo de \$150 y de \$187,800 como máximo. Esto se ve también reflejado en un valor elevado de la desviación estándar, 16,310, que nos habla de la variación o desigualdad que hay en los ingresos.

También observamos que la edad promedio es de casi 30 años, y ésta va desde los 5 hasta los 90 años.

Los años en escuela en promedio son 9, es decir, esto corresponde a tener terminado hasta el 1° año de secundaria<sup>3</sup>.

Por último, la tabla 18 muestra información sobre las variables binarias.

**Tabla 18. Frecuencias y porcentajes de las variables binarias**

Variabes		Frecuencia	Porcentaje
Sexo	Mujer	1135	51.59
	Hombre	1065	48.41
Trabajo	No	1119	50.86
	Sí	1081	49.14
Estrato	Baja dens.	570	25.91
	Alta dens.	1630	74.09
Vehículos	No	1244	56.55
	Sí	956	43.45
	Total	2200	100

<sup>3</sup> La descripción de los años a los que corresponde cada grado se encuentran en el Anexo D.

Como ya habíamos dicho, hay 1135 mujeres y 1065 hombres. El 50.86% del total de individuos no trabajan; la mayoría se encuentran en una zona de alta densidad, y un 56.55% no tiene automóvil en su hogar.

## 2.2 Modelo probit

Para análisis individuales sobre la decisión de utilizar un modo de transporte público en lugar de otro, decimos que la decisión es discreta. Es decir, la variable dependiente toma sólo dos valores, 0 y 1, por lo que un modelo de regresión lineal no nos es tan útil. En cambio, para estos análisis se utilizan modelos logit o probit, siendo estos muy similares. Estos expresan la probabilidad de que una respuesta ocurra como función de una serie de variables explicativas. En este caso utilizaremos un modelo probit.

Después de determinar la relevancia de las variables en las regresiones lineales, se trabajará con un modelo no lineal, un modelo Probit, el cuál nos permite identificar la probabilidad de utilizar un medio de transporte en específico.

Algo importante de aclarar es que en este modelo, para el caso del automóvil, la variable dependiente es una variable binaria que representa si utiliza el automóvil para transportarse, y no es la misma variable que representaba en los modelos anteriores si existía o no un automóvil en el hogar. Para generar esta nueva variable comparamos los gastos totales en transporte público y los gastos en gasolina. Todos los individuos presentaron que sólo incurren en uno de esos gastos, lo que indica que sólo utilizan un medio de transporte: el público o el privado, pero dentro del público pueden utilizar uno o más medios. Por lo tanto, nuestra nueva variable *auto* es una variable binaria que indica si el individuo utiliza o no el automóvil.

Aunque en este trabajo nuestro interés principal está en la probabilidad de utilizar el automóvil, también calculamos las probabilidades de los otros modos, a fin de comparar los resultados con los de las regresiones lineales; por lo que, de igual manera, generamos nuevas variables binarias para cada uno de los modos de transporte. Es decir, cuando los individuos presentan un gasto positivo en un modo, la nueva variable generará un 1, y de manera contraria un 0. De esta forma todas las variables dependientes son discretas.

El modelo general es el que sigue:

$$\Pr(y_i = 1 | X) = \Phi(\mathbf{b}_0 + \mathbf{b}_1 \text{ingtri} + \mathbf{b}_2 \text{edad} + \mathbf{b}_3 \text{ed\_formal} + \mathbf{d}_1 \text{sexo} + \mathbf{d}_2 \text{trabajo} + \mathbf{d}_3 \text{estrato} + \mathbf{d}_4 \text{vehiculo})$$

donde:

$y_i$ : igual a 1 si utiliza el medio  $i$ , 0 de lo contrario.

$\Phi$ : Función de distribución acumulada de la normal estándar.

\* *Todas la demás variables son las mismas que en el modelo lineal.*

El lado izquierdo del modelo nos indica la probabilidad de utilizar el medio de transporte  $i$  dadas las demás variables ( $X$ ). En otras palabras, la distribución logística en el modelo probit da una función de probabilidad de la siguiente manera:

$$P = \frac{1}{1 + \exp(\mathbf{b}_0 + \mathbf{b}_1 x_1 + \dots + \mathbf{b}_n x_n)}$$

Así, tenemos que la transformación inversa de la distribución logística acumulada es:

$$\log\left(\frac{P}{1-P}\right) = \mathbf{b}_0 + \mathbf{b}_1 x_1 + \dots + \mathbf{b}_n x_n$$

Aquí, las betas miden el cambio en los logaritmos de las probabilidades de escoger un modo de transporte.

En estos modelos esperamos que las variables explicativas tengan el mismo efecto que señalamos en los modelos lineales, sólo que esta vez estarán representados por porcentajes.

Por último, en las siguientes tablas tenemos la información de las nuevas variables generadas:



**Tabla 19. Frecuencias y porcentajes de la variable *auto***

Auto	Frecuencia	Porcentaje
No	1662	75.55
Sí	538	24.45
Total	2200	100

Vale la pena comparar esta información con la presentada por la variable *vehiculos*, que representaba que existiera al menos un vehículo en el hogar.

Anteriormente obtuvimos que un 43.45% poseía un automóvil. Ahora podemos ver que un 24.45% utiliza el automóvil como medio de transporte. Esto nos dice entonces que el tener un automóvil no significa que precisamente sea el mismo individuo el que lo utilice. Esto es, el automóvil en el hogar puede pertenecer a otro miembro de la familia.

**Tabla 20. Frecuencias y porcentajes para las variables dependientes binarias**

	Respuesta	Frecuencia	Porcentaje
<b>Metro</b>	No	1959	89.05
	Sí	241	10.95
<b>Autobus</b>	No	1935	87.95
	Sí	265	12.05
<b>Trolebús</b>	No	2183	99.23
	Sí	17	0.77
<b>Colectivo</b>	No	1247	56.68
	Sí	953	43.32
<b>Taxi</b>	No	2014	91.55
	Sí	186	8.45

Vemos que con la información de la muestra obtenemos una participación modal similar a la que presenta la SETRAVI, en donde el modo de transporte más utilizado es el sistema colectivo, con un 43%, y el menos usado es el trolebús, con casi 1%.