



41
2-eg

Universidad Nacional Autónoma de México

FACULTAD DE CIENCIAS

UNA APLICACION DEL MUESTREO PARA GENERAR
INFORMACION SOBRE EL USO DEL AUTOMOVIL
PARTICULAR EN MEXICO

TESIS PROFESIONAL

Que para obtener el Titulo de

A C T U A R I O

P r e s e n t a

MARGRET HEDWIG OETTLER SOMMER

TESIS CON
FALLA DE ORIGEN

MEXICO, D. F.

1 9 9 0



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

INTRODUCCION	1
CAPITULO 1. FUNDAMENTACION TEORICA	5
1.1. Selección de la Muestra	6
1.1.1. Muestreo en dos Etapas	6
1.1.2. Conglomerados desiguales	6
1.1.3. Selección de los Elementos de la Población	10
1.2. El Estimador	11
1.2.1. Presentación	11
1.2.2. Propiedades	14
CAPITULO 2. DISEÑO CONCEPTUAL Y OPERATIVO DE LA ENCUESTA	25
2.1. Población	25
2.2. Marco muestral	31
2.3. Tamaño de la Muestra	35
2.3.1. Marco Teórico	36
2.3.2. Determinación del Tamaño de Muestra .	44
2.4. Cuestionario	59
2.5. Operativo de Campo	63

CAPITULO 3	ANALISIS DE RESULTADOS	66
	3.1. Características del Parque Vehicular	67
	3.1.1. Cilindraje Promedio	68
	3.1.2. Edad Promedio	70
	3.2. Hábitos de Uso del Parque Vehicular	76
	3.2.1. Kilometraje Anual Recorrido	76
	3.2.2. Consumo Semanal	81
CONCLUSIONES		87
ANEXO		89
REFERENCIAS BIBLIOGRAFICAS		94

INTRODUCCION

"En su sentido más amplio la finalidad de una encuesta por muestreo es obtener información para satisfacer una necesidad definida."¹

Esta proposición se va a ilustrar en el presente trabajo en el que se utiliza una encuesta por muestreo para generar información confiable sobre diversas características y hábitos de uso de los automóviles particulares.

Los automóviles particulares participan de manera muy importante en el consumo de energía del sector transporte, ya que constituyen cerca del 95% de los vehículos que transportan pasajeros; sin embargo, realizan sólo el 20% de los viajes urbanos.²

A pesar de la importancia de dicho medio de transporte, la información que se tiene al respecto es muy escasa ya que no se dispone de estadísticas relacionadas al consumo por vehículo, su recorrido anual promedio o algún otro dato corespondiente al uso del parque vehicular.

1. RAJ, Teoría del Muestreo, pag. 34.

2. Datos proporcionados por el Departamento del Distrito Federal.

El registro de características del parque vehicular (cilindrje, modelo, etc.) recopiladas a través del Registro Federal de Automóviles no se actualiza regularmente, por lo que no proporciona datos veraces del parque vehicular en circulación en un momento dado.

Debido a lo expuesto anteriormente, se decidió levantar una encuesta a nivel nacional para suplir la falta de información existente. Esto se hizo mediante la aplicación de un método de muestreo probabilístico.

Dicho método consiste en tomar una muestra de unidades o elementos de la población con las que se realizan inferencias de la población en su totalidad. Se consideran tres componentes básicos en dicho método: selección de la muestra, recolección de los datos e inferencia estadística acerca de la población.³

Para poder llevar a cabo lo expuesto se necesitan emplear métodos de muestreo probabilístico, en los que cada unidad de la población tenga una probabilidad conocida y distinta de cero de ser incluida en la muestra; solo así podemos obtener información válida acerca de nuestra población.⁴

3. RAJ, La Estructura de las Encuestas por Muestreo, pag. 17-18.

4. Ibid, pag. 25.

En este trabajo se presentará el diseño de la encuesta por muestreo realizada para obtener información acerca del automóvil particular, así como algunos de los resultados obtenidos.

La información se captó mediante una encuesta de cobertura nacional en estaciones de servicio, que consistió en colocar un encuestador en cada puesto de venta de gasolina para entrevistar a todos los conductores de los vehículos que acudieran a cargar gasolina en un período determinado. Se pensó en las gasolineras por presentarse ahí las siguientes ventajas:

- Captar en un mismo lugar una gran cantidad de información a un costo relativamente bajo (comparado por ejemplo con encuesta de hogares).
- Aprovechar el tiempo ocioso que pasa el automovilista en la gasolinera sin interrumpirlo en sus actividades.
- Aplicar la encuesta solo a personas que conducen un automóvil y en consecuencia pueden indicar sus hábitos de uso.

A partir de la encuesta se pueden obtener conclusiones de los vehículos que asisten a la gasolinera y por consiguiente se encuentran en movimiento, dichos vehículos constituyen el parque vehicular de interés para el estudio.

Los resultados no son aplicables a vehiculos que aunque estén registrados no se encuentran en circulación.

Otro de los elementos fundamentales del trabajo es la presentación de un estimador que permite obtener información confiable dado el proceso de selección de las unidades en gasolineras, donde no todos los vehículos tienen la misma probabilidad de ser encuestados. Dicho estimador tiene las siguientes propiedades:

- Sólo requiere el conocimiento de la probabilidad de cada elemento de pertenecer a la muestra. Dicha probabilidad se obtuvo a partir de la frecuencia de carga de cada vehículo.
- Converge en probabilidad al parámetro poblacional, es decir, es asintóticamente insesgado.⁶

De esta forma se pretendió con el presente trabajo hacer una aplicación integral del muestreo probabilístico para generar información sobre características y hábitos de uso de los automóviles particulares.

6. Esta propiedad se demuestra en el capítulo 1.

CAPITULO 1
FUNDAMENTACION TEORICA

El objetivo del presente capítulo es mostrar las bases teóricas sobre las que se diseñó la encuesta. Esto es de suma importancia ya que se trata de una encuesta por muestreo y en este tipo de encuesta se incorporan reglas bien definidas para la selección de la muestra. Es decir, que no hay una selección caprichosa de las unidades.

Dichas reglas para seleccionar las unidades que integrarán la muestra se van a precisar en el presente capítulo.

Una vez definido lo anterior se va a precisar también la manera en que se van a inferir conclusiones de la muestra a toda la población bajo estudio. Esto se hace a través de alguna función de los valores muestrales a la que se le denomina estimador.¹

Dicho estimador también se va a presentar en este capítulo, y con ello queda establecida la fundamentación teórica necesaria para el diseño de la encuesta.

1. ABAD, Introducción al Muestreo, pag. 31.

1.1. SELECCION DE LA MUESTRA

1.1.1. MUESTREO EN DOS ETAPAS

El primer punto a considerar para la selección de la muestra es que nos encontramos con un muestreo bietápico. La primera etapa corresponde a la selección de las estaciones de servicio y la segunda a la selección de los automóviles particulares que se encuentran en dicha estación.

A este tipo de métodos de selección se les llama muestreo de conglomerados y se caracterizan porque su unidad de selección contiene más de un elemento de la población; es decir, la unidad de muestreo es un conglomerado de elementos.²

En este caso la unidad de selección es la gasolinera, en ella se van a seleccionar varios elementos de la población que son los automóviles particulares.

1.1.2. CONGLOMERADOS DESIGUALES

En la mayoría de las muestras conglomeradas se trabaja con conglomerados de tamaños desiguales. Esto también sucede en el

2. KISH, Muestreo de Encuestas, pag. 182.

presente estudio, ya que no acude el mismo número de vehículos a cada estación de servicio.

En los conglomerados desiguales la población de $N = X$ elementos se agrupa en A conglomerados, con $N_i = X_i$ elementos en el conglomerado i -ésimo, de manera que: $\sum_1^N X_i = X = N$. Entonces Y_{ij} es el valor de la variable Y para el elemento j -ésimo en el conglomerado i -ésimo; y la población de los valores Y_{ij} se puede representar como:³

$$X_1 \bar{Y}_1 = Y_1 = Y_{11} + Y_{12} + Y_{13} + \dots + Y_{1j} + \dots + Y_{1N_1}$$

$$X_2 \bar{Y}_2 = Y_2 = Y_{21} + \dots + Y_{2j} + \dots + Y_{2N_2}$$

⋮

$$X_i \bar{Y}_i = Y_i = Y_{i1} + Y_{i2} + \dots + Y_{ij} + \dots + Y_{iN_i}$$

⋮

$$X_A \bar{Y}_A = Y_A = Y_{A1} + \dots + Y_{Aj} + \dots + Y_{A N_A}$$

La media por elemento de la población es:⁴

$$\bar{Y} = \frac{Y}{X} = \frac{1}{X} \sum_1^N \sum_j^N Y_{ij} = \frac{1}{X} \sum_1^N Y_i = \frac{\sum_1^N Y_i}{\sum_1^N X_i} \quad (i)$$

3. Ibid, pag. 220.

4. Ibid, pag. 221.

Esta notación es un poco compleja, sobre todo el uso de x para contar elementos: X en lugar de N en la población y x en lugar de n en la muestra. Sin embargo, esta notación sirve para varios propósitos, entre los que podemos mencionar los siguientes:⁵

- El uso de x en lugar de n para el tamaño de muestra destaca el hecho de que la varianza y/x no es simplemente $\text{var } y/x^2$, sino una estadística más compleja.
- Si sencillamente se cuentan elementos en los conglomerados, la variable X_i es uno para cada elemento de la población, y X_i representa el número de elementos en el conglomerado. Sin embargo, $X_i=1$ puede también contar los miembros de cualquier subclase de la población, y hacerse cero para los no miembros; entonces X_i denota el número de miembros de la subclase en el conglomerado.
- Además, X_i no necesita restringirse a variables de conteo; sino que puede denotar los valores de alguna variable auxiliar que se introduzca para mejorar la estimación.

5. Idem.

Hay que considerar que la selección de conglomerados desiguales ocasiona problemas de distinta índole, como son los siguientes:⁶

- El tamaño de la muestra no es fijo convirtiéndose en una variable aleatoria, que depende de la selección al azar de conglomerados mayores o menores.
- El estimador de la media no es un estimador insesgado porque es un cociente de dos variables aleatorias. Sin embargo, como se analizará posteriormente, es un estimador consistente que converge en probabilidad al parámetro poblacional.
- Las fórmulas de la varianza no son estimaciones insesgadas de la verdadera varianza, sino constituyen buenas aproximaciones en muestras bien diseñadas.

Sin embargo, es tan grande la ventaja en cuanto al volumen de información que se puede obtener a un costo relativamente bajo al levantar la encuesta en estaciones de servicio, que a pesar de los problemas que se presentan en la selección de conglomerados desiguales, se decidió usar este método de selección.

6. Ibid, pag. 121 - 122.

1.1.3. SELECCION DE LOS ELEMENTOS DE LA POBLACION

En cada ciudad se seleccionaron las estaciones de servicio mediante un muestreo aleatorio simple sin reemplazo.

Después se levantó un censo en las estaciones de servicio seleccionadas, colocando un encuestador en cada bomba de gasolina de dichas estaciones para entrevistar a todos los vehículos que acudieran ahí en un periodo de tiempo determinado a cargar combustible.

En cada gasolinera se encuestó un día, ya fuera martes, miércoles o jueves desde las 7:00 hasta las 21:00 horas.

De este modo, entre más grande fuera la gasolinera habría más encuestadores y se levantaría un mayor número de encuestas. Esto significa que la probabilidad de seleccionar un cierto automóvil no se vería afectada por la gasolinera en que se le encuestara.

Sin embargo, hay que tomar en cuenta que, toda la información que obtengamos para los elementos de la población, es decir para los automóviles, se verá influenciada por la frecuencia con que éstos asistan a la gasolinera en un cierto periodo de tiempo.

Por ejemplo, un vehículo que recorre 5,000 km. en un año asiste menos frecuentemente a cargar gasolina que uno que recorre

20,000 km. Si no tomamos en cuenta este hecho y calculamos la media directa, estaremos sobreestimando el recorrido anual promedio del parque vehicular.

Es decir, se debe considerar el hecho de que no todos los vehículos van a cargar gasolina con la misma frecuencia y un mismo vehículo puede cargar gasolina más de una vez durante el levantamiento de la encuesta.

Esto lleva a hacer un submuestreo de los elementos de la población con probabilidades desiguales y con reemplazo.

De esta manera queda establecido el criterio de selección de los elementos de la muestra y se puede proponer un estimador adecuado a dicha selección.

1.2. EL ESTIMADOR

1.2.1. PRESENTACION DEL ESTIMADOR

En el presente estudio es de interés conocer algunas de las características y hábitos de uso del parque vehicular como lo son la edad promedio, el kilometraje promedio anual recorrido, el consumo promedio, etc. Todos estos datos que se pretenden obtener con la encuesta se refieren a promedios por los que se va a analizar el estimador de la media de la población.

La media poblacional \bar{Y} de una característica de la población esta dada por:

$$\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N}$$

Donde:

N = Total de elementos en la población

Y_i = Valor de la característica bajo estudio en el i -ésimo elemento

Recordando la ecuación (i) del subcapítulo anterior, la media poblacional en el caso de conglomerados desiguales se puede denotar por:

$$\bar{Y} = \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N X_i} \quad (i)$$

Donde X_i puede denotar los valores de alguna variable auxiliar que se introduzca para mejorar la estimación.

Se propuso entonces el siguiente estimador para valuar la media de la característica Y , considerando que el submuestreo de los elementos de la población es con probabilidades desiguales y con repetición:⁷

7. ARAUJO, Una Aplicación de la Teoría Estadística a la Resolución de Problemas de Información Energética en Países en Vías de Desarrollo, pag. 50.

$$\bar{y} = \frac{\sum_{i=1}^n \left(y_i * \frac{1}{P_i} \right)}{\sum_{i=1}^n \frac{1}{P_i}} \quad (ii)$$

Donde:

n = Total de vehículos en la muestra

P_i = Probabilidad de que el i -ésimo vehículo acuda a una estación de servicio

y_i = Valor de la característica bajo estudio en el i -ésimo vehículo

Para este estimador se necesita conocer la probabilidad que tienen los vehículos de ser incluidos en la muestra. Esta probabilidad se obtuvo preguntando a los automovilistas la frecuencia con la que acuden a cargar combustible en una determinada unidad de tiempo. Para verificar este dato proporcionado por los automovilistas se puede hacer un seguimiento vehicular.

Cada vehículo está multiplicado por el inverso de dicha frecuencia de carga para poder obtener las características reales de la población. Así se está dando un menor peso a vehículos que acuden frecuentemente a la gasolinera para compensar con los que cargan combustible menos seguido y tienen menor posibilidad de ser incluidos en la muestra.

1.2.2. PROPIEDADES DEL ESTIMADOR

El estimador mencionado (ii) tiene la propiedad de que converge en probabilidad al parámetro poblacional $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$. Es decir, es asintóticamente insesgado.⁸

Demostración: El estimador propuesto

$$\bar{y} = \frac{\sum_{i=1}^n \left(y_i * \frac{1}{P_i} \right)}{\sum_{i=1}^n \frac{1}{P_i}} \quad (ii)$$

es un cociente de variables aleatorias. Por simplicidad de notación sea $Z = \sum_{i=1}^n y_i/P_i$ y sea $X = \sum_{i=1}^n 1/P_i$.

Una vez acordada esta notación, se va encontrar la esperanza de dicho cociente de variables aleatorias Z/X . Esta se puede obtener desarrollando el cociente en serie de Taylor alrededor de los valores esperados de Z y X .⁹

La serie de Taylor para una función $f(x,y)$ alrededor de $x=a$ $y=b$ se puede expresar como:¹⁰

8. Ibid, pag. 52 - 54.

9. MOOD, Introduction to the Theory of Statistics, pag. 181.

10. Ibid, pag. 533 - 534.

$$f(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) + \frac{1}{2!} [f_{xx}(a,b)(x-a)^2 + 2f_{xy}(a,b)(x-a)(y-b) + f_{yy}(a,b)(y-b)^2] + c$$

Donde:

$c =$ términos de orden superior a 2

$$f_x(a,b) = \frac{\partial f}{\partial x} \Big|_{x=a, y=b}$$

$$f_{xy}(a,b) = \frac{\partial^2 f}{\partial x \partial y} \Big|_{x=a, y=b}$$

y de manera similar para las demás.

Tenemos que:

$$f(Z,X) = \frac{Z}{X}$$

$$f(E[Z], E[X]) = \frac{E[Z]}{E[X]}$$

$$f_x(E[Z], E[X]) = \frac{-E[Z]}{E^2[X]}$$

$$f_z(E[Z], E[X]) = \frac{E[X]}{E^2[X]}$$

$$f_{xx}(E[Z], E[X]) = \frac{2E[Z]}{E^3[X]}$$

$$f_{zz}(E[Z], E[X]) = 0$$

$$f_{zx}(E[Z], E[X]) = \frac{-1}{E^2[X]}$$

Entonces se puede expresar la esperanza de Z/X como:

$$\begin{aligned}
 E \left[\frac{Z}{X} \right] &= E \left[\frac{E[Z]}{E[X]} + \frac{E[X](Z-E[Z]) - E[Z](X-E[X])}{E^2[X]} + \right. \\
 &\quad \left. \frac{1}{2} \left(- \frac{2(Z-E[Z])(X-E[X])}{E^2[X]} + \frac{2E[Z](X-E[X])^2}{E^3[X]} \right) \right] + c \\
 &= E \left[\frac{E[Z]}{E[X]} + \frac{E[X]Z - E[Z]X}{E^2[X]} - \frac{(Z-E[Z])(X-E[X])}{E^2[X]}, \right. \\
 &\quad \left. \frac{E[Z](X-E[X])^2}{E^3[X]} \right] + c \\
 &= \frac{E[Z]}{E[X]} - \frac{\text{Cov}[X,Z]}{E^2[X]} + \frac{E[Z] \text{Var}[X]}{E^3[X]} + c \quad (iii)
 \end{aligned}$$

Donde c es un término de orden superior a 2.

La ecuación (iii) representa la esperanza del estimador de razón definido por la ecuación (ii), en todas las muestras posibles de tamaño n .

Para poder analizar el comportamiento del estimador (ii) en todas las muestras posibles, se extendieron sus sumatorias a toda la población.

Con esta finalidad se introdujo una variable auxiliar λ_i que fuera el número de veces que la i -ésima unidad se presentara en una muestra de tamaño n , con $i = 1, 2, \dots, N$.¹¹ Esta variable tiene la característica de que $\sum_{i=1}^N \lambda_i = n$.

Entonces, si se extrae una muestra de tamaño n del conjunto de vehiculos que se encuentran en circulación en una ciudad, dado que dicho conjunto constituye una población finita de tamaño N , la variable λ_i puede ser multiplicada por cada término del numerador y del denominador del estimador dado por (ii). Lográndose así extender las sumatorias del estimador a toda la población.

Por otra parte, considerando que el muestreo es con reemplazo y que P_i es la probabilidad de observar la i -ésima unidad de la población se tiene que la distribución conjunta de $\lambda_1, \lambda_2, \dots, \lambda_N$ es una distribución binomial cuya función de densidad es:¹²

$$f(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{n!}{1! 2! \dots N!} \prod_{i=1}^N P_i^{\lambda_i}$$

Además en una distribución binomial tenemos que:¹³

-
11. RAJ, Teoría del Muestreo, pag. 136.
 12. MOOD, Op.Cit., pag. 137.
 13. Ibid, pag. 89, 197.

$$E[\lambda_i] = nP_i$$

$$\text{Var}[\lambda_i] = nP_i(1-P_i)$$

$$\text{Cov}[\lambda_i, \lambda_j] = -nP_iP_j$$

Si se extienden las sumatorias a toda la población se tiene que

$$Z = \sum \frac{Y_i}{P_i} \lambda_i \quad \text{y que} \quad X = \sum \frac{1}{P_i} \lambda_i$$

Entonces:

$$E[Z] = E\left[\sum \frac{Y_i}{P_i} \lambda_i\right]$$

$$= \sum \frac{Y_i}{P_i} E[\lambda_i]$$

$$= n \sum Y_i P_i$$

$$= nN\bar{Y}$$

(iv)

$$E[X] = E\left[\sum \frac{1}{P_i} \lambda_i\right]$$

$$= \sum \frac{1}{P_i} E[\lambda_i]$$

$$= nN$$

(v)

$$\begin{aligned}
\text{Var}[Z] &= \text{Var}\left[\sum \frac{Y_i}{P_i} \lambda_i\right] \\
&= \sum \frac{Y_i^2}{P_i^2} \text{Var}[\lambda_i] + \sum_{i \neq j} \frac{Y_i Y_j}{P_i P_j} \text{Cov}[\lambda_i \lambda_j] \\
&= \sum \frac{n(1-P_i) Y_i^2}{P_i} - \sum_{i \neq j} n Y_i Y_j \\
&= \sum \frac{n(1-P_i) Y_i^2}{P_i} - n(\sum Y_i)^2 + n \sum Y_i^2 \\
&= n \left[\sum \frac{Y_i^2}{P_i} - (\sum Y_i)^2 \right] \quad \text{(vi)}
\end{aligned}$$

$$\begin{aligned}
\text{Var}[X] &= \text{Var}\left[\sum \frac{1}{P_i} \lambda_i\right] \\
&= \sum \frac{1}{P_i^2} \text{Var}[\lambda_i] + \sum_{i \neq j} \frac{\text{Cov}[\lambda_i \lambda_j]}{P_i P_j} \\
&= \sum \frac{n(1-P_i)}{P_i} - \sum_{i \neq j} n \\
&= \sum \frac{n(1-P_i)}{P_i} - n(\sum 1)^2 + n \sum 1^2 \\
&= n \left[\sum \frac{1}{P_i} - N^2 \right] \quad \text{(vii)}
\end{aligned}$$

Como último paso en la demostración de que el estimador propuesto por la ecuación (ii) es asintóticamente insesgado se sustituyeron los valores de $E[Z]$, $E[X]$, $\text{Var}[Z]$ y $\text{Var}[X]$, dados por las ecuaciones (iv), (v), (vi) y (vii) en la ecuación (iii).

$$E \left[\frac{Z}{X} \right] = \frac{nN\bar{Y}}{nN} - \frac{\text{Cov}[X, Z]}{n^2 N^2} + \frac{nN\bar{Y} n \left[\frac{\sum_{i=1}^N (1/P_i) - N^2}{n^3 N^3} \right]}{n^3 N^3} + c$$

$$= \bar{Y} \left[1 + \frac{1}{nN^2} \left(\text{Cov}[X, Z]/\bar{Y}n + \frac{\sum_{i=1}^N (1/P_i) - N^2}{N^2} \right) + c \right]$$

Para la covarianza entre Z y X no se usó su expresión matemática porque ésta resulta un poco compleja, se utilizó la siguiente desigualdad para estudiar el estimador:

$$- \sqrt{\text{Var}[Z] \text{Var}[X]} \leq \text{Cov}[X, Z] \leq \sqrt{\text{Var}[Z] \text{Var}[X]} \quad (\text{viii})$$

Entonces el estimador queda expresado por la siguiente desigualdad

$$\bar{Y} \left(1 + \frac{1}{nN^2} [B_1 - B_2] + c \right) \leq E \left[\frac{Z}{X} \right] \leq \bar{Y} \left(1 + \frac{1}{nN^2} [B_1 + B_2] + c \right) \quad (\text{ix})$$

Donde:

$$B_1 = \frac{N}{\sum_{i=1}^N (1/P_i) - N^2}$$

$$B_2 = \frac{1}{\bar{Y}} \left\{ \left[\frac{N}{\sum_{i=1}^N (Y_i^2/P_i)} - \left(\frac{N}{\sum_{i=1}^N Y_i} \right)^2 \right] \left[\frac{N}{\sum_{i=1}^N (1/P_i) - N^2} \right] \right\}^{1/2}$$

En la desigualdad (ix) se expresa que el valor esperado del estimador en todas las muestras posibles converge a la media poblacional \bar{y} a través de dos factores que tienden a cero conforme se incrementa el tamaño de la muestra.

Es decir, el estimador propuesto

$$\bar{y} = \frac{\sum_{i=1}^n \left(Y_i * \frac{1}{P_i} \right)}{\sum_{i=1}^n \frac{1}{P_i}}$$

es asintóticamente insesgado.

Una vez hecho lo anterior se procedió a calcular la varianza del estimador. Expresándola en términos de la esperanza se tiene la siguiente ecuación

$$\begin{aligned} \text{VAR} \left[\frac{Z}{X} \right] &= E \left[\left(\frac{Z}{X} - \frac{E[Z]}{E[X]} \right)^2 \right] \\ &= E \left[\left(\frac{E[Z]}{E[X]} \left(\frac{ZE[X]}{XE[Z]} - 1 \right) \right)^2 \right] \\ &= \frac{E^2[Z]}{E^2[X]} E \left[\left(\frac{ZE[X] - XE[Z]}{XE[Z]} \right)^2 \right] \quad (x) \end{aligned}$$

El término $\left(\frac{ZE[X] - XE[Z]}{XE[Z]} \right)^2$ se va a desarrollar en serie de

Taylor, alrededor de los valores esperados de X y de Z, reteniendo sólo los términos de primer orden.

Tenemos que

$$f(Z, X) = \left(\frac{ZE[X] - XE[Z]}{XE[Z]} \right)^2$$

$$f(E[Z], E[X]) = 0$$

$$f_x(E[Z], E[X]) = \frac{-1}{E[X]}$$

$$f_z(E[Z], E[X]) = \frac{1}{E[Z]}$$

Sustituyendo en (x) se tiene que:

$$\begin{aligned} \text{Var} \left[\frac{Z}{X} \right] &= \frac{E^2[Z]}{E^2[X]} E \left[\left(\frac{-(X-E[X])}{E[X]} + \frac{Z-E[Z]}{E[Z]} \right)^2 \right] + c \\ &= \frac{E^2[Z]}{E^2[X]} E \left[\frac{(X-E[X])^2}{E^2[X]} + \frac{(Z-E[Z])^2}{E^2[Z]} - \right. \\ &\quad \left. \frac{2(X-E[X])(Z-E[Z])}{E[X]E[Z]} \right] + c \\ &= \frac{E^2[Z]}{E^2[X]} \left[\frac{\text{Var}[X]}{E^2[X]} + \frac{\text{Var}[Z]}{E^2[Z]} - \frac{2\text{Cov}[X, Z]}{E[X]E[Z]} \right] + c \quad (\text{xi}) \end{aligned}$$

La ecuación (xi) representa la varianza del estimador de razón definido por la ecuación (ii), en todas las muestras posibles de tamaño n.

Para poder analizar el comportamiento de la varianza del estimador (ii) en todas las muestras posibles, se extendieron nuevamete sus sumatorias a toda la población.

Esto se hizo recurriendo en primer a lugar la desigualdad (viii) para la Cov{X,Z}, que nos lleva a una nueva expresión para la ecuación (xi). En dicha expresión se sustituyeron los valores de E{Z}, E{X}, Var{Z} y Var{X}, dados por las ecuaciones (iv), (v), (vi) y (vii).

Despejando la Cov{X,Z} en la ecuación (xi) se tiene:

$$\text{Cov}\{X,Z\} = \left[\frac{E^2\{Z\}\text{Var}\{X\}}{E^4\{X\}} + \frac{\text{Var}\{Z\}}{E^2\{X\}} - \text{Var}\left[\frac{Z}{X}\right] \right] \frac{E^3\{X\}}{2E\{Z\}} + c$$

Utilizando la desigualdad (viii)

$$-\sqrt{V\{Z\}V\{X\}} \leq \left[\frac{E^2\{Z\}\text{Var}\{X\}}{E^4\{X\}} + \frac{\text{Var}\{Z\}}{E^2\{X\}} - \text{Var}\left[\frac{Z}{X}\right] \right] \frac{E^3\{X\}}{2E\{Z\}} + c \leq \sqrt{V\{Z\}V\{X\}}$$

De ahí que:

$$\begin{aligned}
 \text{Var} \left[\frac{Z}{X} \right] &\leq \left[\frac{E^2[Z] \text{Var}[X]}{E^4[X]} + \frac{\text{Var}[Z]}{E^2[X]} + \frac{2E[Z]}{E^3[X]} \sqrt{V[Z]V[X]} \right] \\
 &\leq \frac{E^2[Z]}{E[X]} \left[\frac{\text{Var}[X]}{E^2[X]} + \frac{\text{Var}[Z]}{E^2[Z]} + \frac{2 \sqrt{V[Z]V[X]}}{E[X]E[Z]} \right] \\
 &\leq \frac{E^2[Z]}{E^2[X]} \left[\frac{\sqrt{\text{Var}[X]}}{E[X]} + \frac{\sqrt{\text{Var}[Z]}}{E[Z]} \right]^2 \\
 &\leq \bar{Y}^2 \left[\frac{(n[\sum(1/P_i) - N^2])^{1/2}}{nN} + \frac{(n[\sum(Y_i^2/P_i) - (\sum Y_i)^2])^{1/2}}{nN\bar{Y}} \right]^2 \\
 &\leq \frac{\bar{Y}^2}{n} \left[\left[\sum(1/N^2 P_i) - 1 \right]^{1/2} + \left[\sum(Y_i^2/N^2 \bar{Y}^2 P_i) - 1 \right]^{1/2} \right]^2 \quad \text{(xii)}
 \end{aligned}$$

La ecuación (xii) muestra que si n tiende a infinito la varianza del estimador tiende a cero. Quedando así demostradas dos propiedades de suma importancia del estimador.

CAPITULO 2

DISEÑO CONCEPTUAL Y OPERATIVO DE LA ENCUESTA

Después de haber precisado la manera de llevar a cabo tanto el proceso de selección como el de estimación, pasamos a otra fase que es el diseño de la encuesta propiamente.

En esta parte del trabajo se presentarán los aspectos concretos de la encuesta que se llevó a cabo como lo son: la población que se seleccionó para ser encuestada, el tamaño de la muestra, el cuestionario aplicado, etc.

Estos aspectos son sumamente importantes, ya que gran parte del éxito de la encuesta dependerá de ellos.

2.1. POBLACION

Considerando que el objetivo del presente estudio es obtener información acerca del uso del automóvil particular en México, podemos decir que nuestra población objetivo son todos los automóviles de uso particular que circulan en el país.

Sin embargo, esta población se verá restringida por diversas razones, obteniéndose así una población bajo muestreo distinta a la población considerada inicialmente.

En primer lugar, para efecto de la selección de las ciudades en las que se levantaría la encuesta, se consideraron sólo poblaciones grandes en las que se pudiera recopilar un volumen considerable de información. Esto se hizo para optimizar los recursos disponibles, ya que el costo de encuestar en una población pequeña es muy alto en relación a la cantidad de información que se puede obtener.

Con la ayuda de datos de la Secretaría de Programación y Presupuesto se eligieron las 80 ciudades más importantes del país. En ellas se concentra el 50% de la población total, se localiza el 80% de la producción industrial, circula el 88% de los vehículos, se cuenta con el 89% de los teléfonos instalados y se llevaron a cabo el 84% de las ventas de automóviles y camiones nuevos. Estos datos son del año 1984.

Esta es la primera restricción a la población objetivo, ya que no quedan incluidas en la población bajo muestreo las ciudades que no se encuentren en la lista de dichas 80 ciudades.

Las ciudades seleccionadas se agruparon de acuerdo a su número de habitantes y su tasa de motorización. Esta tasa es el cociente del número de habitantes entre el número de automóviles registrados.

El número de automóviles registrados se obtuvo de datos de la Dirección General de Estadística de la Secretaría de Programación

y Presupuesto, el último dato disponible era de 1982. Por lo tanto se tomó una proyección de la población al mismo año. Esto se hizo utilizando datos de los censos de población de 1970 y 1980, considerando la tasa media anual de incremento en dos años y ajustando contra las proyecciones por estado del Consejo Nacional de Población.

Los grupos quedaron como se menciona a continuación:

Grupo 1 Distrito Federal

Grupo 2 Poblaciones con más de 600,000 habitantes.

Grupo 3A Poblaciones con más de 300,000 habitantes y una alta tasa de motorización.

Grupo 3B Poblaciones con más de 300,000 habitantes y una baja tasa de motorización.

Grupo 4A Poblaciones con más de 50,000 habitantes y una alta tasa de motorización.

Grupo 4B Poblaciones con más de 50,000 habitantes y una baja tasa de motorización.

Se hizo esta clasificación para poder captar y analizar posibles diferencias por un lado, entre ciudades muy pobladas y ciudades menos pobladas; y por otro, entre ciudades más motorizadas y ciudades menos motorizadas. La clasificación de cada una de las ciudades dentro de los grupos mencionados se puede consultar en la tabla 1 del anexo.

Con base en los datos mencionados se calculó una tasa media de motorización de .125, lo que corresponde a un automóvil por cada 8 personas. Las poblaciones con una tasa menor a ésta quedaron clasificadas como ciudades con una baja tasa de motorización y las que tuvieron una tasa mayor como ciudades con una alta tasa de motorización.

Esta tasa se ve fuertemente influenciada por los datos del Distrito Federal, ya que ahí se encuentran registrados cerca del 40% de todos los automóviles, debido a esto también podemos decir que nuestras ciudades están clasificadas en ciudades con una tasa de motorización mayor que la del Distrito Federal y ciudades con una tasa menor que la del Distrito Federal.

Por razones de costo desde el inicio del estudio se planteó que las encuestas se levantarían en diez ciudades en el interior de la República y en la Ciudad de México. Dichas ciudades se seleccionaron de la lista mencionada en el anexo.

El hecho de incluir necesariamente la Ciudad de México se debe a que en ella se encuentra concentrado aproximadamente el 21% de la población, el 24% de la población económicamente activa y el 30% de toda la actividad industrial y comercial.¹ Por esta razón no puede quedar fuera de ningún estudio a nivel nacional.

Una vez clasificadas las ciudades se tomaron en cuenta varios criterios para hacer la selección, entre los que podemos citar los siguientes:

- a) Considerar ciudades de cada uno de los grupos.
- b) Tener cobertura geográfica.
- c) Incluir ciudades con distintas actividades económicas.
- d) Tomar en cuenta que en las ciudades seleccionadas se deben levantar además otras encuestas para recopilar información acerca del transporte de carga y el transporte público.
- e) Aprovechar las ciudades en las que se tuvieran contactos que pudieran ayudar en la organización del trabajo de campo y en la recopilación de información.

1. Datos proporcionados por la Secretaria de Programación y Presupuesto.

Los últimos dos criterios de selección aplicados responden a procurar la mejor asignación de los recursos disponibles.

Puesto que se tuvieron que contemplar todos estos factores no se hizo una selección aleatoria. Se organizó una reunión a la que asistieron todos los interesados en el proyecto y se decidió en que ciudades se levantaría la encuesta.

La selección de las ciudades por grupo fue la siguiente:

Grupo 1 Distrito Federal

Grupo 2 Guadalajara
 León

Grupo 3A Ciudad Juárez

Grupo 3B Veracruz
 Saltillo
 Hermosillo

Grupo 4A Cuernavaca
 Tuxtla Gutiérrez

Grupo 4B Mazatlán
 Campeche

Debido a que la selección se hizo de esta manera, no conocemos la probabilidad que tuvieron las ciudades de pertenecer a la muestra. Entonces, a nivel nacional no vamos a poder obtener un estimador estadístico, sino que nos vamos a tener que limitar al uso de un indicador que nos permita tener una idea de la situación global, aunque desconozcamos su precisión.

Considerando lo anterior, podemos decir que nuestra población bajo muestreo se verá reducida a los automóviles que circulan en las ciudades mencionadas. Desde el punto de vista estadístico vamos a considerar cada ciudad como una población independiente de la cual se va a tomar una muestra.

2.2. MARCO MUESTRAL

Debe haber un marco, ya sea una lista, un mapa u otra descripción del material, a partir del cual pueda hacerse la selección de la muestra.²

Se descartó la idea de hacer la selección de la muestra basada en una lista de automóviles debido a las siguientes razones:

2. RAJ, La Estructura de las Encuestas por Muestreo, pag. 174.

- Es sumamente difícil su obtención, ya que nunca se encuentran al día los datos y es prácticamente imposible hacer una actualización.

- Sólo se puede obtener una lista de los automóviles registrados en la ciudad y no de los que circulan en ella. Hay automóviles que aunque estén registrados no se encuentran en circulación y otros que circulan en la ciudad pero están registrados en otro lugar.

- Para identificar cada automóvil se tendrían que hacer visitas domiciliarias en zonas muy dispersas de la ciudad, ya que al escoger los automóviles de una lista, éstos pueden estar ubicados en cualquier punto de la ciudad. Además se debe considerar que un número considerable de vehículos ya cambiaron de propietario o los propietarios cambiaron de domicilio y este no coincide con el registrado. Todo esto es demasiado costoso y no permite obtener un gran volumen de información.

Se consideró entonces el muestreo de área, en el cual no se tiene la necesidad de una lista de todas las unidades, sino que éstas sólo pueden identificarse por su posición geográfica.³ Es decir,

3. Ibid, pag. 179.

se escogen ciertas zonas y se levanta la encuesta a cada uno de los automovilistas de dichas zonas.

Se planteó la posibilidad de dividir cada ciudad en zonas y hacer una encuesta domiciliaria. Sin embargo, aún este tipo de encuesta domiciliaria es muy costoso. Entonces se planteó la posibilidad de levantar la encuesta en estaciones de servicio por presentarse ahí las siguientes ventajas:

- Captar en un mismo lugar una gran cantidad de información a un costo relativamente bajo (comparado por ejemplo con la encuesta domiciliaria).
- Aprovechar el tiempo ocioso que pasa el automovilista en la gasolinera sin interrumpirlo en sus actividades.
- Aplicar la encuesta sólo a personas que conducen un automóvil y en consecuencia pueden indicar sus hábitos de uso.

Para levantar encuestas en estaciones de servicio en primer lugar se tuvo que hacer la selección de dichas estaciones. Para esto se constituyó un marco muestral que incluiría todas las gasolineras de las ciudades previamente seleccionadas.

El marco muestral quedó constituido por todas las estaciones de

servicio que se encuentran en las ciudades seleccionadas. Estas estaciones son abastecidas por las siguiente agencias de PEMEX:

0605	Campeche
0608	Ciudades Juárez
0616	Cuernavaca
0620	Guadalajara
0625	Hermosillo
0630	León
0635	Mazatlán
0653	Saltillo
0661	Tuxtla Gutiérrez
0664	Veracruz
0682	Azcapotzalco
0683	Satélite Oriente
0684	Satélite Sur

Las gasolineras son las unidades primarias de muestreo, y sobre ellas se diseñó el marco muestral. Las unidades de muestreo de la segunda etapa son los automóviles, pero sobre ellas no se puede elaborar un marco muestral porque no se sabe con anticipación que vehículos acudirán a cargar gasolina a las estaciones de servicio en cuestión el día de la encuesta.

No se va a hacer una selección previa de los automóviles, sino que se escuestará a aquéllos que pasen a cargar gasolina cuando se

esté llevando a cabo el levantamiento. Durante el levantamiento de la encuesta se procurará encuestar a todos los automovilistas, es decir, se hará un censo de las unidades de la segunda etapa.

2.3. TAMAÑO DE MUESTRA

Al planear una encuesta nos encontramos con que uno de los problemas fundamentales es determinar el tamaño de la muestra. Esto es sumamente importante, ya que gran parte de su éxito depende de esta decisión.⁴

En efecto, si tomáramos una muestra demasiado pequeña, no obtendríamos resultados suficientemente precisos; y si por el contrario, tomáramos una muestra demasiado grande, esto podría implicar un desperdicio de recursos.⁵

En nuestro caso, determinar el tamaño de muestra presenta una dificultad particular, ya que se trata de un proceso bietápico. Dicho proceso contempla una primera etapa relativa a la selección de las estaciones de servicio en cada ciudad, y una segunda correspondiente a la selección de los automóviles en las estaciones de servicio.

4. COCHRAN, Técnicas de Muestreo, pag. 103.

5. Idem

2.3.1. MARCO TEORICO

Antes de tratar de resolver el problema sobre el tamaño adecuado de la muestra, nos debemos hacer el siguiente cuestionamiento: ¿ Pueden hacerse inferencias confiables acerca de un parámetro desconocido de la población utilizando solamente una muestra aleatoria de tamaño n ? La respuesta es afirmativa, y se puede demostrar probando la Ley Débil de los Grandes Números.⁶

Esta ley afirma lo siguiente: Se puede determinar una n tal que, si se toma una muestra aleatoria de tamaño n o mayor, de una población de varianza finita y media \bar{x} , la probabilidad de que la media muestral \bar{x}_n difiera de la media poblacional \bar{x} en menos de una cantidad determinada ϵ , arbitrariamente pequeña, llega a hacerse tan próxima a uno como se desee.⁷

Designemos por $f(x)$ una función de densidad con media \bar{x} y varianza finita σ^2 y sea \bar{x}_n la media de una muestra aleatoria de $f(x)$ de tamaño n . Sean ϵ y δ dos números pequeños dados tales que $\epsilon > 0$ y $0 < \delta < 1$. Entonces si n es cualquier entero mayor que $\sigma^2/\epsilon^2\delta$:⁸

6. MOOD, Introducción a la Teoría de la Estadística, pag. 169.

7. Idem

8. Ibid, pag. 170.

$$\Pr [| \bar{x}_n - \bar{X} | < \epsilon] \geq 1 - \delta$$

Para demostrar esta ley nos vamos a basar en la Desigualdad de Tchebycheff, que se expone a continuación:⁹

Sea $f(x)$ una función de densidad con media \bar{X} y varianza finita S^2 , α un número positivo cualquiera y \bar{x}_n una variable aleatoria con una desviación estándar $\sigma(\bar{x}_n)$ que representa la media de una muestra al azar de $f(x)$ de tamaño n . Entonces:

$$\Pr [| \bar{x}_n - \bar{X} | \geq \alpha \sigma(\bar{x}_n)] \leq \frac{1}{\alpha^2}$$

Demostración: Tenemos que \bar{x}_n es una variable aleatoria y \bar{X} es una constante, por lo tanto

$$V(\bar{x}_n) = \sum_i (\bar{x}_i - \bar{X})^2 p_i$$

$$V(\bar{x}_n) = \sum_j (\bar{x}_j - \bar{X})^2 p_j + \sum_k (\bar{x}_k - \bar{X})^2 p_k$$

Donde la primera suma se refiere a todas las j , para las cuales $|\bar{x}_j - \bar{X}| < \lambda$, y la segunda suma incluye todas las k para las cuales $|\bar{x}_k - \bar{X}| \geq \lambda$. Siendo λ una constante positiva cualquiera.

9. RAJ, Teoría del Muestreo, pag. 28.

Entonces:

$$V(\bar{x}_n) \geq \sum_k (\bar{x}_k - \bar{X})^2 p_k \geq \lambda^2 \sum_k p_k$$

$$V(\bar{x}_n) \geq \lambda^2 \sum_k p_k = \lambda^2 \Pr [|\bar{x}_n - \bar{X}| \geq \lambda]$$

$$\frac{V(\bar{x}_n)}{\lambda} \geq \Pr [|\bar{x}_n - \bar{X}| \geq \lambda]$$

Si $\lambda = \alpha \sigma(\bar{x}_n)$

$$\Pr [|\bar{x}_n - \bar{X}| \geq \alpha \sigma(\bar{x}_n)] \leq \frac{1}{\alpha^2} \quad (i)$$

Quedando así demostrada la Desigualdad de Tchebycheff. Si \bar{x}_n es una variable de varianza finita la desigualdad anterior también se puede expresar como sigue:¹⁰

$$\Pr [|\bar{x}_n - \bar{X}| < \alpha \sigma(\bar{x}_n)] \geq 1 - \frac{1}{\alpha^2} \quad (ii)$$

Además tenemos que si \bar{x}_n es la media de una muestra aleatoria de tamaño n de una función de densidad $f(x)$ de varianza finita σ^2 , la varianza de \bar{x}_n , que representamos por $\sigma^2(\bar{x}_n)$, es σ^2/n . Entonces podemos representar la desigualdad (ii) como:

10. MOOD, Introduction to the Theory of Statistics, pag. 71.

$$\text{Pr} \left[\left| \bar{x}_n - \bar{x} \right| < \alpha \frac{\sigma}{\sqrt{n}} \right] \geq 1 - \frac{1}{\alpha^2} \quad (\text{iii})$$

Una vez demostrada la Desigualdad de Tchebycheff, podemos utilizarla para demostrar la Ley Débil de los Grandes Números¹¹

Sea $\alpha = 1/\sqrt{\delta}$, sustituyendo en (iii) tenemos,

$$\text{Pr} \left[\left| \bar{x}_n - \bar{x} \right| < \frac{\sigma}{\sqrt{\delta} \sqrt{n}} \right] \geq 1 - \delta$$

Vamos a sustituir una c en la desigualdad, tal que c sea mayor que $\sigma/\sqrt{\delta} \sqrt{n}$ esto no altera la desigualdad. Además con esto tenemos una n tal que $n > \sigma^2/c^2\delta$. Finalmente obtenemos

$$\text{Pr} \left[\left| \bar{x}_n - \bar{x} \right| < c \right] \geq 1 - \delta$$

Quedando así demostrada la Ley Débil de los Grandes Números, con lo cual podemos afirmar que se pueden hacer inferencias confiables acerca de un parámetro desconocido de la población utilizando sólo una muestra aleatoria de tamaño n . Cabe señalar que esta ley no involucra una distribución específica para la variable.

11. MOOD, Introducción a la Teoría de la Estadística, pag. 171.

Una vez comprobado lo anterior, la pregunta que surge inmediatamente después es cuál debe ser el tamaño n de dicha muestra aleatoria que nos va a llevar a hacer inferencias confiables acerca de la población.

Para determinar esto se consideró en primer lugar la precisión deseada. La cual se define en términos de un error e y de la confianza asociada a dicho error.

El error e es el máximo error involucrado en el muestreo que se está dispuesto a aceptar en la encuesta, y está dado en términos de un determinado porcentaje del parámetro poblacional que se va a analizar en la encuesta. Es decir $e = (a\%)\bar{x}_n$. La confianza asociada al error es la probabilidad de no cometer un error mayor al máximo error e que se está dispuesto a aceptar en la encuesta.¹²

Por ser la precisión el factor más importante para determinar el tamaño de muestra, vamos a encontrar una expresión que nos relacione estos dos elementos.¹³

12. ABAD, Introducción al Muestreo, pag. 61-63.

13. COCHRAN, Op.Cit., pag. 109.

Entonces, si se conviene en un margen de error menor a e en la estimación de alguna característica de la población, por ejemplo la media, y estamos dispuestos a correr un pequeño riesgo r de que el error real supere a e , se quiere que:

$$\Pr [| \bar{x}_n - \bar{X} | < e] \geq 1 - r$$

Por la Ley Débil de los Grandes Números sabemos que existe una n que cumple con lo anterior.

Para determinar el tamaño de dicha n nos vamos a auxiliar del Teorema del Límite Central, el cual nos dice que si \bar{x}_n es la media de una muestra aleatoria de una población de varianza finita σ^2 y media \bar{X} , a medida que n aumenta la distribución de \bar{x}_n tiende a una distribución normal con varianza σ^2/n y media \bar{X} . Este teorema se cumple para cualquier distribución que tenga la población de la que provenga la muestra aleatoria.¹⁴

Si \bar{x}_n se distribuye normalmente existe una probabilidad $c = 1 - r$ de que \bar{x}_n se encuentre en el intervalo

$$\left(\bar{X} - t \frac{\sigma}{\sqrt{n}} , \bar{X} + t \frac{\sigma}{\sqrt{n}} \right)$$

14. SNEDECOR, Métodos Estadísticos, pag. 75 - 76.

Donde t es el valor de la abscisa en la distribución normal que nos deje en la parte central de la curva un área igual a la confianza requerida c .¹⁵

Es decir, se quiere que

$$\text{Pr} \left[\left| \bar{x}_n - \bar{X} \right| < t \frac{\sigma}{\sqrt{n}} \right] \geq 1 - r$$

Si se quiere que el máximo error del estimador \bar{x}_n de \bar{X} sea $e(\bar{x}_n)$ se debe verificar que

$$e(\bar{x}_n) \geq t \frac{\sigma}{\sqrt{n}}$$

Despejando n de esta desigualdad se puede obtener una expresión para encontrar el tamaño mínimo de una muestra aleatoria en términos de la precisión deseada.

$$n \geq \frac{t^2 \sigma^2}{e^2(\bar{x}_n)}$$

Teniendo esta expresión para obtener el tamaño de la muestra en términos de la precisión se debe determinar para qué variable o variables se va a fijar dicha precisión. Esto es importante porque en la encuesta hay varios parámetros que se quieren

15. ABAD, Op.Cit., pag. 63.

estimar, y se tiene que fijar un mismo tamaño de muestra para toda la encuesta.

Para hacer esto se puede tomar una única variable, la que se considere más importante, para que gobierne el tamaño de muestra y las demás variables se introducen sin importar la precisión con que sean estimadas.¹⁶

Otra forma para determinar un tamaño adecuado de muestra es tomando varias variables que se consideren importantes, calcular el tamaño de muestra para cada una de ellas y elegir el mayor. Así una de las variables tendrá la precisión requerida y las demás una precisión mayor.¹⁷

En este caso se consideró el segundo criterio, es decir, se eligieron las variables que se consideraron más importantes para determinar el tamaño de muestra.

Otro factor que hay que tomar en cuenta es que se trata de un proceso bietápico. Esto significa que se tienen que determinar dos tamaños de muestra, uno para cada etapa.

16. Ibid, pag. 71.

17. Idem.

En cada ciudad se tiene una primera etapa que es la selección de las gasolineras, para esto tenemos que determinar un tamaño de muestra n para ver en cuántas se va a levantar la encuesta. La segunda etapa consiste en seleccionar los automóviles que se van a encuestar en cada estación, esto nos lleva a encontrar un tamaño de muestra m_i para saber el número de automóviles que se va a encuestar en cada estación i en que se levante la encuesta.

Finamente se analizaron los recursos disponibles, para ver si éstos eran consistentes con el tamaño de muestra propuesto. Para hacer esto se hizo una cuidadosa estimación del costo y el tiempo disponibles para la encuesta con las dimensiones sugeridas.¹⁸

2.3.2. DETERMINACION DEL TAMAÑO DE MUESTRA

Para llevar a cabo la primera etapa del proceso de selección se escogieron las gasolineras mediante el método de muestreo aleatorio simple sin reemplazo.

La varianza de la media \bar{x}_n de una muestra aleatoria simple sin reemplazo, esta dada por :¹⁹

18. COCHRAN, Op.Cit., pag. 105.

19. RAJ, Teoria del Muestreo, pag. 51.

$$V(\bar{x}_n) = \frac{S^2}{n} \left[1 - \frac{n}{N} \right]$$

Demostración:²⁰

$$V(\bar{x}_n) = E [\bar{x}_n - \bar{X}]^2$$

$$V(\bar{x}_n) = E \left[\frac{1}{n} \sum_{i=1}^n x_i - \bar{X} \right]^2$$

$$V(\bar{x}_n) = E \left[\frac{1}{n} \left(\sum_{i=1}^n x_i - n\bar{X} \right) \right]^2$$

$$V(\bar{x}_n) = \frac{1}{n^2} E \left[\sum_{i=1}^n (x_i - \bar{X}) \right]^2$$

$$V(\bar{x}_n) = \frac{1}{n^2} E \left[\sum_{i=1}^n (x_i - \bar{X})^2 + \sum_{i \neq j}^n (x_i - \bar{X})(x_j - \bar{X}) \right]$$

$$V(\bar{x}_n) = \frac{1}{n^2} E \sum_{i=1}^n (x_i - \bar{X})^2 + \frac{1}{n^2} E \sum_{i \neq j}^n (x_i - \bar{X})(x_j - \bar{X})$$

$$V(\bar{x}_n) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^N \frac{1}{N} (x_j - \bar{X})^2 + \frac{1}{n^2} E \sum_{i \neq j}^n (x_i - \bar{X})(x_j - \bar{X})$$

$$V(\bar{x}_n) = \frac{1}{n^2} n \sigma^2 + \frac{1}{n^2} \sum_{i \neq j}^n \frac{1}{N(N-1)} \sum_{k \neq i}^N (x_k - \bar{X})(x_i - \bar{X})$$

$$V(\bar{x}_n) = \frac{\sigma^2}{n} + \frac{1}{n^2} \frac{n(n-1)}{N(N-1)} \sum_{k \neq i}^N (x_k - \bar{X})(x_i - \bar{X})$$

20. Idem

$$V(\bar{x}_n) = \frac{\sigma^2}{n} + \frac{(n-1)}{nN(N-1)} \left[\left(\sum_{k=1}^N (x_k - \bar{X}) \right)^2 - \sum_{k=1}^N (x_k - \bar{X})^2 \right]$$

$$V(\bar{x}_n) = \frac{\sigma^2}{n} + \frac{(n-1)}{nN(N-1)} \left[-N\sigma^2 \right]$$

$$V(\bar{x}_n) = \frac{\sigma^2}{n} + \left[1 - \frac{n-1}{N-1} \right] \quad (\text{iv})$$

La varianza de los valores x_i en una población finita usualmente se define como:²¹

$$\sigma^2 = \sum_{i=1}^N \frac{1}{N} (x_i - \bar{X})^2$$

Sin embargo, para llegar a una expresión más sencilla, los resultados en muestreo generalmente se presentan en términos de una expresión ligeramente diferente, en donde se utiliza como divisor $(N - 1)$ en lugar de N . Entonces tenemos:²²

$$S^2 = \sum_{i=1}^N \frac{1}{N - 1} (x_i - \bar{X})^2$$

Si relacionamos σ^2 con S^2 tenemos que:

$$\sigma^2 = \frac{N - 1}{N} S^2$$

21. COCHRAN, Op.Cit., pag. 47.

22. Idem

Sustituyendo esta igualdad en (iv) la nueva expresión de la varianza es:

$$V(\bar{x}_n) = \frac{N-1}{N} \frac{S^2}{n} + \left[1 - \frac{n-1}{N-1} \right]$$

$$V(\bar{x}_n) = \frac{S^2}{n} + \left[1 - \frac{n}{N} \right]$$

En la ecuación anterior S^2 representa la varianza de las x_i en la población, es decir, la varianza poblacional, mientras que S^2/n es la varianza de la media muestral en poblaciones infinitas. Por tratarse en este caso de una población finita, esta varianza aparece reducida por el factor $(1-n/N)$. Este factor se conoce como factor de corrección para población finita, el efecto de ignorar la corrección es sobreestimar la varianza de \bar{x}_n .²³

Una vez encontrada una expresión para la varianza de la media \bar{x}_n se puede determinar el tamaño de la muestra en la primera etapa de muestreo en términos de la precisión deseada, recordando que el máximo error $e(\bar{x}_n)$ del estimador \bar{x}_n de \bar{X} debe ser tal que:

$$e(\bar{x}_n) \geq t \sqrt{V(\bar{x}_n)}$$

$$e(\bar{x}_n) \geq t \frac{S}{\sqrt{n}} \sqrt{1 - n/N}$$

23. ABAD, Op.Cit., pag. 48.

Despejando n en esta desigualdad obtenemos:

$$n \geq \frac{S^2}{\frac{e^2(\bar{x}_n)}{t^2} + \frac{S^2}{N}}$$

Donde:

n = Número de estaciones de servicio seleccionadas en la ciudad para el levantamiento de la encuesta

N = Número de estaciones de servicio registradas en la(s) agencia(s) que surte(n) a la ciudad

S^2 = Varianza de las x_i en la población

$e(\bar{x}_n)$ = Error del estimador \bar{x}_n

t = Valor del desvío normal correspondiente a la probabilidad de confianza deseada

Con esta expresión se puede determinar el número de estaciones de servicio a encuestar en cada ciudad.

En la encuesta se decidió trabajar con una confianza del 95%, lo que determina un valor de 1.96 para t. Otro dato conocido es el número de estaciones de servicio registradas en cada ciudad, sin

embargo, para determinar el valor de n , aun quedan dos variables que se desconocen, $e(\bar{x}_n)$ y S^2 .

Siendo S^2 una variable poblacional no se puede conocer su verdadero valor, teniéndose que hacer una estimación. Esta se hizo a partir de datos de la encuesta piloto. Se tomaron en cuenta los valores de la desviación estándar de las variables más importantes en la encuesta.

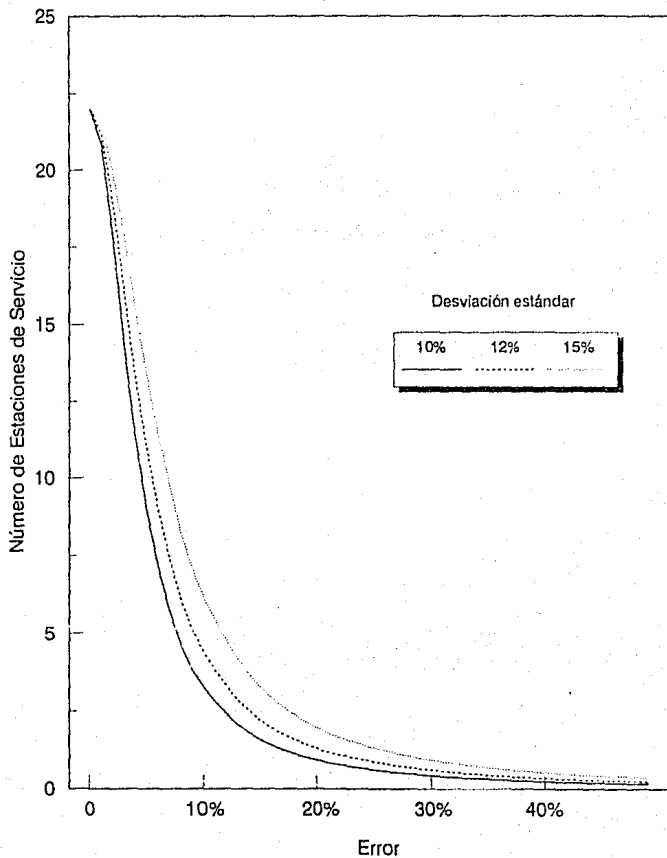
Las variables consideradas fueron: modelo del vehículo, cilindraje, frecuencia de carga de combustible, cantidad de litros cargados y kilometraje recorrido. Los valores obtenidos para la desviación estándar de dichas variables entre las gasolineras de la ciudad piloto oscilaron entre el 3% y el 10% de la respectiva media.

La encuesta piloto se llevó a cabo en 4 gasolineras de distintas características en el D.F.. Para calcular el tamaño de muestra se consideró el mayor valor obtenido, es decir, una desviación del 10% de la media.

Utilizando todos los datos mencionados se trazó una gráfica por ciudad para determinar el tamaño de muestra en cada caso. Un ejemplo de dichas gráficas se puede apreciar en la gráfica que se encuentra a continuación.

TAMAÑO DE MUESTRA

1a. ETAPA (HERMOSILLO)



En estas gráficas tenemos el número de estaciones de servicio a encuestar n en el eje de las ordenadas y el error del estimador $e(\bar{x}_n)$ en el eje de las abscisas, con lo que se puede apreciar como disminuye el error al aumentar el número de unidades encuestadas.

En las gráficas están trazadas tres curvas, una corresponde a una desviación estándar del 10% y las otras a desviaciones mayores. Esto se hizo para poder apreciar la magnitud del error en caso de ocurrir una desviación distinta a la supuesta.

El error del estimador $e(\bar{x}_n)$ no se fijó de antemano, debido a que por la forma de la curva de dichas funciones, hay casos en los que al aumentar el número de estaciones de servicio se gana mucho en precisión. Se consideró este hecho para aumentar en la medida de lo posible el número de unidades primarias en la muestra.

La única restricción que se puso al error es que no debía ser mayor a un 10%.

En base a las gráficas se determinó el siguiente número de unidades a encuestar por ciudad, siendo N el número de estaciones de servicio registradas en la(s) agencia(s) que surte(n) a la ciudad; y n el número de estaciones de servicio seleccionadas en cada ciudad para llevar a cabo el levantamiento de la encuesta:

	N	n
Campeche	5	3
Mazatlán	6	3
Tuxtla Gutiérrez	7	4
Cuernavaca	11	4
Hermosillo	22	7
Saltillo	13	5
Veracruz	11	4
Ciudad Juárez	70	11
León	13	5
Guadalajara	66	10
Distrito Federal	334	15

Una vez determinado el número de estaciones, éstas se eligieron por medio de un muestreo aleatorio simple.

El siguiente paso en cuanto a tamaño de muestra se refiere a determinar el número de coches a encuestar en cada gasolinera, para esto se analizó en primer lugar el proceso de selección de los automóviles.

Este consistió en colocar un encuestador en cada una de las bombas de gasolina de las estaciones de servicio seleccionadas para que encuestara a todos los automovilistas que pasaran a cargar gasolina a dicha bomba. Sólo se consideraron las bombas de gasolina Nova que estuvieran en funcionamiento. No se

consideraron las de gasolina Extra porque a nivel nacional su volumen de ventas es despreciable.

Debido a este proceso de selección se decidió no fijar con anterioridad el tamaño de muestra en esta segunda etapa, sino levantar el mayor número de encuestas posibles en cada gasolinera.

Para tener una estimación del número de encuestas que se levantarían en cada ciudad, se hizo un cálculo aproximado del número de vehículos que se podrían encuestar en cada gasolinera.

Este cálculo se hizo analizando el volumen de ventas de 1984 obtenido de las estadísticas de PEMEX de 60 gasolineras del país. Se utilizó esta información porque las ventas son directamente proporcionales al número de vehículos que acuden a cargar gasolina.

Las 60 gasolineras para hacer el análisis se seleccionaron aleatoriamente en las ciudades donde se levantaría la encuesta. De Campeche, Mazatlán, Tuxtla Gutierrez y Cuernavaca se seleccionaron 3 gasolineras; de Saltillo, Hermosillo, Veracruz y Cd. Juárez 4; de León y Guadalajara 8 y finalmente 16 de D.F.

Para determinar el posible número diario de automóviles que llegan a cargar gasolina en una estación de servicio a partir del volumen anual de ventas se utilizó la siguiente expresión:

$$M = \frac{\frac{I_a}{365} \times \%Ap}{\bar{I}}$$

Donde:

M = Número de automóviles diarios que cargan gasolina en una estación de servicio

I_a = Volumen anual de ventas de gasolina en una estación de servicio

%Ap = Porcentaje de litros vendidos a automóviles particulares

\bar{I} = Promedio de litros cargados por automóvil

El volumen anual de ventas se dividió entre 365 para tener una estimación del volumen diario. Cabe señalar, que no se tienen las mismas ventas toda la semana, sin embargo, la encuesta en todas las ciudades se levantó en martes, miércoles y jueves que son días típicos en los que se puede considerar que en un día se vende 1/365 de la venta anual.

Los datos del porcentaje de litros vendidos a automóviles y el promedio de litros cargados por automóvil se obtuvieron de la encuesta piloto y fueron 78% y 27 litros respectivamente.

Aunque teóricamente debían de ser encuestados todos los conductores de vehículos particulares que acudieran en un día a la gasolinera, no todos son encuestados debido a diversas razones como por ejemplo: personas que se niegan a contestar o personas que cargan en la noche cuando no se esta encuestando.

Considerando esto no se determinó el tamaño de muestra en base al posible número de vehículos que llegaran, sino en base al número de encuestas que se podrían levantar en una gasolinera determinada, utilizando la siguiente expresión:

$$A = M \times \%Ga \times \%Av \times \%Ev$$

Donde:

A = Posible número de automóviles a encuestar en una estación de servicio. (Cuestionarios válidos).

%Ga = Tiempo que se encuentran los encuestadores en la estación de servicio en proporción al tiempo que ésta permanece abierta

%Ae = Porcentaje de automóviles encuestados del total que acude a cargar gasolina.

%Ev = Porcentaje de cuestionarios válidos.

En esta expresión se consideró en primer lugar que la mayoría de las estaciones de servicio permanecen abiertas de las 6:00 a las 22:00 hrs. y que se contratarían dos turnos de 7 horas para levantar las encuestas, es decir, se cubrirían 14 de las 16 horas. Además se tomó en cuenta que hay gasolineras que abren las 24 horas. En base a lo anterior se calculó que se encuestaría el 75% del tiempo que permanecen abiertas las gasolineras.

Por experiencias anteriores y por la encuesta piloto se consideró que sólo el 80% de los conductores accedería a contestar el cuestionario.

Por último, en base también a experiencias anteriores se calculó que sólo 80% de los cuestionarios aplicados se pueden considerar como válidos.

Los resultados del posible número de cuestionarios válidos en una estación a partir del volumen de ventas para cada una de las 60 gasolineras se pueden apreciar en la tabla 2 del anexo al final de este trabajo. De estos resultados se puede concluir que en promedio se pueden levantar 300 encuestas válidas por gasolinera.

Para esta etapa se trazaron gráficas similares a las de la primera etapa para ver si el total de vehículos encuestados en una ciudad era suficiente para no tener un error global mayor del 10% en los resultados para las diferentes variables.

Para trazar dichas gráficas se utilizó inclusive la misma expresión para el tamaño de muestra que para las gráficas de la etapa anterior.

Esto no es muy riguroso porque en esta etapa el proceso de selección es distinto, se trata de un muestreo con probabilidades desiguales y con repetición. Sin embargo, como en las gráficas sólo se está analizando si para un tamaño de muestra dado el error del estimador no sobrepasa un cierto límite, éstas se consideraron válidas. Además no se debe olvidar que no se conoce la probabilidad de inclusión de las unidades con anterioridad.

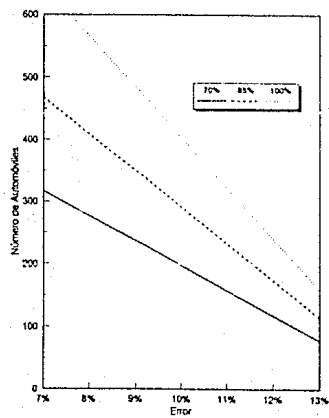
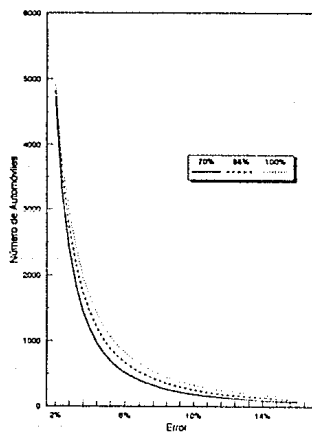
Al igual que para la etapa anterior se consideró una confianza del 95%, los valores de la desviación estándar dentro de la gasolinera de las distintas variables consideradas, obtenidos también de la encuesta piloto, oscilaron entre el 20% y el 70% de la respectiva media. Se consideró entonces el 70% de la media como desviación estándar para trazar las gráficas.

Como en las gráficas para determinar el tamaño de muestra en la primera etapa se trazaron curvas con una desviación mayor para poder ver la magnitud del error en el caso de tener una varianza mayor a la supuesta.

Un ejemplo de las gráficas con que se trabajó en esta segunda etapa se puede apreciar en la gráfica a continuación.

TAMAÑO DE MUESTRA

2a. ETAPA (HERMOSILLO)



2.4. CUESTIONARIO

Existen diversos métodos para obtener la información necesaria para una encuesta: entrevista personal, observación directa, cuestionario por correo o pruebas de laboratorio. La decisión acerca de cuál de dichos métodos usar dependerá del tipo de información buscada y de las características de la población estudiada.²⁴

Considerando que el propósito de la presente investigación es obtener por un lado información acerca de las características del parque automotor y por otro los hábitos relacionados al uso de dicho parque se diseñaron unas formas en que se combinó la observación directa y la entrevista personal.

El cuestionario por correo se descartó desde un principio por el alto índice de no respuesta que suele observarse en este tipo de cuestionarios.

Se usó la observación directa en donde fuera factible por ser éste un muy buen método por estar libre de ciertos errores debido a los entrevistados como son la mala memoria y la exageración.²⁵

24. RAJ, La Estructura de las Encuestas por Muestreo, pag.160.

25. Idem

Sin embargo, la mayor parte de la información se obtuvo por el método de la entrevista, por ser la observación directa muy limitado para los alcances deseados en el estudio.

En la entrevista se utilizó un cuestionario, que es un conjunto estándar de preguntas que se plantea de la misma forma a cada uno de los encuestados.²⁶ Esto se hace para tener la certeza de que los resultados obtenidos sean comparables.

En los cuestionarios de acuerdo a la forma de las respuestas, las preguntas se pueden dividir en preguntas cerradas, cuya respuesta se limita a alternativas establecidas y preguntas abiertas en que el entrevistado tiene la libertad de contestar a su modo.²⁷

Se prefirieron las preguntas cerradas en el cuestionario, que tienen la ventaja de poder comparar siempre todas las respuestas. Sin embargo se debe tomar en cuenta que para ciertas preguntas existen respuestas no contempladas en el conjunto original de respuestas y que por consiguiente caen en la categoría de "otros", donde se pierde el nivel de detalle.

En la presente investigación para la recolección de datos, como ya

26. Ibid, pag. 163.

27. Ibid, pag. 169-170.

se había mencionado, se recurrió en primer lugar a la observación directa por parte del encuestador. Esto se hizo para la obtención de información sobre algunas características del parque automotor como modelo o cilindraje de los vehículos y algunos hábitos de uso como la cantidad de litros cargados o el número de pasajeros.

Si el encuestador por la simple observación no sabía algún dato, como el del cilindraje, lo debía preguntar al automovilista.

La mayoría de los datos no se obtuvieron por observación directa, sino que debían ser preguntados al automovilista. Como lo son la frecuencia de carga, la frecuencia de salida de la ciudad, el uso del transporte colectivo, etc.

Todos los datos, tanto los que se preguntaron a los automovilistas como los de observación directa, se vaciaron en unas formas, ocupándose una forma por unidad encuestada.

Un ejemplar de dichas formas se puede apreciar en la página a continuación.

Todos los encuestadores recibieron un cuidadoso entrenamiento para poder llenar adecuadamente las formas del cuestionario. Además se les entregó un manual para que lo consultaran en caso de que el supervisor no estuviera presente en el momento en que surgiera alguna duda.

ESTACION No. _____ FECHA: _____ SEPT. 85 HORA: _____
PLACAS DEL ESTADO DE: _____ No. DE PASAJEROS: _____

1. ¿Vive aquí en Veracruz? SI NO
2. ¿Está cargando gasolina para salir a Veracruz? SI NO

3. ¿A dónde va? _____
Localidad Edo.

¿A qué distancia? _____
4. ¿Cuál es el motivo de su viaje?
PASEO 1 TRABAJO 2 OTRO 3

5. ¿Cada cuándo sale de Veracruz este vehículo? _____
6. ¿Cada cuándo carga gasolina? _____
7. ¿Qué año, marca y modelo es su vehículo? _____

- año marca modelo
8. ¿Cuántos cilindros tiene su vehículo?
9. ¿Siempre ha funcionado el marcador de kilómetros? SI NO
10. ¿En qué vuelta está? _____ ¿Cuánto Marca? _____
11. ¿Ha calculado recientemente el rendimiento de su vehículo?
SI NO
12. ¿Cuál es? _____ Cd. _____ Carretera
13. ¿Cuánto tiempo pasa usted diariamente en el vehículo? _____
14. ¿A quien pertenece el vehículo?

PARTICULAR 1 EMPRESA PRIVADA 2 ORGANISMO PUBLICO 3

15. ¿Cuál es su ocupación?

DIRECTIVO O EMPRESARIO 1 PROFESIONISTA INDEPENDIENTE 2
PROFESIONISTA EMPLEADO 3 TRABAJADOR POR SU CUENTA 4
EMPLEADO 5 OBRERO 6 PEON 7 AMA DE CASA 3
ESTUDIANTE 9 DESEMPELADO 10 C T R O 11

16. ¿Usa frecuentemente el transporte colectivo?

NO PORQUE: TIENE COCHE 1 INSEGURIDAD 2 INCOMODIDAD 3
FALTA DE TRANSPORTE 4 LENTO 5 OTRO 6
SI PORQUE: PROBLEMAS VIALES 7 COSTO 8 TIEMPO 9
NO TIENE COCHE PROPIO 10 O T R O 11

17. ¿Qué problemas tiene para usted usar el transporte colectivo?

LENTO 1 TARDA EN PASAR 2
INCOMODO 3 NINGUNO 4 NO SE 5

18. Cantidad de litros cargados: _____

19. Cola de espera: 0 1 2 3 4 5 6 7 8 9 10

2.5. OPERATIVO DE CAMPO

En cada una de las ciudades seleccionadas se llevó a cabo el mismo procedimiento para el levantamiento de las encuestas para contar con datos comparables.

Considerando la importancia de poder comparar los datos sobre una misma base todas las encuestas se levantaron en martes, miércoles y jueves en un mismo mes para evitar fluctuaciones, como por ejemplo las ocasionadas por fines de semana o por vacaciones.

El primer paso que debía seguir el supervisor para la realización de la encuesta fué hacer una visita a cada una de las gasolineras previamente seleccionadas para explicar al responsable de la gasolinera el motivo de la encuesta y que ésta no interferiría con el movimiento normal de la misma.

Además se aprovechó esta visita para observar el movimiento de la gasolinera y el número de bombas de gasolina Nova en funcionamiento, para poder determinar el número de encuestadores que se necesitarían en cada caso.

Posteriormente se procedió a contratar a los encuestadores, para esto se buscaron personas con el siguiente perfil:

- Ambos sexos entre 20 y 30 años.
- Nivel de estudios mínimo preparatoria o equivalente.
- Experiencia en el levantamiento de encuestas.
- Buena presentación.
- Estar familiarizado con automóviles (Marcas, modelos, cilindraje, odómetro, etc.).
- Ser extrovertido

A todos los interesados se les dió una amplia explicación de los objetivos de la encuesta, así como del procedimiento para aplicar el cuestionario, detallando la manera en que debía ser contestada cada pregunta. Además individualmente se les aplicó un pequeño examen en el que el supervisor simulaba ser un automovilista a encuestar.

A partir de los resultados de este examen y del perfil antes mencionado se escogieron los mejores candidatos para participar en el levantamiento de la encuesta.

A cada encuestador se le asignó la gasolinera en la que debería levantar las encuestas, así como el turno, ya fuera en la mañana

de 7.00 a 14.00 Hrs. o en la tarde de 14.00 a 21.00 Hrs. y se les pidió que estuvieran en el lugar correspondiente media hora antes. Además se escogió un suplente por si algún encuestador llegara a faltar.

Una vez en la gasolinera a cada encuestador se le dió un cuaderno con 100 cuestionarios y se le indicó cual sería su bomba para levantar las encuestas, lugar que no debería abandonar, encuestando a todos los vehículos que cargaran gasolina ahí.

Para cada dos gasolineras había un supervisor para cuidar del adecuado levantamiento de las encuestas, ayudando a los encuestadores a resolver cualquier dificultad que se les presentara.

CAPITULO 3

ANALISIS DE RESULTADOS

En este capítulo se van a presentar algunos de los resultados que se obtuvieron en la encuesta, con el objeto de ilustrar el proceso completo de la obtención de información acerca del automóvil particular en México.

No es el objetivo del presente trabajo hacer una descripción exhaustiva de todos los resultados obtenidos a través de los datos recopilados, sino mostrar, en base a algunos resultados la utilización del estimador antes descrito.

Se van a comparar algunos resultados que se obtuvieron con el estimador con los resultados que se conseguirían si no se considerara el hecho de que cada automóvil tiene una probabilidad distinta de pertenecer a la muestra; es decir, bajo el supuesto de que todos los vehículos encuestados tuvieran la misma probabilidad de ser incluidos en la muestra.

Con esto se puede apreciar la magnitud de los errores si no se hubieran considerado los datos ponderados con la frecuencia de carga.

También se muestra que tan sensibles son las distintas variables a la frecuencia de carga. Cabe mencionar que esto difiere entre una

ciudad y otra, dependiendo de las características y hábitos de uso propios de cada ciudad. Esto también se va a ilustrar en este capítulo.

3.1. CARACTERISTICAS DEL PARQUE VEHICULAR

Entre las características mas importantes del parque vehicular que se analizaron en la encuesta tenemos:

Cilindraje

Edad

Para ilustrar con mayor claridad estas dos variables se presentará un análisis comparativo de dos ciudades que tienen características distintas.

Por un lado tenemos la ciudad de Saltillo que tiene un parque vehicular relativamente viejo. Esto nos lleva también a un número mayor de vehículos de 8 y 6 cilindros, debido a que anteriormente se fabricaba una mayor cantidad de este tipo de vehículos, mientras que en años recientes se fabrican en su mayoría vehículos de 4 cilindros.

Por otro lado tenemos a la ciudad de Veracruz con un parque vehicular más nuevo y por consiguiente con un cilindraje menor.

Cabe mencionar que debido a estas características Veracruz tiene un parque más homogéneo. Mientras que la desviación estándar para la edad promedio es de 6.5 para Saltillo es de sólo 4.6 para Veracruz.

3.1.1. CILINDRAJE

Para la ciudad de Saltillo tenemos que la frecuencia semanal de carga de los vehículos de 4, 6 y 8 cilindros está dada como sigue:

Número de Cilindros	Frecuencia semanal de carga	Número de Vehículos
4	2.55	631
6	3.39	544
8	3.77	445

Los vehículos de 8 cilindros cargan más frecuentemente debido a su mayor consumo de combustible.

La hipótesis de que la frecuencia semanal media de carga sea igual para los vehículos de 4, 6 y 8 cilindros es rechazada al hacer un análisis de varianza.

Esto nos lleva a concluir que si no ponderamos con el inverso de la frecuencia de carga vamos a obtener un dato distinto al real. Es decir, como los vehículos de 8 cilindros cargan más frecuentemente va a parecer que el cilindraje de nuestro parque es mayor al que es realmente.

De hecho se obtuvieron los siguientes resultados para la media de cilindraje:

Sin corregir	5.8
Corrigiendo con la frecuencia de carga	5.5

La frecuencia semanal media de carga para Veracruz está dada como sigue:

Número de Cilindros	Frecuencia semanal de carga	Número de Vehículos
4	2.62	737
6	2.97	323
8	3.61	276

La hipótesis de que la frecuencia semanal media de carga sea igual para todos los vehículos también fue rechazada al hacer un análisis de varianza.

Los resultados obtenidos para la media de cilindraje fueron los siguientes:

Sin corregir	5.3
Corrigiendo con la frecuencia de carga	5.2

La diferencia fue menor en el caso de Veracruz por tratarse de una población más homogénea. Ahí más de la mitad de los vehículos son de 4 cilindros y menos del 20% son de 8 cilindros, por lo que el resultado no se ve tan afectado como en el caso de Saltillo donde los vehículos de 8 cilindros son cerca de la tercera parte del total.

3.1.2 EDAD

Para analizar la frecuencia de carga se dividió la población en los siguientes grupos de edad:

70	vehículos modelo \leq 1970
80	vehículos modelo 1971-1980
85	vehículos modelo 1981-1985

El último grupo llega hasta 1985, año en que se llevó a cabo el levantamiento de la encuesta.

En Saltillo los resultados obtenidos fueron los siguientes:

Modelo del Vehículo	Frecuencia semanal de carga	Número de Vehículos
70	3.62	307
80	3.14	941
85	2.88	394

Los vehículos más viejos cargan con mayor frecuencia debido a que consumen más combustible. Los motores nuevos están diseñados para dar un mayor kilometraje por litro.

La hipótesis de que la frecuencia semanal de carga es igual para los vehículos de distintos modelos también es rechazada al hacer un análisis de varianza.

Los resultados obtenidos para la media de modelo del parque vehicular fueron:

Sin corregir	75.3
Corrigiendo con la frecuencia de carga	75.8

Utilizando el estimador se puede apreciar que el parque vehicular es más nuevo de lo que parece al observar simplemente los vehículos que acuden a cargar gasolina.

Por otro lado en Veracruz la hipótesis de que la frecuencia semanal de carga es igual para los vehículos de los distintos modelos no pudo ser rechazada haciendo un análisis de varianza. Sin embargo, en la siguiente tabla sí podemos apreciar una diferencia, que aunque sea pequeña sí nos muestra una tendencia en los vehículos más viejos a cargar con mayor frecuencia.

Modelo del Vehículo	Frecuencia semanal de carga	Número de Vehículos
70	3.05	98
80	2.98	701
85	2.81	545

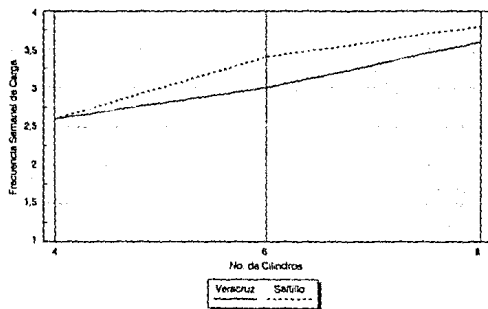
Los resultados obtenidos para la media de modelo del parque vehicular en la ciudad de Veracruz fueron:

Sin corregir	78.5
Corrigiendo con la frecuencia de carga	78.7

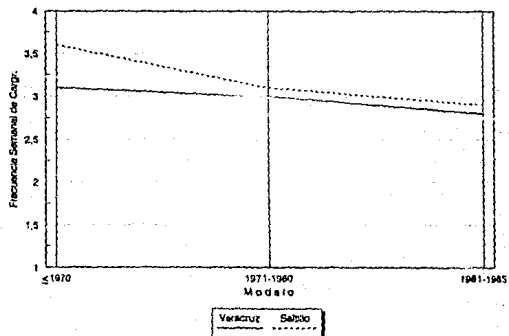
Al igual que para el cilindraje la diferencia fue menor en el caso de Veracruz. Esto se debe a que también en lo referente a la edad del parque vehicular la población es más homogénea. Sólo alrededor de un 5% de los vehículos de Veracruz son modelo 1970 o anterior, mientras que en el caso de Saltillo este grupo lo constituye casi el 20% de los vehículos.

FRECUENCIA DE CARGA

a) Por Cilindraje

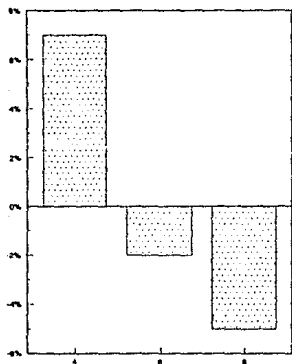


b) Por Edad

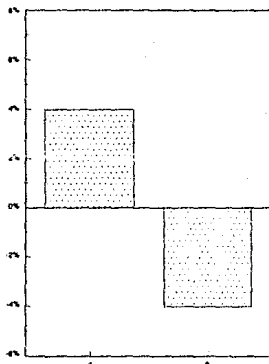


CILINDRAJE

CORRECCION POR FRECUENCIA DE CARGA

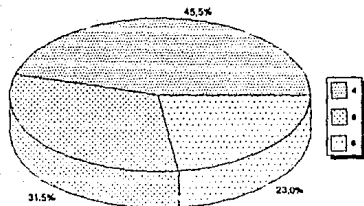


Jalisco

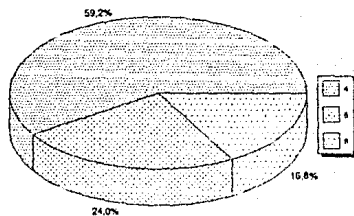


Veracruz

COMPOSICION CORREGIDA DEL PARQUE VEHICULAR



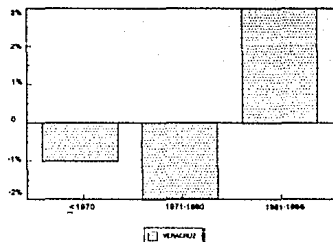
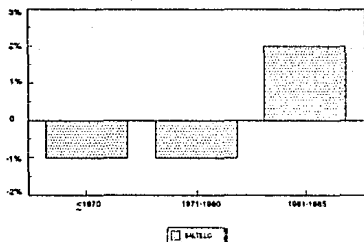
Jalisco



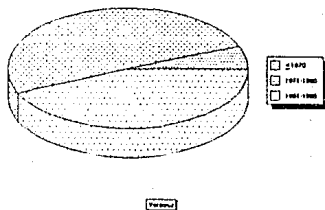
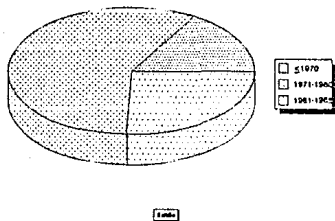
Veracruz

EDAD

CORRECCION POR FRECUENCIA DE CARGA



COMPOSICION CORREGIDA DEL PARQUE VEHICULAR



3.2. HABITOS DE USO DEL PARQUE VEHICULAR

Los hábitos de uso se analizaron por separado porque no son inherentes a la flota vehicular en sí, sino que dependen de factores de tipo externo, como por ejemplo las características de la ciudad en que circulan los vehículos.

En lo referente a los hábitos de uso se presentan a continuación dos que son de gran importancia:

Kilometraje Anual Recorrido

Consumo Semanal

Al igual que para las características del parque vehicular se van a estudiar ambas variables en dos ciudades con características distintas para hacer un análisis comparativo.

Se van a comparar los resultados obtenidos en Tuxtla Gutiérrez, que es una ciudad pequeña, con los de Guadalajara, que es la segunda ciudad más grande de México.

3.2.1. KILOMETRAJE ANUAL RECORRIDO

Este dato se calculó individualmente para cada vehículo a partir de su modelo y de la lectura directa del odómetro. Cabe mencionar que en la gran mayoría de los odómetros instalados en los

vehículos en circulación al momento de la encuesta sólo marcaban hasta 99,999 Km., volviendo a empezar su cuenta en cero. Debido a esto también se consideró la vuelta del odómetro para estos vehículos.

Entonces el kilometraje anual recorrido de un vehículo se calculó como sigue:

$$KM = \left[\frac{(V - 1) (100,000) + D}{M} \right] 12$$

Donde:

KM = Kilometraje anual recorrido

V = Vuelta del odómetro

D = Lectura del odómetro

M = Meses en circulación del vehículo

Los meses en circulación se consideraron como sigue:

$$M = 12 (84 - A) + 14.5 \quad \text{si } A \neq 85$$

$$M = 4.5 \quad \text{si } A = 85$$

Donde:

A = Modelo del vehículo

Los meses en circulación se calcularon considerando que un vehículo entra en circulación en promedio el mes de junio de cada año. Entonces para mediados de septiembre, cuando se levantaron las encuestas, se encontraban los automóviles 14.5 meses en circulación a partir de 1984.

Para ilustrar como cambia la frecuencia de carga en relación al kilometraje que recorren anualmente se dividió a la población en:

5,000	Vehículos que recorren hasta 10,000 km.
15,000	Vehículos que recorren entre 10,000 y 20,000 km.
25,000	Vehículos que recorren entre 20,000 y 30,000 km.
35,000	Vehículos que recorren entre 30,000 y 40,000 km.
50,000	Vehículos que recorren más de 40,000 km.

Para estos grupos se obtuvieron los siguientes resultados en la ciudad de Tuxtla Gutiérrez:

Kilometraje anual recorrido	Frecuencia semanal de carga	Número de Vehículos
5,000	2.31	163
15,000	2.36	376
25,000	2.70	234
35,000	2.91	78
50,000	3.29	83

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

En esta tabla se puede apreciar que los vehiculos que recorren más kilómetros cargan con mayor frecuencia porque tienen mayores requerimientos de combustible.

La hipótesis de que la frecuencia semanal de carga es igual para todos los vehiculos independientemente del kilometraje recorrido es rechazada al hacer un análisis de varianza, al igual que para las variables anteriores.

De hecho los resultados obtenidos en Tuxtla Gutiérrez para el kilometraje anual medio recorrido fueron los siguientes:

Sin corregir	20,723
Corrigiendo con la frecuencia de carga	19,027

Aquí se puede apreciar que de no corregir por el inverso de la frecuencia de carga se sobreestima el kilometraje anual medio recorrido. La diferencia es de casi dos mil kilómetros, lo cual no es una cantidad despreciable.

Por otro lado se analizó la ciudad de Guadalajara, donde los resultados obtenidos para la frecuencia de carga por kilometraje anual recorrido fueron los siguientes:

Kilometraje anual recorrido	Frecuencia semanal de carga	Número de Vehiculos
5,000	2.08	236
15,000	2.12	536
25,000	2.29	249
35,000	2.40	81
50,000	2.47	46

Se puede notar una clara tendencia ascendente en la frecuencia de carga conforme aumenta el kilometraje anual medio recorrido.

Sin embargo, la hipótesis de que la frecuencia semanal de carga es igual para todos los vehículos no pudo ser rechazada. De hecho las diferencias en la frecuencia de carga son pequeñas.

En Guadalajara se obtuvieron los resultados expuestos a continuación para el kilometraje anual medio recorrido:

Sin corregir	17,924
Corrigiendo con la frecuencia de carga	17,312

Se aprecia que la diferencia es menor en el caso de Guadalajara. Esto se debe probablemente a que existe una mayor uniformidad en la frecuencia de carga.

También cabe resaltar que en una ciudad grande como Guadalajara se recorran menos kilómetros que en una pequeña como lo es Tuxtla Gutiérrez.

Lo anterior se debe principalmente a los viajes interurbanos que son más frecuentes en localidades pequeñas.

3.2.2. CONSUMO SEMANAL

El análisis de esta variable es indudablemente de los más importantes. Ya que para implementar medidas de racionalización en el consumo de combustible hay que conocer en primer lugar a cuánto asciende dicho consumo.

Para calcular el consumo semanal por vehículo se consideró el producto de la cantidad de litros cargados por la frecuencia semanal de carga. Bajo el supuesto de que la carga de combustible el día de la encuesta haya sido representativa de la carga de combustible promedio del vehículo.

De hecho se observó que más de un 60% de los vehículos cargaron por una cantidad fija de dinero. En septiembre de 1985 el 15% de los vehículos cargó 9.1 lt de gasolina que eran \$ 500.--, el 32% de la población cargó 18.2 lt. que eran \$ 1,000.-- y el 14% cargó 36.4 lt. que eran \$ 2,000.--

Además el 15% de la población cargó tanque lleno. Esto nos muestra un cierto patrón de comportamiento de los automovilistas al cargar combustible.

Los grupos que se hicieron para ilustrar la frecuencia de carga en relación al consumo semanal fueron los siguientes:

20	Vehículos que consumen hasta 30 lt.
50	Vehículos que consumen entre 30 y 60 lt.
80	Vehículos que consumen entre 60 y 90 lt.
110	Vehículos que consumen más de 90 lt.

En Tuxtla Gutiérrez se obtuvieron los siguientes resultados:

Consumo semanal de combustible	Frecuencia semanal de carga	Número de Vehículos
20	1.21	461
50	1.90	477
80	3.09	223
110	5.18	315

Efectivamente, los vehículos que consumen más combustible acuden con mayor frecuencia a cargar gasolina. De hecho ésta es la variable en la que se pueden apreciar las diferencias más pronunciadas.

La hipótesis de suponer una igual frecuencia de carga para los distintos consumos de gasolina es rechazada al hacer un análisis de varianza.

La marcada diferencia en las frecuencias de carga queda reflejada en los resultados obtenidos en Tuxtla Gutiérrez para el consumo semanal de combustible:

Sin corregir	67
Corrigiendo con la frecuencia de carga	40

Para esta variable la utilización del estimador es de suma importancia para no sobreestimar el consumo en más de un 65%.

Por otro lado en Guadalajara los resultados obtenidos son los que aparecen en la tabla a continuación:

Consumo semanal de combustible	Frecuencia semanal de carga	Número de Vehiculos
20	1.20	578
50	1.74	696
80	3.10	309
110	4.99	308

La hipótesis de medias iguales en la frecuencia de carga fue

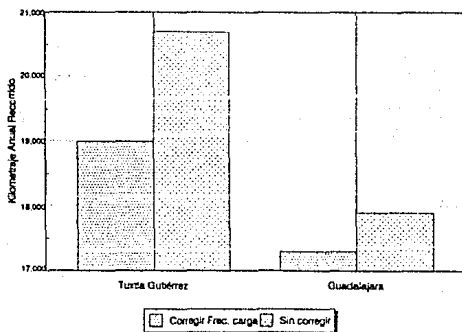
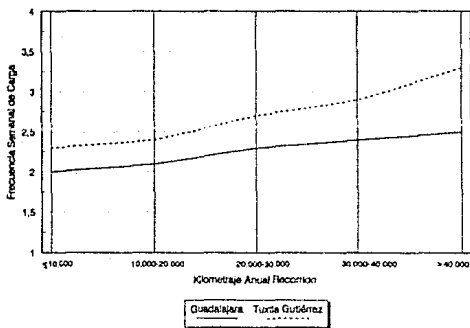
rechazada al igual que para Tuxtla Gutiérrez al hacer un análisis de varianza.

También aquí se observa una gran diferencia en el consumo de combustible al usar el estimador:

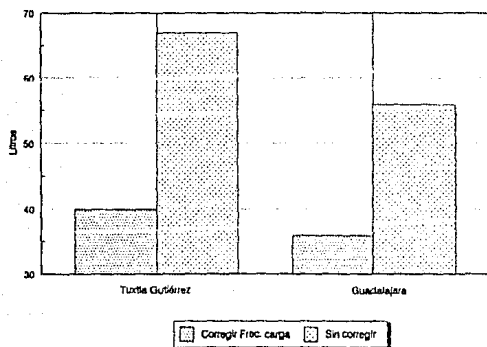
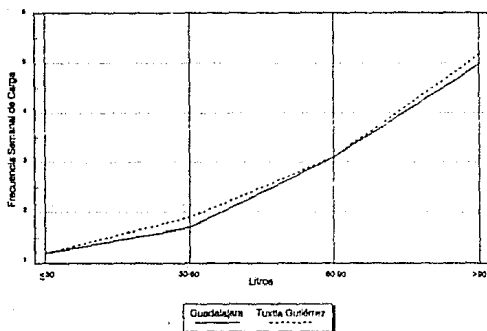
Sin corregir	56
Corrigiendo con la frecuencia de carga	36

Al igual que para el kilometraje anual medio recorrido la diferencia es mayor en Tuxtla Gutiérrez. Esto se debe probablemente a que aquí se utiliza más el automóvil para hacer recorridos interurbanos.

KILOMETRAJE ANUAL RECORRIDO



CONSUMO SEMANAL DE COMBUSTIBLE



CONCLUSIONES

En primer lugar se puede señalar que las ventajas aparentes que mostraban las gasolineras como centros para generar información confiable y de manera económica fueron corroboradas a lo largo del levantamiento.

Las horas-encuestador fueron aprovechadas en un porcentaje muy elevado para generar información útil, a diferencia de otras formas de generación de información como pueden ser las encuestas domiciliarias, en las que los encuestadores dedican una parte importante de su tiempo a buscar información. Lo anterior tiene como consecuencia un muy bajo costo por observación generada en nuestro caso.

Por otra parte el hecho de que los automovilistas tengan tiempo ocioso antes de cargar gasolina permitió que colaboraran con los encuestadores de manera muy favorable.

Por lo que respecta al estimador propuesto se demostró teóricamente su insesgamiento y prácticamente debido a los volúmenes de información generados se constató que el sesgo es despreciable.

A título ilustrativo se presentaron en el trabajo algunas estimaciones de características y hábitos de uso de los

automóviles. Para fines de establecer políticas sobre uso de los vehículos y de consumo de combustible se pueden obtener otras estimaciones que resulten pertinentes a través de un proceso análogo al mostrado.

Con base en lo anterior se podría diseñar y afinar un levantamiento periódico que permitiera dar un seguimiento sistemático y en forma permanente a las características y hábitos de uso del parque automotor a fin de instrumentar oportunamente medidas en el ámbito del transporte urbano.

Lo anterior resultaría conveniente dada la escasa información confiable existente sobre el parque vehicular, así como su gran impacto energético y ambiental.

ANEXO

TABLA 1: CLASIFICACION DE CIUDADES

	Ciudad	Número de Habitantes	Automóviles registrados
Grupo 1	Distrito Federal *	14'961,300	1'911,461
Grupo 2	Guadalajara *	2'353,000	407,090
	Monterrey *	2'133,600	205,749
	Puebla *	1'069,300	154,666
	León *	814,600	36,661
	Torrcón *	654,600	60,155
	Total	7'025,100	864,321
Grupo 3A	Ciudad Juárez	559,900	182,513
	Tijuana *	494,600	136,009
	Mexicali	379,500	124,426
	Total	1'434,000	442,948
Grupo 3B	San Luis Potosí *	523,600	32,847
	Mérida *	497,600	52,387
	Tampico *	470,000	27,646
	Chihuahua *	431,100	33,142
	Veracruz *	391,100	29,091
	Morelia	368,600	25,788
	Acapulco	355,900	22,256
	Culiacán	355,200	21,606
	Hermosillo	332,200	27,432
	Aguascalientes	317,000	31,526
	Saltillo	316,300	38,374
	Total	4'358,600	342,095
Grupo 4A	Toluca	252,300	44,980
	Querétaro	244,700	35,529
	Reynosa	210,800	53,576
	Cuernavaca	205,100	48,426
	Tuxtla Gutiérrez	152,800	20,674
	Coatzacoalcos	141,800	18,287

* Se incluye toda la zona metropolitana.

	Ciudad	Número de Habitantes	Automóviles registrados
Grupo 4A (cont.)	Ensenada	138,800	76,490
	La Paz	107,400	20,900
	Colima	92,700	25,577
	Piedras Negras	69,700	17,842
	Nogales	69,100	13,137
	Guasave	66,800	9,177
	Total	1'752,000	384,595
Grupo 4B	Durango	284,300	16,525
	Orizaba	249,200	8,027
	Mazatlán	229,000	10,302
	Jalapa	224,100	16,053
	Nuevo Laredo	216,400	21,689
	Matamoros	202,900	20,220
	Villa Hermosa	195,700	17,245
	Irapuato	189,800	12,019
	Oaxaca	183,800	20,156
	Ciudad Obregón	179,600	12,531
	Poza Rica	175,900	7,837
	Celaya	164,700	12,485
	Tepic	163,700	11,458
	Ciudad Victoria	156,900	10,089
	Uruapan	145,600	8,650
	Los Mochis	142,600	9,177
	Campeche	126,600	9,590
	Monclova	124,100	12,971
	Pachuca	116,000	12,055
	Minatitlán	115,400	7,664
	Salamanca	109,700	7,720
	Córdoba	103,700	9,356
	Zamora	103,500	6,001
	Tapachula	93,600	6,437
	Zacatecas	90,900	6,686
	Tehuacán	87,500	10,134
	San Luis Río Colorado	84,100	8,790
	Chilpancingo	80,800	5,934
	Parral	77,500	4,797
	Ciudad Mante	76,000	4,172
	Iguala	75,100	3,648
Ciudad del Carmen	73,300	7,001	
Ciudad Valles	70,900	4,161	
Navojoa	68,100	4,968	
Delicias	66,500	7,463	
Apatzingán	63,400	2,468	
Chetumal	63,300	6,570	
Ciudad Guzmán	63,100	3,430	

	Ciudad	Número de Habitantes	Automóviles registrados
Grupo 4B (cont.)	Tuxpan	61,200	3,971
	Fresnillo	60,700	3,323
	Río Bravo	59,800	5,169
	Tulancingo	57,600	3,481
	Atlixco	55,400	4,012
	La Piedad	55,200	4,431
	Zitácuaro	54,700	4,345
	Guaymas	54,600	6,006
	Guanajuato	53,700	3,675
	Ocotlán	51,600	2,859
	Total	5'601,800	407,751

TABLA 2:

MUESTRA DE ESTACIONES DE SERVICIO PARA DETERMINAR EL POSIBLE NUMERO DE AUTOMOVILES A ENCUESTAR EN CADA GASOLINERA

Ciudad	Gasolinera No.	Litros anuales vendidos	Posible No. de Automóviles(A)
Campeche	365	6'854,000	260
	375	10'154,892	386
	2778	6'259,544	238
Mazatlán	1762	9'198,000	349
	1764	3'250,000	123
	1765	11'025,000	419
Tuxtla Gutiérrez	336	4'165,530	158
	352	7'020,000	267
	356	8'520,000	324
Cuernavaca	468	6'810,000	259
	3040	12'720,000	483
	3079	6'892,632	262
Hermosillo	1643	12'555,120	477
	1645	3'487,274	132
	1649	10,650,000	405
	3431	5'842,642	222
Saltillo	2240	4'770,000	181
	2247	6'972,177	265
	2251	10'020,000	381
	2255	8'619,097	327
Veracruz	1005	5'245,000	199
	1031	9'873,289	375
	1033	13'981,174	531
	1037	6'425,944	244
Cd. Juárez	1138	3'045,000	116
	1142	3'561,306	136
	1148	4'965,000	189
	1151	5'820,000	221

Ciudad	Gasolinera No.	Litros anuales vendidos	Posible No. de Automóviles (A)	
León	1710	9'039,830	343	
	1712	6'750,000	256	
	1714	8'850,000	336	
	1716	10'920,000	415	
	1724	4'713,505	179	
	1725	7'602,790	289	
	3050	5'151,317	196	
	3053	11'745,000	446	
	Guadalajara	1515	7'119,655	270
		1519	8'085,000	307
1520		5'310,000	202	
1532		7'755,000	295	
1541		13,560,000	515	
1542		10'740,000	408	
1550		5'655,000	215	
2529		12'003,482	456	
Distrito Federal		19	6'210,000	236
	44	12'120,000	460	
	69	5'760,000	219	
	75	9'915,000	377	
	107	16'155,000	614	
	135	9'420,000	358	
	148	11'640,000	442	
	184	7'005,000	266	
	202	18,195,000	691	
	212	10'305,000	391	
	217	7'095,000	270	
	301	10'650,000	405	
	2785	13'140,000	499	
	2927	24'381,250	926	
	3383	3'780,000	144	
	3451	10'231,500	389	
Resumen de Resultados		Promedio	329	
		Máximo	926	
		Mínimo	116	

REFERENCIAS BIBLIOGRAFICAS

ABAD ADELA & SERVIN LUIS
Introducción al Muestreo
Limusa, México D.F., 1978.
p.p. 200

ARAUJO CARLOS & CANTABRESES JOSE LUIS
Una Aplicación de la Teoría Estadística a la Resolución de Problemas de Información Energética en Países en Vías de Desarrollo.
Revista Energética de la Organización Latinoamericana de Energía
Año 8 6/84, pag. 49-56.

COCHRAN WILLIAM G.
Técnicas de Muestreo
Trad. Andrés Sestier Bouclier
C.E.C.S.A., México D.F., 1984.
p.p. 513

KISH LESLIE
Muestreo de Encuestas
Trad. Ricardo Vinós Cruz López
Trillas, México D.F., 1975.
p.p. 739

MOOD ALEXANDER M. & GRAYBILL FRANKLIN A.
Introducción a la Teoría de la Estadística
Trad. Rafael Pro Bermejo
Aguilar, Madrid, 1978.
p.p. 536

MOOD ALEXANDER M., GRAYBILL FRANKLIN A. & BOES DUANE C.
Introduction to the Theory of Statistics
Mac Graw Hill, New York, 1981.
p.p. 564

RAJ DES
La Estructura de las Encuestas por Muestreo
Trad. Eduardo L. Suárez
Fondo de Cultura Económica, México D.F., 1979.
p.p. 475

RAJ DES

Teoría del Muestreo

Trad. Roberto R. Reyes-Mazzoni

Fondo de Cultura Económica, México D.F., 1980.

p.p. 305

SNEDECOR GEORGE W. & COCHRAN WILLIAM G.

Metodos Estadísticos

Trad. J. A. Reinos Fuller

C.E.C.S.A., México D.F., 1984.

p.p. 703