

Otra copia:

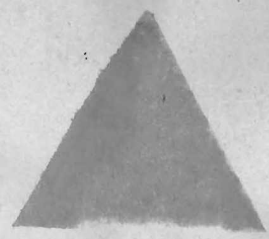
- ① Elizaga
- ② Mauro
- ③ Lopez
- ④ Sylvia

8/10/77 20  
 ① ~~Dr. Louge~~  
 ② ~~Dr. Barros~~  
 ③ DOCPAL

CELADE  
 DOCUMENTO  
 MICROFILMADO  
 DOCPAL

RESERVADO

Abril de 1979



INFORME DE LA MISION DE ASESORIA REALIZADA EN LA  
 REPUBLICA DE MEXICO DESDE EL 28 DE FEBRERO  
 HASTA EL 20 DE MARZO DE 1979

Carlos Cavallini

Asesor Regional en Muestreo para Estadísticas Demográficas  
 adscrito a la CEPAL

"El autor de este informe es el único responsable de las opiniones expresadas en él. Se han presentado copias a la Oficina de Cooperación Técnica de las Naciones Unidas, la cual, a su debido tiempo, podrá comunicar al Gobierno sus propias conclusiones y recomendaciones."

04112.00=No per

CAVALLINI, Carlos (Au)

Informe de la mision de asesoria realiza

28 de Febrero hasta el 20 de Marzo

Abril 1979; Pags:28

Editorial: CEPAL. Santiago CL

Idicma:Es Distr:Confid Impresion:Mime

Pais/region principal:MX Paises trata

Descriptor:<PLAN DE MUESTREO\*> <METODO

<ENCUESTA DE HOGARES>

Categ. Revista:<PEEL:MICOL>

Fechas datos demogr:9999-9999 No. de

Analisis de los problemas de confia  
Encuesta de Hogares de la Direccion Gene  
marco muestral data del censo de poblacion  
estadistico para conocer las distintas fi  
describela elaboracion de un nuevo marco  
Hoja Resmen de Empadronador Censal a emp  
1980, as como el diseno muestral (muest  
encuesta de hogares que reemplazara a la  
(If. interna para DCCPAL: ISIS=09

## MISION EN LA REPUBLICA DE MEXICO

1. En atención a una solicitud de cooperación técnica estadística, emanada del Gobierno Mexicano y cursada a la CEPAL a través del Representante Residente del Programa de las Naciones Unidas para el Desarrollo - PNUD - de dicho país, se cumplió una misión oficial de asesoramiento en la Dirección General de Estadística - DGE - de México.
2. La misma se realizó desde el 28 de febrero hasta el 20 de marzo de 1979, es decir por un período de 3 semanas.
3. Los alcances de la misión y los delineamientos generales del trabajo a realizar fueron establecidos con la Lic. Leticia Calzada de Vielle, Sub-Directora General de Planeación de la DGE. La labor específica a desarrollar se trató, fundamentalmente, con el Est. Raúl Hudlet, Ph.D., University of Wisconsin, Jefe de la Sub-Dirección Auxiliar de Investigación y Proyectos de la DGE. La contraparte nacional estuvo integrada, además del señor R. Hudlet, por los señores Jesús Romo, Dr. en Estadística, Imperial College of London, Jefe del Depto. de Análisis Estadístico, Fabián Hernández, Ph.D., University of Wisconsin, Jefe del Depto. de Muestreo y por la señorita Mat. Analía Babinsky, Auxiliar de Investigación y Proyectos de la Sub-Dirección.
4. En distintas oportunidades se mantuvieron reuniones y se hicieron consultas, relativas a los distintos temas del trabajo a realizar, con los señores Ing. Luis Eugenio Miranda V., Jefe de la Dirección General de Estudios del Territorio Nacional - DETENAL -; Act. Saul Kletzel, Sub-Director General de Apoyo a la Coordinación del Sistema Nacional de Información de la Dirección General de Sistemas y Procesos Electrónicos; Fis. Javier Games, Sub-Director Auxiliar de la Dirección General de Sistemas y Procesos Electrónicos; Act. Jesús Salinas, Coordinador de Censos; y Lic. José Luis Cerda, Coordinador del Censo de Población.
5. En todo momento se mantuvo informado del progreso de los trabajos al Lic. Luis Eduardo Rosas Landa, Director General de la DGE.
6. En distintas oportunidades se mantuvieron reuniones con los señores Benjamín Gurman, Adjunto del PNUD; Gert Rosenthal, Director de la Sub sede de la CEPAL en México; Horacio Santamaría, experto de la CEPAL y Stephen A. Grun, oficial de programa del PNUD, quienes de una u otra forma facilitaron la labor del experto.

7. El programa de trabajo a desarrollar comprendió los siguientes puntos principales:

- i) Encuesta de Hogares basada en el censo de 1970;
- ii) Diseño para la confección de un Marco Muestral basado en el próximo censo de 1980;
- iii) Diseño de una nueva Encuesta de Hogares basada en el Marco Muestral de 1980.

Encuesta de Hogares basada en el censo de 1970

8. La Encuesta de Hogares, que actualmente se está llevando a cabo, está basada en el censo de población y vivienda del año 1970. La misma que cuenta con un tamaño muestral que se puede considerar sobredimensionado para este tipo de investigaciones, alrededor de 9 000 viviendas particulares en el Area Metropolitana de México son medidas cada 3 meses, se estima que carece de una confiabilidad adecuada, además de presentar problemas operativos y técnicos de agotamiento de las manzanas. Es decir, debido a los 9 años transcurridos del levantamiento censal, que sirvió como marco de selección muestral, la muestra ha perdido actualización y está dejando de poseer la representatividad que se desea. Por ello, de todas las alternativas que se presentaron para contar con una muestra más eficiente, se decidió, por razones técnicas y de costo que era conveniente continuar con esta muestra hasta el próximo censo de 1980, aplicándosele algunas modificaciones que permitan en forma simple y rápida conocer la medida de su confiabilidad. Mientras tanto, se diseñará una nueva muestra que estará basada en el marco muestral que proveerá el censo de 1980 y que reemplazará a la actual muestra a partir de ese momento.

9. Para conocer la confiabilidad de los principales resultados de la actual encuesta, se propuso replicar la investigación. Esta técnica provee, principalmente, las siguientes ventajas:

- i) fácil cálculo de la varianza del estimador;
- ii) hacer un análisis de la varianza, de tal manera de conocer las fuentes de variación que han influido en la variación total, cuantificar los efectos causados por estas fuentes de variación y estimar el grado de significación de cada uno de ellos con el objeto de modificar y adecuar los tamaños muestrales y aumentar la eficiencia por unidad de costo;

/iii) poseer

iii) poseer varias submuestras, todas ellas muestras representativas de la población en igual grado, las cuales podrán ser utilizadas en forma separada o conjunta según las características que se deseen medir, los recursos que se dispongan, etc.;

iv) medir el sesgo del o de los entrevistadores;

v) obtener varias estimaciones para una característica dada, lo cual otorga a simple vista una apreciación de la confiabilidad de la investigación.

10. El modelo que aquí se presenta se aplicará en el Area Metropolitana de México, para un estrato determinado, hasta la ejecución del censo de población de 1980. En el caso de no considerarse la estratificación, el modelo abarcará al Area Metropolitana de México como un todo. Suponiendo tener  $r$  replicaciones de igual tamaño de viviendas particulares,  $v$  entrevistadores y  $w$  viviendas por celda replicación-entrevistador, el modelo lineal estadístico aditivo que se postula para el estudio de cualquier variable observable de la encuesta, que presente variación aleatoria en sus valores es,

$$y_{hij} = u + b_h + d_{hi} + e_{hij}$$

donde

$y_{hij}$             valor observado en la hij-vivienda  
con  $h = \overline{1;r}$     replicación  
           $i = \overline{1;v}$     entrevistador  
           $j = \overline{1;w}$     vivienda

$u$                 media parámetro  
 $b_h$               efecto entre replicaciones  
 $d_{hi}$              efecto entre entrevistadores dentro de la replicación  
 $e_{hij}$             error aleatorio por las características específicas de la hij-observación.

11. Este modelo permitirá conocer cómo contribuye a la variación total cada uno de los componentes, o efectos de variación, que responden a una determinada fuente de variación. En este caso las fuentes de variación son las replicaciones, los entrevistadores y el error aleatorio. Si se puede

/determinar cual

determinar cual de estas fuentes contribuye significativamente a la variación total se podrá adecuar el tamaño muestral, por ejemplo, disminuyendo los tamaños donde la variación es poca y aumentándolos donde sea grande. Además, se podrá controlar la variación debido al efecto del o de los entrevistadores.

12. Para cuantificar los distintos componentes de variación se establece la siguiente identidad algebraica que desglosa a la varianza total en sus componentes de variación,

$$\begin{aligned} & \sum_h^r \sum_i^v \sum_j^w (y_{hij} - \bar{y})^2 \\ = & \sum_h^r \sum_i^v \sum_j^w (y_{hij} - \bar{y}_{hi})^2 \\ + & w \sum_h^r \sum_i^v (\bar{y}_{hi} - \bar{y}_h)^2 \\ + & wv \sum_h^r (\bar{y}_h - \bar{y})^2 + 0 \quad (\text{los dobles productos se anulan por} \end{aligned}$$

haber en cada sumando sumas de desvíos iguales a cero),

donde,

$$\bar{y}_{hi} = \frac{1}{w} \sum_j^w y_{hij} \quad \text{media de la hi-celda}$$

$$\bar{y}_h = \frac{1}{vw} \sum_i^v \sum_j^w y_{hij}$$

$$= \frac{1}{v} \sum_i^v \bar{y}_{hi} \quad \text{media de la h-replicación}$$

$$\bar{y} = \frac{1}{rvw} \sum_h^r \sum_i^v \sum_j^w y_{hij}$$

$$= \frac{1}{r} \sum_h^r \bar{y}_h \quad \text{media del esquema.}$$

13. Se puede confeccionar así el siguiente cuadro del análisis de la varianza.

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios
Entre replicaciones	$r - 1$	$A = wv \sum_h^r (\bar{y}_h - \bar{y})^2$	$\frac{A}{r-1}$
Entre entrevistadores dentro de la replicación	$(v-1)r$	$B = w \sum_h^r \sum_i^v (\bar{y}_{hi} - \bar{y}_h)^2$	$\frac{B}{(v-1)r}$
Error	$(w-1)rv$	$C = \sum_h^r \sum_i^v \sum_j^w (y_{hij} - \bar{y}_{hi})^2$	$\frac{C}{(w-1)rv}$
Total	$rvw-1$	$A+B+C$	

14. Evaluados los cuadrados medios se podrá aplicar una docimasia o test de hipótesis para determinar si alguna de las fuentes de variación consideradas ha tenido una influencia significativa en la variación total de la variable considerada.

15. Por otro lado, cada replicación proporciona una estimación del parámetro utilizando la misma fórmula del estimador que la diseñada para la muestra completa.

Llamando con  $\hat{Y}_h$  al estimador del total para la h-replicación es

$$\hat{Y} = \frac{1}{r} \sum_h^r \hat{Y}_h$$

el estimador total del esquema, cuya varianza estimada es

$$\text{var}(\hat{Y}) = \frac{1}{r(r-1)} \sum_h^r (\hat{Y}_h - \hat{Y})^2$$

/donde se

donde se ha ignorado el factor finito de corrección por ser la fracción de muestreo significativamente pequeña.

16. Evaluado  $\hat{Y}$  y su varianza, bajo el supuesto que los  $\hat{Y}_h$  siguen una distribución normal, se podrá aplicar por ejemplo la distribución t-Student, para r-1 grados de libertad, a los niveles del 1% y del 1<sup>o</sup>/oo, para conocer la sensibilidad de los resultados, es decir, si las submuestras representan a la población de estudio en igual grado.

17. En base al análisis de la varianza y al test de la t-Student, se podrá conocer, principalmente, la confiabilidad de los resultados de la actual encuesta, modificar los tamaños de la misma de ser necesario y conocer la eficiencia de la información recogida por los entrevistadores. Además otorgará experiencia técnica y operacional que será valiosa para aplicarla en la nueva encuesta basada en el censo de 1980.

#### Marco muestral

18. El avance constante de las técnicas estadísticas, que tienden a "dar más de menos" o en otras palabras, a aumentar la confiabilidad de la información por unidad de costo, están basadas, muchas de ellas, en un correlativo mejoramiento de la infraestructura estadística. Esta infraestructura comprende, entre otros, la cartografía, los marcos muestrales, estratos y dominios de estudio tanto geográficos como socioeconómicos bien identificados, técnicos estadísticos, agentes de campo eficientes y los equipos de procesamiento. Los resultados que se obtienen son, generalmente, función de la acuracidad de estos elementos y en casos, cuando estos elementos no se tienen o carecen de validez, los resultados pueden ser tambaleantes y económicamente caros en relación a su confiabilidad.

El marco muestral está considerado como una de las herramientas más importantes para los trabajos estadísticos.<sup>1/</sup>

El marco ha sido definido como la entrada al universo o población que se quiere investigar. Las muestras que se seleccionen de marcos inadecuados o anacrónicos que no representen a la población de estudio, arrojarán resultados valederos para esos marcos, pero no para la población para la cual

---

<sup>1/</sup> S.S. Zarcovich. "Quality of Statistical Data". FAO, Rome 1966.



se quiere inferir. Así, por ejemplo, la muestra actual está basada en el censo de 1970, que sirvió de marco muestral de selección. Este marco no se ha mantenido actualizado y se conoce que las grandes manchas urbanas en México han tenido zonas de crecimiento lento y otras zonas de crecimiento rápido. Si una muestra no contempla en su diseño este fenómeno, podrá a través del tiempo, perder representatividad. Uno de los grandes valores que tienen los censos es el de proveer marcos, que de mantenerlos actualizados permitirán durante todo el período inter-censal seleccionar muestras adecuadas. Si bien la actualización de un marco censal puede resultar costosa, se puede mantener actualizado un sub-marco censal o muestra representativa del marco censal. Esta actualización se puede realizar, por ejemplo, a través de las encuestas que se llevan normalmente, o a través de trabajos o encuestas ad-hoc, o a través de conocimientos derivados de otras fuentes, etc.

19. México tiene programado levantar su censo de población y vivienda del 28 de mayo al 11 de junio de 1980. Será un censo de derecho. En el mismo se han de utilizar alrededor de 1.4 millón de empadronadores con una carga de trabajo de unas 10 viviendas-empadronador por día. Para tratar la confección de un marco muestral, el día 14 de marzo de 1979 se llevó a cabo una reunión con los funcionarios encargados de los censos, señores Jesús Salinas, José Luis Cerda, Miguel Chavira y Carlos Salinas y por la otra parte la Sra. Leticia Calzada de Vielle, el Sr. Raúl Hudlet y el suscrito. En la misma se acordó que la Hoja Resumen del Empadronador Censal - HREC - será la unidad básica primera para conformar un marco muestral. Esta HREC tendrá también como objetivos el de servir como control de la labor de campo del empadronador y para proveer los primeros resultados del censo, en lo que respecta, principalmente, al número de personas por sexo y área geográfica y al número de viviendas por área geográfica.

20. Como borrador para discusión se presentó el siguiente diseño que podría tener la HREC.

Hoja Resumen del Empadronador Censal

Empadronador: Nombre: \_\_\_\_\_ Localidad: Nombre: \_\_\_\_\_  
 N°: \_\_\_\_\_ N°: \_\_\_\_\_  
 Jefe de Manzana: Nombre: \_\_\_\_\_ Municipio: Nombre: \_\_\_\_\_  
 N°: \_\_\_\_\_ N°: \_\_\_\_\_  
 Supervisor: Nombre: \_\_\_\_\_ Estado: Nombre: \_\_\_\_\_  
 N°: \_\_\_\_\_ N°: \_\_\_\_\_

Area N°: \_\_\_\_\_

Vivienda		Dirección				Personas			Nombre del Jefe
N°	Tipo	Calle, Camino, Senda, etc.	N°	Piso	Depto.	Tot.	Var.	Muj.	
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
Total									

Total viviendas particulares ocupadas:

Total viviendas particulares desocupadas:

Total viviendas colectivas:

donde

Vivienda N°

Vivienda tipo

Número secuencial de la vivienda particular ocupada comenzando de 1  
 P particular ocupada  
 D particular desocupada  
 C colectiva

/Dirección

Dirección, calle, etc.

escribir la calle, camino, terracería, senda, canal o referencia que permita guiar a dicha vivienda. Utilice más de una línea si es necesario

Dirección N°

escribir el número de la vivienda si ésta lo tiene escrito, caso contrario escribir sin número - SN -

Piso

escribir el piso en el cual está ubicada la vivienda - PB, 1, 2, etc.

Depto.

si es un Departamento escribir el número o letra del Departamento o escribir SN si no tiene número

Personas

escribir la cantidad de varones, mujeres y total por vivienda

Nombre del Jefe

escribir el nombre del jefe de la vivienda. Si hubiese varios hogares en dicha vivienda escribir el nombre de sólo uno de ellos.

21. Dada la importancia que reviste en las investigaciones conocer el costo de accesibilidad a las unidades muestrales, lo cual permite ajustar los costos de la investigación a niveles más óptimos de beneficio, será conveniente insertar al final de la HREC una serie de preguntas, preferentemente cerradas, para que conteste el empadronador, de tal manera que las mismas provean un factor de costo que permita una posterior estratificación del marco muestral en función de este factor de costo. En especial estas preguntas se harán para el área rural. Se sugieren, por ejemplo:

- i) Nombre del centro poblado más cercano al sector de empadronamiento;
- ii) Distancia aproximada en kilómetros entre este centro poblado más cercano y el sector de empadronamiento: .....km;
- iii) Tiempo aproximado en horas que se tarde en llegar de este centro poblado más cercano al sector de empadronamiento: .....horas;
- iv) El camino para llegar es:

pavimentado	<input type="checkbox"/>	senda	<input type="checkbox"/>	no hay camino	<input type="checkbox"/>
de tierra	<input type="checkbox"/>	un canal	<input type="checkbox"/>		
mejorado	<input type="checkbox"/>	.....			

/v) Se

v) Se puede llegar en:

- |                    |                          |           |                          |
|--------------------|--------------------------|-----------|--------------------------|
| locomoción pública | <input type="checkbox"/> | caminando | <input type="checkbox"/> |
| auto de alquiler   | <input type="checkbox"/> | canoa     | <input type="checkbox"/> |
| caballo, burro     | <input type="checkbox"/> | .....     |                          |

vi) En el centro poblado hay:

- |         |                          |                    |                          |
|---------|--------------------------|--------------------|--------------------------|
| hotel   | <input type="checkbox"/> | .....              |                          |
| pensión | <input type="checkbox"/> | no hay alojamiento | <input type="checkbox"/> |

vii) A este sector se puede llegar:

- todo el año
- imposible llegar por las lluvias en los meses de:  
.....

viii) A qué altura aproximada en metros se encuentra este sector: . . . .  
..... mts.

22. La sigla AGEB que se menciona en la HREC significa Area Geoestadística Básica. Estas AGEB son pequeñas áreas, en las cuales ha sido dividido el territorio nacional, para facilitar el proceso de las investigaciones estadísticas. Existen AGEB urbanas y AGEB rurales. En este último caso se ha conversado sobre la posibilidad de dividir las AGEB rurales, que tienen en promedio unas 10 000 hectáreas, en sub-AGEB. Esto reducirá los costos de actualización de estas áreas, dado que las AGEB serán utilizadas como unidades muestrales de selección, y actualizar 10 000 hectáreas, aunque sea en forma rápida, puede demandar un elevado gasto de recursos. De allí la conveniencia que DETENAL, que está realizando un meritorio trabajo en la confección del Marco Geoestadístico del Sistema Nacional de Información, en forma conjunta con la DGE, estudien la posibilidad de la creación de estas sub-AGEB.

23. Levantado el censo, las HREC se centralizarán para su crítica, codificación del factor de costo y posterior procesamiento mecánico. Por tal motivo se mantuvo una reunión con los funcionarios de computación, señores Saul Kletzel y Javier Games en la cual participaron además la Sra. Leticia Calzada de Vielle, el señor Jesús Romo y el suscrito. Esta información, alrededor de 1.4 (10)<sup>6</sup> HREC, se registrará en cinta. Este archivo constituirá el Marco Muestral, que posteriormente, una vez procesado el censo, será incrementado con indicadores estadísticos y económicos que se obtengan de la información censal y de las investigaciones muestrales que se lleven a cabo. Se deberá prever que toda información que se registre en el Marco Muestral deberá llevar la fecha del registro y toda información que se retire del Marco Muestral se deberá guardar en un archivo histórico. Asimismo, se estudiará la bondad de tener el Marco Muestral dividido en un conjunto de marcos muestrales por distintas áreas de estudio. Así, por ejemplo, cada Estado podrá tener su propio Marco Muestral, lo mismo que el Area Metropolitana de México y los grandes conglomerados como Monterrey y Guadalajara. Esto facilitará el trabajo independiente en los diseños muestrales que se realicen.

24. Las HREC servirán también para la confección de croquis en los casos en que no se tenga el croquis del HREC seleccionado en la muestra. Con este fin se realizó una experiencia en el área rural y en el área urbana del Area Metropolitana de México, donde se hicieron los croquis a mano alzada, en función de una supuesta HREC obtenida de la actual encuesta, y se registraron los tiempos de los distintos trabajos realizados.

Nueva Encuesta de Hogares basada en el censo de 1980

25. Para el diseño de la nueva encuesta de hogares se deberán considerar, principalmente, las siguientes etapas:

- i) Objetivos;
- ii) Conceptos y definiciones;
- iii) Variables;
- iv) Cuestionarios y formularios;
- v) Personal de campo (enumeradores, entrevistadores, supervisores, jefes de campo, etc.). Cantidad, jerarquías, normas, etc.;
- vi) Organigrama funcional;

/vii) Regiones,

- vii) Regiones, estratos y dominios de estudio para los cuales se quieren estimaciones;
- viii) Diseño muestral;
  - a) unidades muestrales;
  - b) tamaños muestrales y submuestrales;
  - c) métodos de selección;
  - d) estratificación;
  - e) replicaciones;
  - f) estimadores;
  - g) confiabilidad;
  - h) errores no muestrales;
  - i) control de calidad de la información;
  - j) control del movimiento del formulario;
  - k) efecto del diseño
- ix) Cartografía
  - x) Tabulados (distintas tabulaciones por períodos, áreas geográficas, estratos, etc.);
- xi) Análisis de la Información
  - a) Manual
    - entrevistados
    - supervisores
    - jefes de campo
    - oficina
  - b) Mecánica
    - consistencia
    - congruencia
    - coherencia
- xii) Programas de Computación
  - a) diseño lógico del procesamiento
  - b) listados
  - c) tabulados
  - d) coeficientes de expansión enteros (ver punto 53)
- xiii) Propaganda

xiv) Experiencia piloto

- a) diseño
- b) análisis
- c) replaneamiento

xv) Tiempos y costos

xvi) Difusión

26. Se estima, para el Area Metropolitana de México, por trimestre, consultar unas 2 400 viviendas para medir el desempleo. Este valor se obtuvo considerando una tasa de desempleo  $p = .10$  y aceptando un desvío absoluto  $d = .01$  con una confiabilidad del 95% o sea  $t=2$ . Se estimó la población económicamente activa - PEA - en un 30% de la población total, y se aceptó el promedio de 5 personas por vivienda particular. De esta manera se espera contar con una encuesta trimestral que mide a unas 12 000 personas de las cuales unas 3 600 serán PEA. Dado que la carga de trabajo del entrevistador se considera en 5 viviendas por día, se seleccionarán 480 sectores de empadronamiento - SE -.

27. Previamente a los SE, se seleccionarán AGEB como unidades primarias de muestreo. Dentro de las AGEB se seleccionarán los SE como unidades secundarias de muestreo y dentro de los SE se seleccionarán las viviendas como unidades terciarias de muestreo. La cantidad de AGEB a seleccionar aún no ha sido determinada.

28. El diseño adoptará el esquema del muestreo replicado. Se estima construir 4 replicaciones.

29. Considerando 480 SE por trimestre, alrededor de 8 SE por día, se necesitarán 8 entrevistadores, de acuerdo con la carga de trabajo establecida por entrevistador.

30. Para mejorar la muestra se han considerado, dentro del Area Metropolitana de México, 2 estratos geográficos. Uno, el área de crecimiento rápido de viviendas y el otro el área de crecimiento lento de viviendas. La asignación, por estrato hecha en forma proporcional al total de la población, resultó de 1/5 o sea de 96 SE en el estrato 1 de crecimiento rápido y de 4/5 o sea de 384 SE en el estrato 2 de crecimiento lento.

Asignación aleatoria de los SE, por replicación y según entrevistador en el estrato h = 1

31. Considerando 4 replicaciones cruzadas con 8 entrevistadores se tendrán 32 celdas replicación-entrevistador. Dado que se deben asignar al estrato h = 1 96 SE, cada celda contará con 3 SE o sea alrededor de 15 viviendas particulares.

32. Los SE se encuentran ordenados geográficamente y enumerados en forma secuencial: 1, 2, ....., 96. La asignación de estos 96 SE en las 32 celdas se hará en forma aleatoria. El método propuesto consiste en seleccionar una celda al azar y a esta celda asignarle el SE número 1. A partir de esta celda asignar los demás SE siguiendo los movimientos del caballo de ajedrez.<sup>2/</sup> Primero se deben agotar las 32 celdas con un SE en cada celda. En el caso en que el movimiento del caballo quede cerrado por estar la celda ya ocupada, se seleccionará una nueva celda desocupada al azar. Una vez asignados los 32 primeros SE, los 64 SE restantes se asignarán sumando las constantes 32 y 64 al SE de cada celda. Ejemplo:

Cuadro 1

Entrevistador	Replicación			
	1	2	3	4
1	2 - 34 - 66	31 - 63 - 95	17 - 49 - 81	4 - 36 - 68
2	10 - 42 - 74		1 - 33 - 65	32 - 64 - 96
⋮	⋮	⋮	⋮	⋮
8		6 - 38 - 70		25 - 57 - 89

33. Luego se deberán asignar los SE por semana y según entrevistador. Por razones prácticas se considerarán 12 semanas en el trimestre y 4 semanas en el mes. Se postula que los SE no deben guardar una interacción mes-semana-replicación. Es decir, que los SE de cada celda replicación-entrevistador no correspondan intencionalmente a la misma semana ni al mismo mes. Una

<sup>2/</sup> W.G. Cochran, "Sampling Techniques". John Wiley & Sons, Inc. 1960.



forma de realizar esta asignación puede ser la siguiente,

i) cada entrevistador tiene 12 SE asignados por replicación (3 SE por cada replicación) y dentro de la replicación los SE están ordenados en forma secuencial ascendente,

ii) se configura el siguiente cuadro 2, mes por semana, donde el mes =  $\overline{1;3}$  y la semana =  $\overline{1;4}$ ,

Cuadro 2

	Columna j											
	1	2	3	4	5	6	7	8	9	10	11	12
Mes	1	2	3	1	2	3	1	2	3	1	2	3
Semana	1	2	3	4	1	2	3	4	1	2	3	4

iii) se selecciona al azar un j, donde  $1 \leq j \leq 12$ ,

iv) se asigna esta columna j al primer SE del entrevistador 1,

v) esto significa que este primer SE debe ser entrevistado en el mes y semana que indica la columna j. Por ejemplo, si  $j = 6$ , este primer SE del primer entrevistador se entrevistará en la segunda semana del tercer mes, o sea, en la semana número 10 del trimestre, según se muestra en el siguiente cuadro,

vi) el segundo SE se entrevistará en la semana  $j + 1$  y así sucesivamente hasta llegar a  $j + k = 12$ . Luego se comenzará con  $j = 1$  hasta agotar al entrevistador 1,

vii) al entrevistador 2 se le asignará como número aleatorio de inicio el  $j + 1$  y se procederá de manera similar a lo descrito,

viii) al entrevistador 3 se le asignará el  $j + 2$  y así con el resto de los entrevistadores.

34. El siguiente cuadro 3 muestra los 12 SE que tiene cada entrevistador, según cuadro 1 anterior, y la semana del trimestre en que debe medir a cada uno de esos SE

Cuadro 3

Entrevistador	Semana del trimestre											
	1	2	3	4	5	6	7	8	9	10	11	12
1	49	63	34	36	68	81	95	66	31	2	4	17
2	1		10					42	74			
⋮												
8												
	$\hat{z}_{11}$	$\hat{z}_{12}$	$\hat{z}_{13}$		$\hat{z}_{21}$							$\hat{z}_{34}$
		$\hat{z}_{1.}$				$\hat{z}_{2.}$			$\hat{z}_{3.}$			

donde

$\hat{z}_{gl}$  estimador del g-mes l-semana

$\hat{z}_{g.}$  estimador del g-mes

35. Se observa en el cuadro 2, que dado que 3 (meses) no es múltiplo de 4 (semanas) se salva el riesgo de producir un ciclo j, lo cual podría introducir un sesgo sistemático debido a alguna asociación entre entrevistador, semana, mes y replicación.

Asignación aleatoria de los SE, por replicación y según entrevistador, en el estrato h = 2

36. Los SE se encuentran ordenados geográficamente y enumerados en forma secuencial: 1, 2, ..., 384. La asignación aleatoria siguiendo los movimientos del caballo de ajedrez se hará en forma similar al descrito para h = 1. Cada entrevistador tendrá 48 SE. Asimismo, la distribución de los SE por semana

/y mes

y mes dentro del trimestre, se hará de acuerdo con los procedimientos descritos para  $h = 1$ .

37. Cada entrevistador deberá, por tanto, medir a 60 SE por trimestre. Es decir, un promedio de un SE por día, o sea, alrededor de 5 viviendas diarias.

#### Rotación de las viviendas

38. Con el objeto de no cansar a las unidades medidas y además por ser conveniente mantener un equilibrio de eficiencia entre la estimación de un cambio ocurrido entre dos períodos y la media aritmética entre esos dos períodos, se hace necesario rotar parte de las unidades muestrales de un período a otro período. La pregunta se plantea en cuál es la proporción de unidades muestrales que se deben rotar (sustituir) de un período a otro para lograr una eficiencia máxima de la medición del cambio y de la medición de la media aritmética.

39. Para estimar el cambio del valor de cierta característica de un elemento de la muestra ocurrido entre dos períodos, la varianza del cambio será:

$$\text{Var}(x_{i1} - x_{i2}) = S_1^2 + S_2^2 - 2rS_1S_2$$

donde  $x_{i1}$  es el valor de la característica en el período 1,  $x_{i2}$  es el valor de la misma característica en el período 2,  $S_1^2$  y  $S_2^2$  son las varianzas poblacionales de la característica en los períodos 1 y 2 respectivamente y  $r$  es el coeficiente de correlación entre  $x_{i1}$  y  $x_{i2}$ .

Se observa que para un  $r > 0$ , para estimar el cambio es más conveniente conservar la misma muestra en los dos períodos que reemplazarla.

40. Para estimar la media aritmética

$$\bar{x} = \frac{1}{2} (x_{i1} + x_{i2})$$

la varianza de  $\bar{x}$  es:

$$\text{Var}(\bar{x}) = \frac{1}{4} (S_1^2 + S_2^2 + 2rS_1S_2)$$

/por tanto

por tanto para  $r = 0$ , es decir la unidad muestral no se conserva, es más conveniente para estimar la media aritmética obtener una nueva muestra para cada período.

41. Estimador compuesto. Llamando:

- $\bar{y}_1$             media de la muestra en el primer período
- $\bar{y}_2$             media de la muestra en el segundo período
- $\bar{y}_{1c}$            media muestral de la parte común de la muestra en el primer período
- $\bar{y}_{2c}$            media muestral de la parte común de la muestra en el segundo período
- $\bar{y}'_{2c}$            media muestral de la parte no común de la muestra en el segundo período

podemos establecer la siguiente ecuación de regresión

$$y_{i2c} = a + b y_{i1c} \quad (1)$$

donde  $y_{i2c}$  es el valor de la característica del  $i$ -elemento de la parte común en el segundo período e  $y_{i1c}$  es el valor de la misma característica del  $i$ -elemento de la parte común en el primer período.

Por tanto:

$$\bar{y}_{2c} = a + b \bar{y}_{1c} \quad (2)$$

y restando (2) de (1) obtenemos

$$y_{i2c} = \bar{y}_{2c} + b (y_{i1c} - \bar{y}_{1c}) \quad (3)$$

El estimador lineal de regresión de la media de la parte común en el segundo período cuando  $y_{i1c} = \bar{y}_{1c}$  será

$$\bar{y}'_{2c} = \bar{y}_{2c} = b (\bar{y}_1 - \bar{y}_{1c}) \quad (4)$$

donde  $b$  es el coeficiente de regresión de las unidades muestrales de la parte común en el primer y segundo período.

/Un estimador

Un estimador compuesto de la media poblacional  $\bar{Y}_2$  en el segundo período se puede establecer mediante una media aritmética ponderada de los estimadores  $\bar{y}_{2c'}$  e  $\bar{y}_{2c}$ , ponderándolos inversamente a sus varianzas. Llamando a este estimador  $\bar{y}_2'$ , será

$$\bar{y}_2' = \frac{p_2 \bar{y}_{2c'} + p_2' \bar{y}_{2c}}{p_2 + p_2'} \quad (5)$$

Imponiendo la condición  $p_2 + p_2' = 1$  es

$$\bar{y}_2' = p_2 \bar{y}_{2c'} + p_2' \bar{y}_{2c} \quad (6)$$

El tamaño total de la muestra en cada período es  $n = c + c'$ , donde  $c$  es el tamaño muestral de la parte común y  $c'$  es el tamaño muestral de la parte no común.

La varianza del estimador  $\bar{y}_{2c'}$ , de la parte no común es

$$\text{Var}(\bar{y}_{2c'}) = \frac{S^2}{c'} = \frac{1}{W_{2c'}} \quad (7)$$

y la varianza del estimador  $\bar{y}_{2c}$  de la parte común es

$$\text{Var}(\bar{y}_{2c}) = \frac{S^2 (1 - r^2)}{c} + r^2 \frac{S^2}{n} = \frac{1}{W_{2c}} \quad (8)$$

con lo cual el estimador (6) se convierte en

$$\bar{y}_2' = \frac{W_{2c'}}{W_{2c'} + W_{2c}} \bar{y}_{2c'} + \frac{W_{2c}}{W_{2c'} + W_{2c}} \bar{y}_{2c} \quad (9)$$

42. Varianza del estimador compuesto. La varianza de  $\bar{y}'_2$  es

$$\begin{aligned} \text{Var} (\bar{y}'_2) &= \frac{W_{2c'}^2}{(W_{2c'} + W_{2c})^2} \text{Var} (\bar{y}_{2c'}) + \frac{W_{2c}^2}{(W_{2c'} + W_{2c})^2} \text{Var} (\bar{y}'_{2c}) \\ &= \frac{W_{2c'}^2}{(W_{2c'} + W_{2c})^2} \cdot \frac{1}{W_{2c'}} + \frac{W_{2c}^2}{(W_{2c'} + W_{2c})^2} \cdot \frac{1}{W_{2c}} \\ &= \frac{1}{W_{2c'} + W_{2c}} \end{aligned} \quad (10)$$

Si en la (10) sustituimos por las (7) y (8) y operamos el álgebra, obtenemos

$$\text{Var} (\bar{y}'_2) = \frac{S^2 (n - c' r^2)}{n^2 - c'^2 r^2} \quad (11)$$

Si  $c' = 0$  no existe parte no común y el apareamiento de la parte común es completo. Si  $c' = n$  no existe parte común. En ambos casos la varianza de  $\bar{y}'_2$  dada por (11) toma la misma expresión. Por tanto para estimaciones actuales se obtiene la misma precisión manteniendo la misma muestra o reemplazándola totalmente en cada período.

43. Fracción óptima de la muestra en la parte común. El óptimo valor de  $c'$  se encuentra minimizando (11). Hallando

$$\frac{d}{dc'} \text{Var} (\bar{y}'_2) = 0, \text{ resolviendo la ecuación de segundo grado para } c'$$

y teniendo en cuenta que  $c' = n - c$ , se obtiene

$$\frac{c}{n} = \frac{\sqrt{1 - r^2}}{1 + \sqrt{1 - r^2}} \quad (12)$$

que da la fracción óptima de la muestra total que debe aparearse en la parte común.

44. Ventajas del reemplazamiento parcial de la muestra. Cuando el óptimo de  $c'$  se sustituye en (11) se obtiene la varianza mínima del estimador que viene dada por

$$\text{Var}(\bar{y}'_2) = \frac{S^2}{2n} (1 + \sqrt{1 - r^2}) \quad (13)$$

es decir, cuanto mayor sea  $r$  tanto mayor es la precisión del estimador. De aquí se deduce que cuando existe correlación entre los elementos muestrales de la parte común el reemplazamiento parcial de la muestra es mejor que conservar la misma muestra o sustituirla en cada período.

45. Ganancia en precisión. Para hallar la ganancia en precisión relativa de un estimador respecto a otro dividimos la inversa de la varianza del primero por la inversa de la varianza del segundo. Así la ganancia en precisión utilizando el estimador  $\bar{y}'_2$  con parte común, respecto a un estimador sin parte común cuya varianza está dada por  $\frac{S^2}{n}$ , es

$$\frac{2n}{S^2 (1 + \sqrt{1 - r^2})} : \frac{n}{S^2} = \frac{2}{1 + \sqrt{1 - r^2}}$$

Por ejemplo para  $r = .50$  la ganancia en precisión es 1.08, o sea un 8% de ganancia. Analizando la curva

$$g = \frac{2}{1 + \sqrt{1 - r^2}} - 1 \quad (14)$$

y multiplicando los valores de esta función por 100 se obtienen las ganancias en porcentajes.

Así para $r = 0$	$g = 0$	(0%)
$r = .50$	$g = .08$	(8%)
$r = 1$	$g = 1$	(100%)

La curva  $g$  es continua, creciente sin puntos de inflexión para  $0 \leq r \leq 1$ .

46. Ganancia en precisión cuando la fracción de muestreo en la parte común sea  $\frac{1}{4}$ . Es decir del primer período al segundo período se reemplazan 3 de 4 unidades muestrales. En este caso

$$\frac{c}{n} = \frac{1}{4} \text{ y } \frac{c'}{n} = \frac{3}{4}$$

La ganancia vendrá dada por la función

$$g = \frac{1}{4} \left( \frac{16 - 9r^2}{4 - 3r^2} \right) - 1 \quad (15)$$

$$\text{Para } r = 0 \quad g = 0$$

$$r = 1 \quad g = .75$$

47. Ganancia en precisión cuando la fracción de muestreo en la parte común sea  $\frac{3}{4}$ . En este caso  $\frac{c}{n} = \frac{3}{4}$  y  $\frac{c'}{n} = \frac{1}{4}$ .

La función de ganancia es

$$g = \frac{16 - r^2}{4(4 - r^2)} - 1$$

$$\text{Para } r = 0 \quad g = 0$$

$$r = 1 \quad g = .25$$

48. Por tanto, y por razones prácticas, se decidió que se rotará  $\frac{1}{4}$  de la muestra de período a período. Dado que cada período es un trimestre, se harán 3 rotaciones en el primer año de la investigación. En cada período se retirarán 120 SE (600 viviendas) y se incluirá un nuevo panel de 120 SE. De acuerdo con ello, la muestra tendrá un tamaño anual, para el Area Metropolitana de México, de 4 200 viviendas particulares, o sea, 840 SE.

49. No ha sido decidido aún, si finalizado el primer año de la investigación se seleccionará una nueva muestra o se continuará con la rotación muestral establecida.

50. Para el resto del país la muestra será seleccionada en forma análoga a la descrita. En principio, se han sugerido las siguientes divisiones del país para dar estimaciones,



Area

Es- ta- do	Urbana				Rural				Total Total	Total Total	
	G C 250(10) <sup>3</sup> y más	Localidades		Total	Zona producción						Total
		de 100(10) <sup>3</sup> hasta 250(10) <sup>3</sup>	resto, pero mayores 2 500		0	1	...	9			
1											
Total 1											
2											
Total 2											
.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	
32											
Total 32											
Total País											

donde

G C                                    grandes conglomerados

Zona de Producción            se ha considerado que podría haber un máximo de 9 zonas. Por ejemplo, zona agrícola, zona ganadera, zona petrolera, etc.

Estimador de total

51.    En el Area Metropolitana de México la muestra será seleccionada en 3 etapas, ver punto 27. A continuación se presenta un cuadro resumen conteniendo, entre otros, el método de selección por etapa y la unidad muestral que se utilizará en cada etapa.

/Area Metropolitana

Area Metropolitana de México. Estrato h

Etapa t	Unidad Muestral	Unidades en Universo	Unidades en Muestra	Método de selección
1	UP	$G_h$	$g_h$	pptcv(s)
2	US	$M_{hi}$	$m_{hi}$	pptav(s)
3	UT	$V_{hij}$	$v_{hij}$	psa(s) o igual

/donde

Probabilidad  
de selección  
de una unidad

Fracción muestral  
o probabilidad que  
la unidad sea selec-  
cionada en la muestra

Subíndices  
de la  
muestra

$$f_t$$

$$P_{hi} = \frac{v_{hi}^o}{v_h^o}$$

$$f_1 = g_h P_{hi}$$

$$h = \overline{1; L}$$

$$i = \overline{1; g_h}$$

$$P_{hij} = \frac{v_{hij}^i}{v_{hi}^i}$$

$$f_2 = m_{hi} P_{hij}$$

$$j = \overline{1; m_{hi}}$$

$$P_{hijk} = \frac{1}{v_{hij}}$$

$$f_3 = v_{hij} P_{hijk}$$

$$k = \overline{1; v_{hij}}$$

donde

UP	unidad primaria de muestreo (AGEB)
US	unidad secundaria de muestreo (SE)
UT	unidad terciaria de muestreo (viviendas)
$G_h$	total poblacional de UP en el h-estrato
$M_{hi}$	total poblacional de US en el hi-UP
$V_{hij}$	total poblacional de UT en el hij-US según registro detallado
$V'_{hij}$	total poblacional de UT en el hij-US según recuento rápido
$V'_{hi} = \sum_j V'_{hij}$	total poblacional de UT en el hi-UP según recuento rápido
$V^o_{hi}$	total poblacional de UT en el hi-UP según censo
$V^o_h = \sum_i V^o_{hi}$	total poblacional de UT en el h-estrato según censo
$g_h$	tamaño muestral de UP en el h-estrato
$m_{hi}$	tamaño muestral de US en el hi-UP
$v_{hij}$	tamaño muestral de UT en el hij-US
$P_{hi}$	probabilidad de seleccionar la hi-UP
$P_{hij}$	probabilidad de seleccionar la hij-US
$P_{hijk}$	probabilidad de seleccionar la hijk-UT
pptcv(s)	probabilidad proporcional al tamaño censal de viviendas en forma sistemática
pptav(s)	probabilidad proporcional al tamaño aproximado de viviendas en forma sistemática
psa(s)	probabilidad simple al azar en forma sistemática

/y considerando

y considerando

$y_{hijk}$  valor de la característica en la hijk-UT

$\frac{v'_{hi}}{v^o_{hi}}$  tasa de crecimiento de la hi-UP

$\frac{v_{hij}}{v'_{hij}}$  tasa de corrección de la hij-US

el estimador del total para el h-estrato es

$$\hat{Y}_h = \frac{v^o_h}{g_h} \sum_i \frac{g_h}{i} \frac{v'_{hi}}{v^o_{hi}} \frac{1}{m_{hi}} \sum_j \frac{m_{hi}}{j} \frac{v_{hij}}{v'_{hij}} \frac{1}{v_{hij}} \sum_k \frac{v_{hij}}{k} y_{hijk}$$

52. Dado que se utilizarán 4 replicaciones, donde cada replicación es una muestra representativa en igual grado de la población, se tendrán 4 estimadores por estrato, uno por cada replicación. Estos 4 estimadores permiten, en forma sencilla, estimar la varianza de  $\hat{Y}_h$  utilizando el criterio dado en punto 15.

53. Expansión de los datos. Por razones prácticas en el proceso de computación se puede utilizar un artificio algebraico para la expansión de los datos muestrales. El mismo consiste en utilizar solamente la parte entera del coeficiente de expansión evitando así el uso de decimales. Una parte de las unidades muestrales son expandidas por defecto y la otra por exceso, en ambos casos el error del coeficiente empleado es menor a un entero. La suma de los totales así estimados de ambas partes de la muestra es igual a la suma del total estimado que se obtiene si se utiliza el coeficiente de expansión sin modificar. El procedimiento es el siguiente. Haciendo

$$\frac{N}{n} =$$

$$\frac{N}{n} = d + \frac{r}{n}$$

donde N es la población de unidades, n el tamaño muestral, d la parte entera del cociente y r el resto, luego

$$n \left( \frac{N}{n} \right) = nd + r$$

y dividiendo la muestra en dos partes de tal manera que

$$n = r + (n - r)$$

queda

$$N = r(d + 1) + (n - r)d$$

Dado que las unidades muestrales entran en el proceso en forma aleatoria, las primeras r unidades son expandidas por (d + 1) y las restantes (n - r) unidades son expandidas por d.

#### Resumen

54. i) a la encuesta de hogares, basada en el censo de 1970, se le aplicará un modelo estadístico que permitirá analizar las distintas fuentes de variación;

ii) la Hoja Resumen del Empadronador Censal del censo de 1980 servirá como base para la construcción de un marco muestral;

iii) la nueva encuesta de hogares basada en el censo de 1980 utilizará un modelo basado en el esquema de muestras replicadas.





