

2  
2 E



# Universidad Nacional Autónoma de México

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES

“ACATLAN”

## “DESCRIPCION DE LA MUESTRA RURAL DEL EMPADRONAMIENTO DE LOS CENSOS ECONOMICOS 1986”

**T E S I S**

QUE PARA OBTENER EL TITULO DE:  
**A C T U A R I O**  
P R E S E N T A :

**Luis de Jesús Cosaretti Castillo**





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

	página
INTRODUCCION	1
CAPITULO I	
DELIMITACION DE LA POBLACION EN ESTUDIO	14
CAPITULO II	
OBTENCION DEL TAMANO DE LA MUESTRA	28
CAPITULO III	
SELECCION DE LA MUESTRA BAJO EL ESQUEMA DE PROBABLIDAD PROPORCIONAL AL TAMANO (PPT) Y- REDUCCION DE LA MISMA	39
CAPITULO IV	
FACTORES QUE AFECTARON EL TAMANO DE LA MUESTRA	58
CAPITULO V	
INFERENCIA DE LOS RESULTADOS	77
CONCLUSIONES	107
APENDICE	118
BIBLIOGRAFIA	122

## INTRODUCCION

El Gobierno de la República Mexicana, a través del Instituto Nacional de Estadística Geografía e Informática (INEGI), órgano desconcentrado de la Secretaría de Programación y Presupuesto, es el encargado de llevar a cabo los trabajos relacionados con los Censos Económicos 1986.

Este evento es uno de los mecanismos más importantes para la obtención de información económica a nivel nacional, pues coadyuvará en la toma de decisiones de políticas económicas para el país.

El presente documento persigue desarrollar los aspectos técnicos involucrados a lo largo de la etapa del levantamiento rural de los Censos Económicos 1986.

Sin embargo es menester para lograr la ambientación con el presente trabajo, desarrollar un panorama general sobre el INEGI y lo relacionado con la actividad censal del país, asimismo describir el lenguaje técnico que se manejará a lo largo de todo el documento con el objeto de definir el marco sobre el cual se trabajará; de esta forma este apartado se divide en dos partes: Actividad Censal en el País,

y División del Territorio para el levantamiento Censal

#### ACTIVIDAD CENSAL EN EL PAIS.

El Instituto Nacional de Estadística, Geografía e Informática fue fundado el 26 de enero de 1983, como un órgano desconcentrado de la Secretaría de Programación y Presupuesto; su creación responde a un proceso de reestructuración administrativa actual, que implementa el Gobierno de la República.

El INEGI está formado por cuatro unidades: Dirección General de Estadística, Dirección General de Política Informática, Dirección General de Análisis de la Información, además de una Coordinación Ejecutiva.

El INEGI tiene como uno de sus objetivos desarrollar las bases para la producción de información, así como también la tarea de elaborar estadísticas de los hechos demográficos, sociales y económicos que permitan definir y establecer políticas adecuadas al desarrollo del país.

Por otra parte, en el contexto de la política de desconcentración se crearon diez Direcciones Regionales del INEGI. Al delegar en éstas las responsabilidades de

de la planeación a detalle de los levantamientos, su ejecución y la captura y procesamiento de la información, los Censos Económicos se significarán en un elemento fundamental para la consolidación de las mismas.

Para poder obtener la información estadística necesaria y conocer la realidad del país, se cuenta con un sistema de información constituido por diversas técnicas de investigación como: registros administrativos, encuestas por muestreo y censos; como se menciona anteriormente, en este caso nos ocuparemos únicamente de los censos.

Los Censos son precisamente una técnica de investigación consistente en identificar, ubicar y enumerar de manera exhaustiva las unidades objeto de estudio, a fin de conocer sus características principales, en un momento dado.

La importancia que tienen los censos es plenamente reconocida por los diferentes países del mundo así como por diversos organismos internacionales, ya que los datos captados de los Censos permiten, entre otras cosas, contar con información estadística que de acuerdo a la Población en estudio tenga una cobertura total y,

por tanto, completa representatividad, que servirá como base para la formulación de señalamientos de tipo gubernamental, en materia de desarrollo social, económico y político.

En México se realizan tres tipos de Censos a nivel nacional.

**El Censo de Población y Vivienda.** A través de éste se realiza un recuento de las viviendas, familias y personas existentes, así como de sus principales características

**El Censo Agrícola, Ganadero y Ejidal.** A través de éste se efectúa un recuento sobre el uso del suelo en las áreas destinadas a las actividades agropecuarias y forestales, así como de las principales características de las unidades de producción.

**Los Censos Económicos.** a Tráves de éstos se hace un recuento de todos los establecimientos productores de bienes y/o servicios, sean públicos o privados, captando información de sus principales características; se excluyen aquí las unidades de producción agropecuaria y forestal ya que su recuento lo lleva a cabo el Censo Agrícola.

Los principales objetivos de estos Censos son:

Captar, procesar y divulgar información básica sobre las unidades económicas productoras de bienes y servicios, que no sean agropecuarias ni forestales referidas al año de 1985 y a la totalidad del territorio nacional con la finalidad de brindar información actual y clasificada para:

- i) -Estudios del sector Público en la planeación económica y social. En particular, servirá de punto de partida para la elaboración de un sistema nacional de identificación de empresas y establecimientos, proporcionando los directorios correspondientes para conformar marcos de referencia tanto para sistemas de encuestas continuas, como para estudios de índole particular.
- ii) -El Servicio Público de Información.
- iii) -Apoyo a la Investigación Científica.
- iv) -Uso de Organismos Internacionales.

El Censo tendrá cobertura nacional y los establecimientos objeto de del Censo son aquellos que no

se encuentren comprendidos dentro de las actividades agrícolas y forestales que, como ya se dijo, son captados por otro tipo de Censo.

La Clasificación de Actividades y Productos (CMAP) registra todas las actividades económicas que se realizan en nuestro País. De esta clasificación únicamente las unidades que están comprendidas en los sectores económicos que a continuación se detallan serán los establecimientos objetos del Censo Económico.

#### INDUSTRIA

Sectores 2, 3 y 4 CMAP

#### COMERCIO

Sector 6 CMAP

#### SERVICIOS

Sectores 8 y 9 CMAP

TRANSPORTE  
Sector 7 CMAP

CONSTRUCCION  
Sector 5 CMAP

PESCA  
Sector 1 CMAP

De acuerdo a lo anterior se ha considerado conveniente captar la información de estos Censos Económicos en dos etapas:

1.- EMPADRONAMIENTO.- Este se realizará con base en dos necesidades: la primera, la creación del sistema de identificación de empresas y establecimientos, y segundo, la decisión de utilizar cuestionarios diferenciados por actividad económica durante la etapa del Levantamiento Censal, lo cual hace necesaria la existencia de un directorio que apoye la asignación de dichos cuestionarios.

2.- LEVANTAMIENTO CENSAL.- Este se realizará contando ya con los cuestionarios diferenciados, los que permitirán captar con mayor especificidad la información por sector económico.

La cobertura será nacional y se han definido tres modalidades para su ejecución:

**MODALIDAD BARRIDO.-** Se cubrirán totalmente las localidades que en 1980 presentaron una población de 2,500 o más habitantes, así como todas las cabeceras municipales que no hayan alcanzado esa población.

**MODALIDAD DIRECTORIOS.-** Se cubrirán por medio de directorios todos los establecimientos con actividad económica relevante que se encuentran fuera de las poblaciones mencionadas y que serán principalmente parques industriales, minas, agroindustrias etc.

## MODALIDAD RURAL.-

Para el levantamiento de las áreas rurales se ha planteado la necesidad de lograr los objetivos de ambas etapas en un solo momento denominado Empadronamiento Rural.

Esta fase es tan importante como la anterior debido a que:

En primer lugar a través de los Censos se pretende conocer la actividad económica de todo el país, y este conocimiento solo se alcanza si se conoce la totalidad: tanto de las áreas de mayor actividad como las de menor.

En segundo lugar dado que las localidades de las zonas rurales se encuentran dispersas: el Empadronamiento Rural se realizará bajo la modalidad de muestreo, lo cual permitirá inferir las características de todo el país.

Los objetivos del Empadronamiento Rural son:

- Aportar elementos que permitan la conformación de un directorio de establecimientos.
- Clasificar a los establecimientos por actividad

económica.

- Actualizar los materiales cartográficos utilizados en el transcurso del Empadronamiento Rural y/o elaborar el material cartográfico de aquellas localidades donde no exista este tipo de material.

Unicamente son objeto de Empadronamiento Rural, los establecimientos que están comprendidos dentro de los sectores que en seguida se detallan de acuerdo a la CMAP.

Industria.- De este sector económico lo que será posible encontrar en localidades rurales serán pequeñas industrias manufactureras, transformadoras de materia prima, así como de productos agrícolas.

Comercio.- De este sector económico se hallarán en zonas rurales actividades tales como:

- Tiendas Rurales y Bodegas CONASUPO.
- El productor que comercializa en un establecimiento su materia prima directamente al público.
- Un comerciante que compra productos elaborados y los distribuye al público.

También serán considerados aquellos

establecimientos cuya actividad económica quede comprendida dentro de los sectores: Servicios, Transportes, Construcción y Pesca.

#### DIVISION DEL TERRITORIO NACIONAL PARA EL LEVANTAMIENTO CENSAL.

El contar con un instrumento para manejar la información estadística dentro de un contexto geográfico y temporal, así como para facilitar las operaciones censales, ha creado la necesidad de establecer un marco de referencia formado por unidades cartográficas definidas a las que se les ha denominado Areas Geoestadísticas Básicas (AGEB).

Dado que para efectos de tratamiento estadístico se requiere homogeneidad desde el momento del levantamiento de la información, hace necesario que las áreas Geoestadísticas sea homogéneas en su constitución; además sus límites son permanentes y fácilmente reconocibles en el terreno a fin de facilitar las operaciones de levantamiento de Censos y Encuestas.

En virtud de que la información es manejada por entidades políticas el marco geográfico para fines estadísticos respeta en principio la división política, por lo cual, dicho marco, es una subdivisión de la

división municipal del país.

Cada AGEB es una división territorial que por su dimensión y características de asentamientos humanos, permite considerarse como unidad en el transcurso de las actividades del levantamiento censal.

Existen dos tipos de AGEB: AGEB Rural y AGEB Urbana.

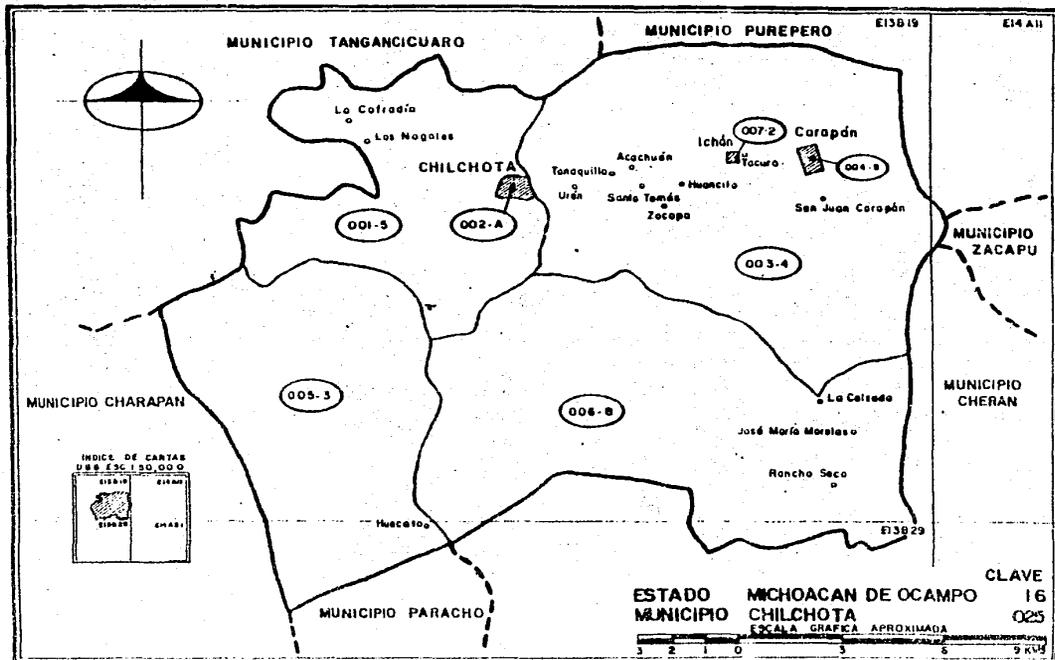
Las AGEB Urbanas son las áreas ocupadas por un conjunto de manzanas (entre 20 y 30) perfectamente delimitadas, cuyo total varía en función del número de habitantes que contiene. Cuando una localidad alcanza 2500 habitantes ya es susceptible de asignarle una AGEB Urbana; si se trata de una localidad que rebasa considerablemente dicha cantidad de habitantes, entonces contendrá tantas AGEB Urbanas como sea necesario.

Las AGEB Rurales tienen en su mayoría, límites coincidentes con la estructura político-administrativa; de fácil identificación en el terreno, a través de brechas, ríos, líneas de transmisión, vías de ferrocarril, etc.

La extensión promedio de estas áreas es de 10,000 hectáreas, variando de acuerdo a las características geográficas y topográficas del terreno. Por lo general de estas AGEB Rurales se excluyen las localidades que presentan una traza urbana bien definida.

El cuadro de la siguiente hoja muestra un "Croquis Municipal" con la división del marco geoestadístico, este grafo como su nombre lo indica, se trata de un croquis de la periferia municipal y dividido en Áreas Geoestadísticas Básicas.

# CROQUIS MUNICIPAL CON LA DIVISION EN AREAS GEOESTADISTICAS BASICAS



DEFINITIVO / 86.

1:86.01.02.012.10.004.05

## CAPITULO I

### DELIMITACION DE LA POBLACION EN ESTUDIO

Dentro del Programa "Censos Económicos 1986", en la fase rural se pretende la captación de la información sobre la actividad económica no agropecuaria de las zonas rurales en el País; para tal efecto es necesario establecer una diferencia entre lo que se define por zona rural y urbana, a fin de recolectar la información únicamente de la primera.

Una zona se podrá calificar como rural cuando presente baja densidad demográfica, escasa diferenciación social, baja movilidad espacial y social y predominio de actividades agropecuarias, todas estas características se asocian al indicador básico que es la densidad de población relativamente baja, de tal suerte que para efectos censales la división de lo urbano y rural se funda en un criterio cuantitativo, de donde se considera urbano la población que habita en localidades con una población mayor o igual a 2,500 habitantes, y como rural la que habita en localidades con población menor a la cifra indicada.

Los conceptos anteriores tienen una larga

tradición censal de ahí que numerosas características de la población se encuentren tabuladas según esta clasificación. Cabe decir que el criterio generalmente usado para fines censales, en clasificar a la población desde ese punto de vista no considera las características de urbanismo, categoría política (excepto que se trate de cabecera municipal), o la administración local, ni la proporción de población dependientes de actividades agrícolas o no agrícolas etc.

Por otra parte, refiriéndose a la actividad económica en el país, se calcula que ésta se encuentra aproximadamente en un 95% dentro del sector urbano, restando solo un 5% en el área rural; asimismo, se estima, que el 98% de las localidades caen dentro del sector rural (123,200 localidades aproximadamente), mientras que únicamente un 2% (2,100 localidades aproximadamente), comprenden el sector urbano; lo anterior traducido a captación de la información, significa un costo igual para ambos sectores, esto es, el costo para captar un 95% de la actividad económica del país dentro el sector urbano es el mismo que se necesitaría para la captación del 5% en zonas rurales.

Por otra parte, la complicación excesiva en la

estrategia de captación y organización debida al tiempo y esfuerzo que se requirió al barrer aproximadamente 2,100 localidades contra lo que se invertiría en 123,200, y si se considera que la gran mayoría de los casos se trata de pequeñas localidades excesivamente dispersas de las cuales en muchas de éstas, la actividad económica no agropecuaria es completamente nula, obliga a plantearse la alternativa de obtener la información mediante un muestreo en las áreas rurales a fin de reducir costos y asegurar un levantamiento ordenado, eficaz y completamente independiente de los resultados de la fase urbana.

El planteamiento de una muestra implica definir la unidad estratégica u operativa, la cual en este caso quedó considerada como un AGEB, pues se trata de la unidad más adecuada para el manejo de información estadística dentro de un contexto geográfico. Enseguida surge la necesidad de establecer la variable o característica de interés en el presente estudio para fines de determinación del tamaño de muestra, para lo cual, se escogió al personal ocupado dentro de AGEB.

Por otra parte el esquema de muestreo seleccionado fue con probabilidad proporcional al tamaño para la variable antes mencionada, sin embargo, por

desconocimiento de la cantidad de personal ocupado por AGEB se empleó como medida de peso a la Población Económicamente Activa (PEA) no Agropecuaria dentro del AGEB, por tratarse de una variable altamente correlacionada con personal ocupado debido a que esta última puede considerarse como un subconjunto de la PEA-No-Agropecuaria; de este modo se pretende que la muestra se dirija a aquellas AGEB con un mayor grado de actividad económica, y posteriormente ejecutar las ponderaciones convenientes a fin de lograr una correcta inferencia de los datos.

Una vez que la unidad operativa junto con la característica de peso en estudio quedaron plenamente definidas se procedió a delimitar la Población en estudio, el cual se obtuvo para cada una de las Entidades Federativas que conforman el País. Esto se logró de la siguiente manera:

La Población por Entidad Federativa quedó integrado por todas las AGEB rurales y urbanas que contenían a localidades con población menor a 2,500 habitantes según el Censo de Población de 1980, que no fueran Cabeceras Municipales y además que no estuvieran contempladas en la fase urbana; de tales AGEB se obtuvo su PEA-No-Agropecuaria, siendo necesario en varios casos hacer

una serie de ajustes en los PEA-No-Agropecuaria para ciertas AGEB en las cuales se suscitaron los siguientes casos particulares:

i) AGEB rurales cuyos limites colindan con localidades con poblacion mayor o igual a 2,500 habitantes o bien con Cabeceras Municipales, que hayan rebasado sus limites introduciendose dentro del AGEB rural (ver esquema # 1).

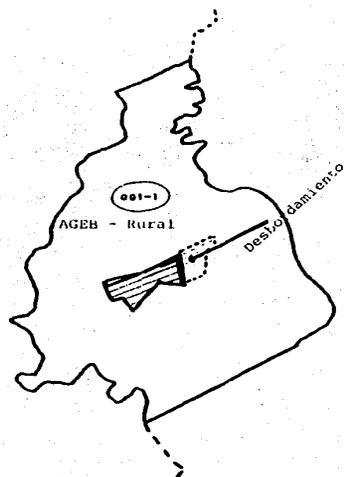
ii) AGEB Rurales que contenian localidades que, aun sin contar con AGEB Urbanas fueron barridas durante el Operativo urbano ( como ejemplo considerense Cabeceras Municipales con poblacion inferior a los 2,500 habitantes. Ver esquema # 2).

Tales ajustes consistieron en excluir del AGEB rural la parte de la PEA-No-Agropecuaria que correspondia a aquellas localidades o AGEB que fueron cubiertas durante la etapa urbana.

Los calculos se efectuaron en forma proporcional por no contar con informacion reportada a nivel localidad. La siguiente seccion se ocupa en describir el material requerido y enseguida explica para cada caso la metodologia empleada.

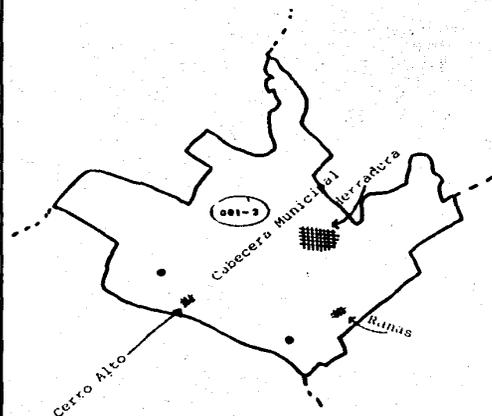
ESQUEMA 1

Localidad Urbana que rebasó los límites, invadiendo la AGEB Rural.



ESQUEMA 2

AGEB Rural que contiene una localidad barrida durante el Operativo Urbano.



## MATERIAL REQUERIDO

-Catálogo de Integración Territorial (CIT) de Población por entidad federativa, municipio y localidad según características de la población de acuerdo a su actividad económica por sectores:

- i) Sector primario.- Correspondiente a la actividad económica Agropecuaria.
- ii) Sector secundario.- Correspondiente a la actividad económica Industrial.
- iii) Sector terciario.- Correspondiente a Servicios.

-Catálogo de cifras ajustadas (CCA), por entidad federativa, el cual provee información desde el nivel manzana hasta el nivel estado sobre el total de población.

-Catálogo de AGEB por municipio y localidad, el cual provee información de AGEB urbanas y rurales que integran un municipio, localidades rurales por AGEB y localidades con AGEB Urbanas.

-Croquis municipales con el Marco Geoestadístico.

-Listados de Cobertura Geográfica por Entidad federativa de los Censos Económicos 1986.

**PROCEDIMIENTO:**

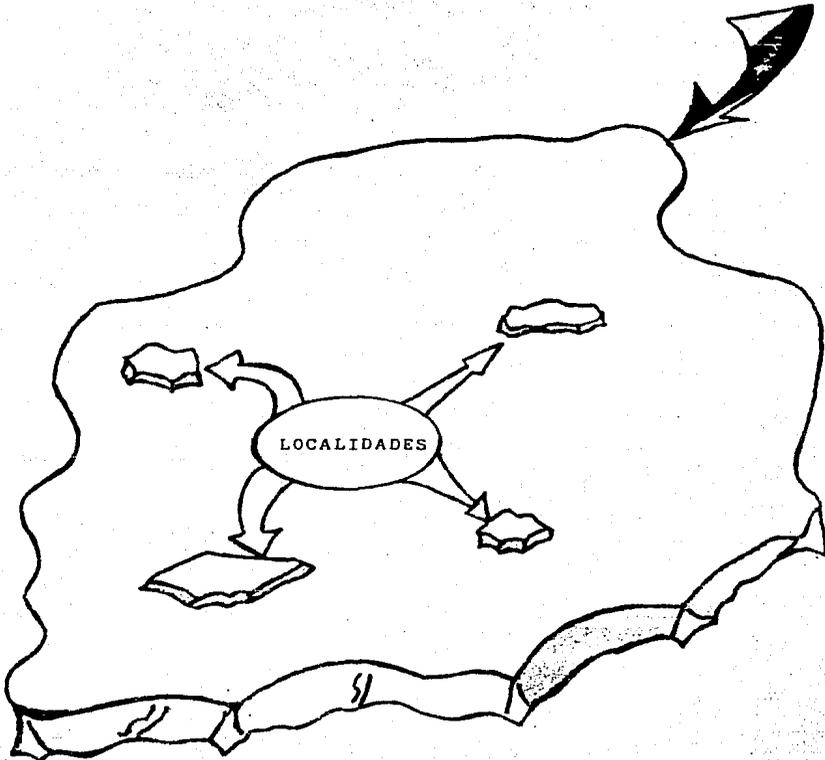
1.- Se cancelaron del Catálogo de Integración Territorial las AGEB urbanas así como sus sectores secundario y terciario correspondientes a las localidades con población mayor o igual a 2,500 habitantes y Cabeceras Municipales contempladas en los listados de cobertura geográfica por entidad federativa.

De esta forma, las AGEB restantes son aquellas que no fueron barridas durante el Operativo Urbano, es decir, se trata de AGEB rurales y urbanas referidas a localidades con población menor a los 2,500 habitantes y que no son cabeceras municipales, ya sea que las contengan como en el caso de las AGEB Rurales (esquema # 3), o estén contenidas las AGEB Urbanas dentro de la localidad. (esquema # 4).

2.- En caso de que una localidad considerada en el barrido urbano no tenga AGEB urbana propia y esté contenida en un AGEB Rural, el tratamiento que se sigue para realizar el ajuste es el siguiente:

ESQUEMA 3

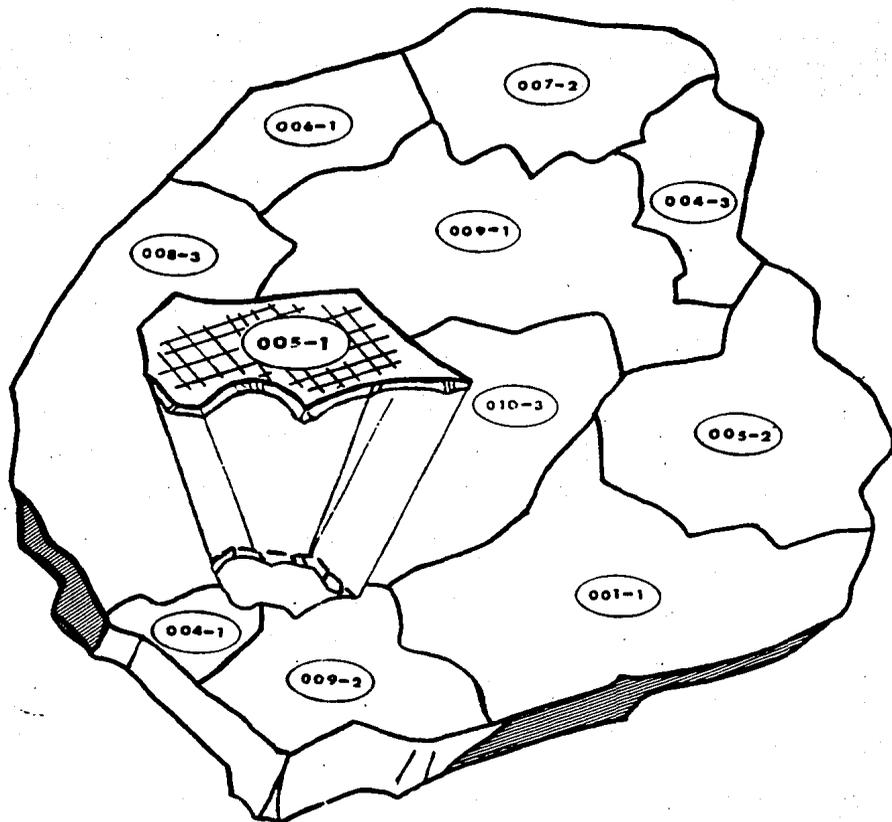
AGEB RUARAL



Gráfica de un AGEB Rural, las localidades están contenidas dentro del AGEB.

ESQUEMA 4

Las AGEB Urbanas estan contenidas dentro de la localidad.



Se restara del AGEB el total de la poblacion de la localidad que se reporte en el Catálogo de Cifras Ajustadas, el resultado es la poblacion rural en el AGEB, la cual se dividira entre la poblacion total a fin de obtener el porcentaje de poblacion rural en el AGEB, finalmente se multiplicara la PEA-No-Agropecuaria por el porcentaje de poblacion rural (factor de ajuste), con lo cual se obtiene la proporcion de PEA-No-Agropecuaria rural estimada dentro del AGEB. De esta forma el ajuste queda concluido.

El siguiente ejemplo facilitara la comprension de lo anterior:

Suponga que al revisar los Croquis Municipales en cierta entidad federativa, en el Municipio 023 "Lagunas", cuya cabecera municipal es 0001 "Laguna" esta se encuentra ubicada dentro del AGEB Rural 002-1 segun grafica # 5. Se observa que es necesario ajustar la poblacion en el AGEB 001-2 pues debe excluirse del AGEB la poblacion de la Cabecera Municipal ya que carece de AGEB propia; entonces como primer paso se busca el municipio 023 en el CCA.

Suponga que el CCA reporta:

AGEB:

CLAVE:

001-2

0001 Laguna

001-2

0002 San Raúl

001-2

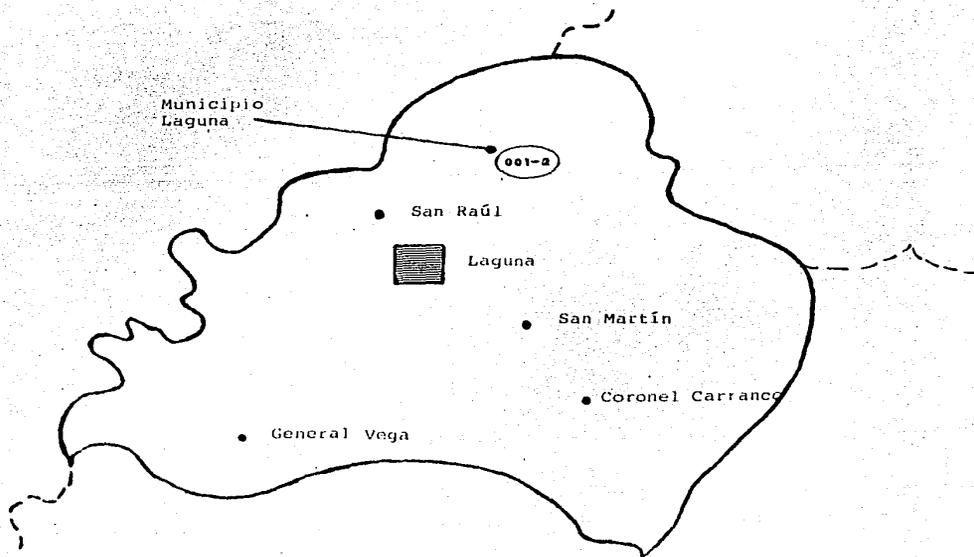
0003 San Martín

001-2

0004 Cor. Carranco

001-2

0005 General Vega



ESQUEMA 5

MUNICIPIO	LOCALIDAD	AGEB	POBLACION
023	0001	001-2	526
023	0002	001-2	361
023	0003	001-2	401
023	0004	001-2	280
023	0005	001-2	100

La población de la cabecera municipal 0001 Laguna es 526 y la población total del AGEB es:

$$526 + 361 + 401 + 280 + 100 = 1668 \text{ Hab.}$$

Entonces la población rural en el AGEB es:

$$1668 - 526 = 1142 \text{ Hab.}$$

Y el porcentaje de la población rural en el AGEB es:

$$\frac{\text{Población rural} = 1142}{\text{Población total} = 1668} = 0.6847$$

Que es el factor de ajuste que multiplicará a la PEA-No-Agropecuaria, esta última se obtiene del CIT sumando la PEA sector secundario y la PEA sector terciario y si en este caso los datos del CIT son los

se presentan a continuación:

AGEB	LAGUNA	PEA - SECTORES		
		POBLACION	PRIMARIO	SECUNDARIO
001-2	1668	211	600	124

La PEA-No-Agropecuaria es:  $600 + 124 = 724$ , entonces la parte proporcional de la PEA-No-Agropecuaria ( esto es, la PEA-No-Agropecuaria excluyendo a la cabecera municipal) se obtiene multiplicando el factor de ajuste ( 0.6847 en este caso) por la PEA-No-Agropecuaria que se obtuvo del CIT, obteniendose:

$$\begin{aligned} &(\text{factor de ajuste}) * (\text{PEA-No-Agro.}) = \text{PEA-No-Agro.} \\ & \hspace{15em} \text{Ajustada} \\ \Rightarrow & (0.6847) * (724) = 495.72 \approx 496. \end{aligned}$$

En caso se existir desbordamiento de localidades urbanas sobre las AGEB rurales el mecanismo para realizar un ajuste a su población se describe enseguida:

Con ayuda de Croquis Municipales y del Catálogo de AGEB por Municipio y Localidad, se procede a identificar las AGEB que colindan con las localidades urbanas ya barridas o bien en Cabeceras Municipales, en general se trata de AGEB envolventes o adyacentes. El objetivo que

se persigue es detectar los posibles crecimientos de localidades urbanas sobre AGEB Rurales para descontar proporcionalmente la población reportada en esta última. Los pasos a seguir son los siguientes:

i) Obtener del CIT la suma de la población correspondiente a la o a las AGEB Urbanas que conforman la localidad; dicha suma se coteja con el total de población reportado en los Listados de Cobertura por localidad; en caso de que ambos totales no coincidan se tratará de un desbordamiento de la población urbana sobre la o las AGEB Rurales por lo cual será necesario hacer un ajuste en la población de dichas AGEB ya que de no ajustarse se incurriría en el error de asignarles una población económicamente activa no agropecuaria mayor, dándole un peso que no le corresponde.

ii) Para realizar el ajuste se recurre al CCA a fin de averiguar cuales son las AGEB que reportaron población para cada localidad contemplada en el barrido urbano y cual es su magnitud, a fin de que esta cantidad se reste del AGEB Rural, de este modo se obtiene la población rural en el AGEB. El procedimiento que se sigue es el mismo que el del caso 2, esto es, se obtiene el factor de ajuste dividiendo población rural entre población total y esto se multiplica por la PEA-No-Agropecuaria

que se obtenga del CIT.

El resultado de dicho producto es la PEA-No-agropecuaria ajustada.

A continuación se ilustra este caso mediante un ejemplo:

Suponga que el AGEB Urbano 001-4 correspondiente a la localidad (0098) "San Miguel", esta contemplada en el listado de cobertura y se halla envuelto por el AGEB Rural 002-2. Al cotejar los totales de población del Listado de Cobertura (5840) contra el del CIT (5000 habitantes), estos no coinciden, por lo cual es necesario ajustar la PEA del AGEB en cuestión.

Suponga que el CIT reporta las siguientes cantidades:

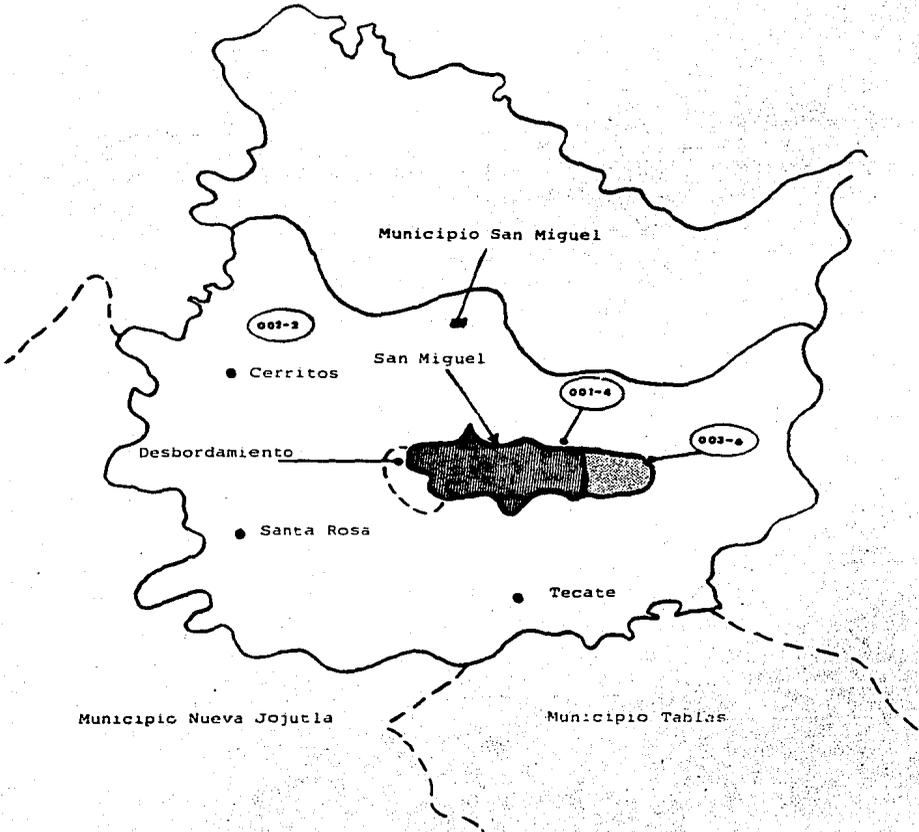
AGEB	POBLACION	PEA - SECTORES		
		PRIMARIO	SECUNDARIO	TERCIARIO
001-4	5000	-----	60	30
002-2	3200	-----	200	100

Del croquis municipal se extrae el esquema #5 donde se observa la situación.

Al no coincidir los resultados del CIT con el listado de cobertura implica que la mancha urbana

ESQUEMA 6

Gráfica del desbordamiento de la localidad urbana San Miguel sobre el AGEB Rural 002-2.



invadió el AGEB Rural.

Consultando el CCA se verifica si efectivamente la AGEB 001-4 es la única que comprende la localidad "San Miguel", o si reporta población en más de un AGEB y se le acumula a dicha localidad. Revisando el CCA encontramos:

LOCALIDAD	AGEB	POBLACION
0098	001-4	5000
0098	002-2	840

De donde la población total para la localidad San Miguel es  $5000 + 840 = 5840$  habitantes.

De aquí se deduce que el CCA reporta 840 habitantes que viven en la localidad "San Miguel" dentro del AGEB Rural envolvente, por lo cual se debe restar esa cantidad, de donde la población rural para el AGEB 002-2 es  $3200 - 840 = 2360$  habitantes.

Recuérdese que para obtener la PEA-No-Agropecuaria, se suma la PEA Secundaria y la PEA terciaria la cual en este ejemplo es:  $200 + 100 = 300$ .

Enseguida se obtiene la Proporción de Población

rural con respecto a la población total con el factor de ajuste  $2360 / 3200 = 0.7375$ , quedando la PEA-  
No-Agropecuaria ajustada igual a  $300 * 0.7375 = 221.25$   
 $= 221$ .

De esta forma queda establecido el peso definitivo de cada una de las AGEB de la Población objeto de muestra.

## CAPITULO II

### OBTENCION DEL TAMANO DE LA MUESTRA

El capitulo anterior se ocupó en su totalidad en la definición de la Población en estudio; ahora el siguiente paso es elegir el tamaño de muestra por entidad.

Como fue mencionado anteriormente, la selección de la muestra se realizó bajo el esquema de probabilidad proporcional al tamaño (PPT), cuyo desarrollo se detalla en el siguiente capítulo, sin embargo, será necesario en varias ocasiones contemplar el muestreo PPT como un factor decisivo dentro de la obtención del tamaño de la muestra. de esta forma, el presente capítulo se relaciona en algunas partes con el siguiente.

Para la obtención del tamaño de muestra fue implementado un sistema de cómputo, el cual fue destinado para la obtención de ciertos parámetros poblacionales y para efectuar un modelo de simulación sobre los efectos de diseño (DEFF), con objeto de averiguar la ganancia por incorporar el PPT sobre el MIA para la selección de la muestra. El sistema contó con una sección de captura, la cual, se alimentó de los

siguientes datos por cada entidad: Población total, PEA Secundaria y PEA terciaria. Con estos datos el sistema calculaba la PEA-No-Agropecuaria y los siguientes parámetros poblacionales: media, varianza, desviación estandar y coeficiente de variación para cada uno de los diferentes tipos de datos capturados, al mismo tiempo se iba calculando la suma acumulada para la PEA-No-Agropecuaria la cual es necesaria para poder llevar a efecto el muestreo con probabilidad proporcional al tamaño. Con todos los datos anteriores se procedió a realizar la simulación dirigida a la estimación del efecto en el Diseño (DEFF). El procedimiento fue el siguiente: por cada entidad federativa capturada se obtenían 20 muestras de tamaño 20, 40 o 60 y 80, con los esquemas de muestreo irrestricto aleatorio (MIA) y con probabilidad proporcional al tamaño (PPT).

El procedimiento empleado bajo la modalidad del MIA consistió en seleccionar las AGEB cuyo número en el arreglo fuera el mismo valor que el aleatorio generado entre 1 y el número total de AGEB almacenados. Mientras que la selección de acuerdo al PPT se realizó en forma sistemática, con intervalos de regular tamaño seleccionando en cada uno el AGEB correspondiente y de acuerdo al número de arranque (este procedimiento se analiza a detalle en el capítulo III).

Una vez concluida la selección para los dos esquemas (MIA y PPT), se procede a calcular la media muestral de la población y de la PEA=No-Agropecuaria, el proceso se realiza 20 veces (por cada una de las muestras), obteniendo finalmente la media y la varianza de las medias del MIA y PPT esto es  $\bar{X}(\text{MIA})$ ,  $\text{VAR}(\bar{X}(\text{MIA}))$ ,  $\bar{X}(\text{PPT})$ ,  $\text{VAR}(\bar{X}(\text{PPT}))$  respectivamente; a seguida se calculo el cociente que resulta de dividir la varianza del PPT entre la del MIA (esto es  $\text{VAR}(\bar{X}(\text{PPT})) / \text{VAR}(\bar{X}(\text{MIA}))$ ), que es precisamente el efecto del Diseño, el cual reflejó la ganancia de incorporar el PPT en lugar del MIA puesto que, al ser la medida de peso la PEA-No-Agropecuaria, se está considerando una variable altamente correlacionada positivamente con personal ocupado en el AGEB lo que implica que si ella crece, entonces se espera que suceda lo mismo con el número de personal ocupado, asimismo por las características del PPT la probabilidad de seleccionar a un AGEB es proporcional al tamaño de la PEA-No-Agropecuaria haciendo a este más preciso que el MIA de tal suerte que en la simulación se observó en general  $0 \leq \text{DEFF} \leq 1$ , lo cual implica que  $\text{VAR}(\bar{X}(\text{MIA})) \geq \text{VAR}(\bar{X}(\text{PPT}))$ .

Con los resultados de la simulación se realizaron estimaciones sobre población total y PEA sector secundario; esto dió como resultado la elección de

distintos valores en el efecto del diseño de acuerdo al número de AGEB contenidos por entidad: tales efectos de diseño se utilizaron como factores que multiplicaron al tamaño de la muestra obtenido de la fórmula del MIA, de este modo se redujo considerablemente el tamaño de la muestra, conservando precisión y al mismo tiempo minimizando los costos de operación.

Los Estados que se vieron involucrados en el programa de simulación fueron: Tlaxcala, Colima, Quintana Roo, Queretaro, Hidalgo, Tabasco, Nayarit, Zacatecas y Michoacán. El cuadro No.1 muestra los efectos de diseño para dichos estados, mientras que el cuadro No. 2 muestra las medias poblacionales junto con las medias muestrales con el objeto de plasmar la bondad del modelo de simulación.

Los diferentes efectos de diseño de acuerdo al número de AGEB Rurales en la entidad se muestran en el siguiente cuadro:

NUMERO DE AGEB POR ENTIDAD	DEFF
1 - 100	0.05
101 - 500	0.10
501 o más	0.35

La diferencia en los efectos de diseño se explica

C U A D R O # 1  
 EFECTOS DE DISEÑO CALCULADOS PARA PEA DEL SECTOR SECUNDARIO Y  
 POBLACION TOTAL POR MEDIO DE SIMULACIONES CON 20 REPLICACIONES

	<u>ENTIDAD</u>	<u>No. DE AGER</u>	<u>TAMANO DE LA MUESTRA</u>	<u>DEFF POR TOT</u>	<u>DEFF PEA SEC</u>
1)	TLAXCALA	63	20	.097	.001
			40	.127	.001
			60	.005	.0004
2)	COLIMA	69	20	.181	.005
			40	.151	.009
			60	.032	.001
3)	QUINTANA ROO	90	20	.113	.078
			40	.096	.018
			60	.071	.028
4)	QUERETARO	129	20	.029	.001
			40	.162	.002
			60	.137	.023
5)	HIDALGO	253	20	.236	.052
			40	.039	.004
			60	.106	.017
6)	TABASCO	269	20	.159	.004
			40	.068	.014
			60	.226	.004
			80	.278	.016

	<u>ENTIDAD</u>	<u>No. DE AGEB</u>	<u>TAMAÑO DE LA MUESTRA</u>	<u>DEF POR TOT</u>	<u>DEF PEA SEC</u>
7)	NAYARIT	280	20	.320	.012
			40	.215	.009
			60	1.010	.004
			80	.347	.019
8)	ZACATECAS	722	20	1.080	.016
			40	.609	.017
			80	.404	.011
9)	MICHOACAN	737	40	1.04781	0.02483
			80	0.85015	0.01586
			100	1.1640	0.01427

CUADRO # 2

VALORES DE  $\bar{X}$ ,  $\bar{X}$ -MIA Y  $\bar{X}$ -PPT DE POBLACION TOTAL POR CADA ENTIDAD QUE INTERVINO EN EL MODELO DE SIMULACION EN 20 REPLICACIONES POR MUESTRA.

ENTIDAD	TAMANO DE LA MUESTRA.	$\bar{X}$	$\bar{X}$ -MIA	$\bar{X}$ -PPT
Colima	20	1332.42	1295.78	1374.54
	40		1362.30	1302.78
	60		1405.10	1327.20
Tlaxcala	20	3498.95	3759.72	3436.31
	40		3376.38	3453.38
	60		3497.38	3465.45
O.Roc	20	934.07	836.34	954.73
	40		993.90	961.54
	60		958.55	920.55
Gueratario	20	3023.05	3216.00	3076.70
	40		2990.40	3151.47
	60		3143.15	3069.91
Tabasco	20	2455.10	2731.96	2495.88
	40		2603.95	2522.04
	60		2491.77	2569.84
Michoacan	20	1750.52	1746.62	1813.04
	40		1925.44	1816.46
	60		1781.00	1729.05
	100		1795.95	1780.90
Hidalgo	20	4105.63	4086.36	4742.39
	40		4001.75	4607.42
Nayarit	20	1115.76	1047.90	1086.27
	40		1104.45	1059.60
	60		1133.55	1145.60
Zacatecas	20	941.58	899.54	986.22
	40		963.27	918.79

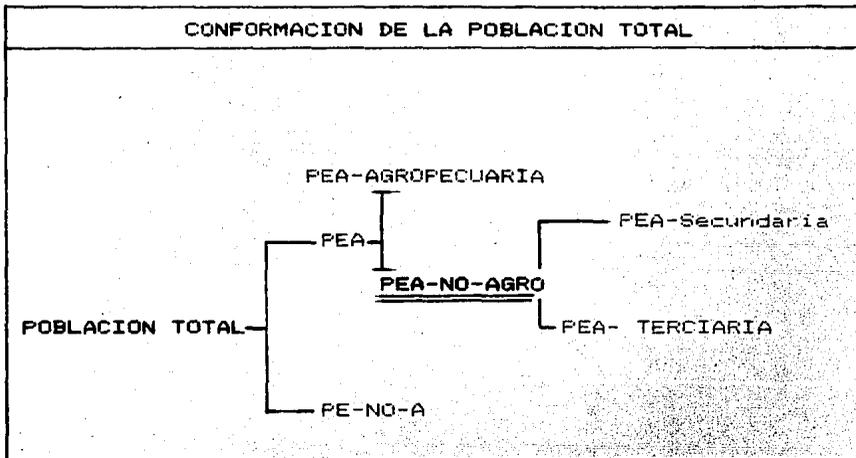
enseguida:

A medida que el número de AGEB disminuye, la ganancia del PPT se acrecenta sobre el MIA, ya que el primero ofrece la ventaja de selección casi sin replazo. Cuando el número de AGEB en la muestra aumenta la varianza disminuye, puesto que los estimadores convergen al valor real del parámetro. Sin embargo el MIA continúa con fluctuaciones.

Además, al aumentar la muestra, las diferencias de peso en la PEA-No-Agropecuaria mayores y menores tienden a ser imperceptibles, de tal suerte que la ganancia obtenida por la incorporación como la medida de peso de la PEA-No-Agropecuaria se reduce.

La obtención de los parámetros para el DEFF se realizó tomando en cuenta el promedio de las medias del DEFF Poblacional total y el DEFF PEA sector secundario, sectorizando por grupos de entidades de acuerdo a su número de AGEB esto es: Entidades con menos de 100; entre 100 y 500 y más de 500 AGEB. Utilizar el promedio de los DEFF se justifica pues si se observa la conformación de la población total como enseguida se muestra, podrá notarse que se esta promediando una medida bastante general (población total) lo que

balancea el promedio obtenido para la PEA-No-Agropecuaria que es una medida muy aproximada al personal ocupado.



De esta forma la fórmula para calcular el tamaño de muestra es:

$$a = DEFF \left( \frac{[ (cv + t) / r ]^2}{1 + 1/A + [ (cv + t) / r ]^2} \right)$$

donde:

a = Tamaño de la muestra

A = Número de AGEB Rurales en la entidad.

t = Absisa correspondiente en la distribución normal al 90% acumulativo

r = Máximo error permitido (.10 en este caso).

cv = Coeficiente de variación de la PEA-Sector Secundario.

Como se habra notado "t" y "r" son constantes para todas las entidades: esto se debe a la relativa homogeneidad de las unidades de selección y a la intención de dar la misma confiabilidad a los resultados sin tomar en cuenta su tamaño. Por otra parte el hecho de haber elegido el coeficiente de variación correspondiente a la PEA del sector Secundario obedece a poder preveer y asimilar variaciones de variables de interés censal que sean subconjunto de otras cuyo comportamiento es semejante al de la PEA-No-Agropecuaria (v.g. personal ocupado por sector; número de establecimientos por tamaño, etc.).

El procedimiento para la obtención del tamaño de la muestra es el siguiente.

De los Catálogos de Integración Territorial (CIT) de cada una de las entidades federativas (de las cuales se hace mención en el capítulo anterior), con los valores de la PEA-No-Agropecuaria completamente

ajustados, se obtienen los siguientes datos: número de AGEB sujetos a muestra media y varianza de la PEA Sector Secundario; con estos dos últimos se obtiene el coeficiente de variación ( $cv = \text{media PEA sector secundario} / \text{desviación estándar de la PEA sector secundario}$ ). Estos datos se substituyen en la fórmula antes expuesta y se obtiene el tamaño deseado.

Cabe aclarar que por no contar con los CIT de los Estados de: Chihuahua, Jalisco, México, Puebla y Veracruz se tuvo que realizar un modelo de regresión con el propósito de estimar los "cv"; donde la variable independiente fue el número de AGEB por entidad y al dependiente el cv por entidad.

En la página siguiente se presenta el Análisis de Varianza (ANOVA), para el modelo en cuestión.

FUENTE DE VARIACION	GRADOS DE LIBERTAD	SUMA DE CUADRADOS	CUADRADOS MEDIOS	F0
Debido a:				
Regresión	1	12.24	12.24	
Error	22	12.27	0.557	
				21.94
Total	23	24.51		

Las pruebas  $H_0: B(1) = 0$  vs  $H_a: B(1) \neq 0$  se hicieron a un 95% y 99% rechazandose en ambas la hipótesis nula, esto es, la pendiente  $B(1)$  es diferente de cero.

Los intervalos de confianza a un 95% para  $B(0)$  y  $B(1)$  son:

$$1.2756 \leq B(0) \leq 1.5410$$

$$0.0021 \leq B(1) \leq 0.0022$$

El cuadro número 3 contiene los datos empleados para la construcción del modelo.

A continuación se muestran las extrapolaciones e intervalos de confianza para los cinco estados involucrados.

ENTIDAD	COEFICIENTE DE VARIACION ESTIMADO	INTERVALOS DE CONFIANZA AL 95%
Jalisco	3.13	(3.03 . 3.23)
Chihuahua	4.58	(4.06 . 5.09)
Veracruz	3.13	(3.03 . 3.23)
México	2.24	(2.19 . 2.28)
Puebla	2.41	(2.36 . 2.45)

Todos los cálculos se presentan en el apéndice al final del documento.

Es necesario aclarar que debido al desconocimiento de número de personal ocupado por establecimiento, la PEA sector secundario se ocupó en el cálculo del CV, sin embargo esto implica el manejo de cierto grado de imprecisión de este dato, de tal suerte que en algunos casos el CV fue demasiado grande. Esto obligó a fijar una cota máxima aceptable (de acuerdo a experiencias

anteriores y/o similares), de tal suerte que el criterio a seguir fue asignar un valor de 2.5 a los CV mayores o iguales a 2.5; el resto (los inferiores a 2.5), conservan su valor.

El cuadro número 4 en la siguiente página muestra por cada una de las entidades federativas el CV definitivo y el tamaño de muestra.

CUACRO No. 4

ENTIDAD	CV	TAMAÑO DE LA MUESTRA	ENTIDAD	CV	TAMAÑO DE LA MUESTRA
Baja California	2.5	26	Guanajuato	1.16	22
Baja California Sur	2.5	30	Querétaro	1.25	10
Sonora	2.5	269	Veracruz	2.5	194
Sinaloa	2.5	152	Hidalgo	1.34	18
Chihuahua	2.5	280	Puebla	2.48	37
Durango	2.5	226	Tlaxcala	1.47	04
Zacatecas	1.78	138	México	2.5	32
Coahuila	2.5	203	Morelos	1.59	4
Nuevo León	2.5	166	Guerrero	2.39	172
Tamaulipas	2.5	185	Oaxaca	1.2	101
Nayarit	2.06	223	Tabasco	1.73	21
Jalisco	2.5	194	Chiapas	2.5	184
Colima	1.55	05	Yucatán	2.35	28
Michoacán	2.21	166	Campeche	2.5	20
San Luis Potosí	1.78	122	Quintana Roo	2.11	5
Aguascalientes	1.16	4			

### CAPITULO III

#### SELECCION DE LA MUESTRA BAJO EL ESQUEMA DE PROBABILIDAD PROPORCIONAL AL TAMAÑO Y REDUCCION DE LA MISMA

El modelo de simulación tratado en el capítulo anterior, demostró la ganancia al incorporar el Muestreo Proporcional al Tamaño (PPT), en lugar del Muestreo Irrestringido Aleatorio (MIA), puesto que al haberse considerado como medida de peso la PEA-No-Agropecuaria se está trabajando con una variable altamente correlacionada con la actividad económica en el AGEB y que permitió orientar la muestra hacia aquellas AGEB con un mayor grado de actividad.

El presente capítulo se dedicará a detallar como se realizó la selección bajo el esquema de muestreo citado.

El muestreo proporcional al tamaño se podrá llevar a efecto si se conocen los tamaños o pesos asociados a cada una de las  $N$  unidades que componen la Población sujeta a muestra.

Cabe señalar que "peso" no se refiere

necesariamente al valor de la variable respuesta sino, que puede tomarse una aproximación a través de una variable auxiliar que se piensa está altamente correlacionada con la característica en estudio (en este caso, la actividad económica no agropecuaria se medirá de acuerdo a la PEA-No-Agropecuaria en el AGEB)

El método a seguir para una selección bajo el esquema PPT se describe a continuación.

Dado un conjunto de AGEB perteneciente a cualquier entidad federativa sea  $M(i)$  el valor de la PEA-No-Agropecuaria para el  $i$ -ésimo AGEB reportada en el Catálogo de Integración Territorial (CIT).

El primer paso es obtener la suma acumulada de la PEA-No-Agropecuaria ( $\sum M(i)$ ) tal y como se muestra en el siguiente cuadro.

UNIDAD	TAMANO	SUMA ACUMULADA	RANGO ASIGNADO
1	$M(1)$	$M(1)$	1 a $M(1)$
2	$M(2)$	$M(1) + M(2)$	$1 + \sum M(i)$ a $\sum^2 M(i)$
3	$M(3)$	$M(1) + \dots + M(3)$	$1 + \sum M(i)$ a $\sum^3 M(i)$
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
A-1	$M(A-1)$	$\sum^{A-1} M(i)$	$1 + \sum^{A-2} M(i)$ a $\sum^{A-1} M(i)$
A	$M(A)$	$M(0) = M(i)$	$1 + \sum^{A-1} M(i)$ a $\sum^A M(i)$

Así, la suma acumulada  $\sum M(i)$  abarcará un conjunto de números que se encuentren en el intervalo cerrado  $[1 + \sum M(i), \sum M(i)]$  y la longitud de este dependerá del valor de  $M(i)$

El procedimiento consiste en extraer un AGEB en forma sistemática (un AGEB por cada  $k$ -ésima de la medida de peso): entonces para una muestra de tamaño "a" tomamos  $k = M(0)/a$  y enseguida se genera un número aleatorio  $r$  ( $r$ : número de arranque) tal que  $1 \leq r \leq k$  a continuación se seleccionan aquellas AGEB cuyo rango asignado contenga a los números  $r, r+k, r+2k, \dots, r+(a-1)k$  o dicho de otra forma se seleccionará el AGEB que tenga la primer Suma Acumulada mayor que  $r+tk$  (para  $t = 0, 1, 2, \dots, (a-1)$ ), (esto es,  $w(t+1) = r+tk$ :  $t = 0, 1, 2, \dots, (a-1)$  se seleccionará el  $j$ -ésimo AGEB [ $j = 0, 1, 2, \dots, (a-1)$ ] si y sólo si la suma acumulada del AGEB es  $S(j)$  donde  $S(j) \geq w(j)$  y además:  $S(j) = \min \{ \sum M(i) ; \sum M(i) \geq w(j) \}$ ,  $w(j) = r+jk$ ,  $k = 1, 2, 3, \dots, A$  ).

El AGEB ( $j$ ) es seleccionado con Probabilidad  $Z(j) = M(j)/M(0)$ . Nótese que  $Z(j)$  varía de acuerdo a la magnitud de  $M(i)$  por lo que las probabilidades son distintas y proporcionales al peso o tamaño de la unidad.

De esta forma se obtiene una muestra en forma sistemática con probabilidad de selección diferente y sin reemplazo.

El esquema dibujado en el cuadro # 1 ilustra este proceso de selección. A lo largo de la recta se ubican los valores de  $\sum M(i)$  de acuerdo a su tamaño.

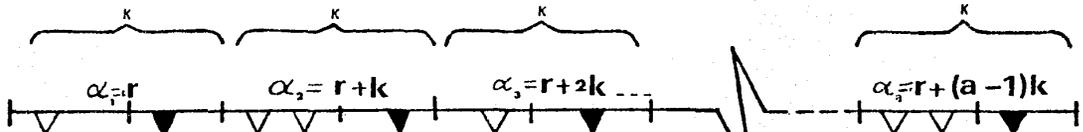
Es importante señalar que en algunos casos no es posible evitar el reemplazo; esto sucede cuando el tamaño de un  $M(i)$  fue lo suficientemente grande para sobrepasar en magnitud al valor de  $k$ , entonces sucede que para uno o varios números de selección, ( $w = r + tk$  es el número de selección), éstos se encuentran dentro del intervalo asociado al mismo AGEB. Esto es para  $w(r)$ ,  $w(r+1), \dots, \dots, w(s)$  donde  $1 \leq s \leq a$  sucede que entonces se selecciona la AGEB con multiplicidad  $(s-r+1)$ . Esta situación se ilustra en el cuadro # 2.

El reemplazo bajo el método señalado puede ser evitado si se sigue el criterio de asignar probabilidad = 1 de selección (esto es, incluirlos directamente en la muestra), a todas aquellas AGEB tales que su tamaño es lo suficientemente grande como para causar el problema de repetición en la muestra, y posteriormente utilizar el esquema PPT con selección en forma sistemática para

CUADRO N° 1

ESQUEMA DE SELECCION SISTEMATICA Y P.P.T.

INTERVALOS



PFA NO AGRO ACUMULADO

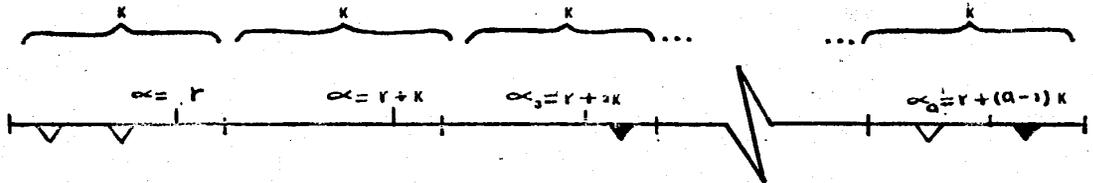


AGEB SELECCIONADA

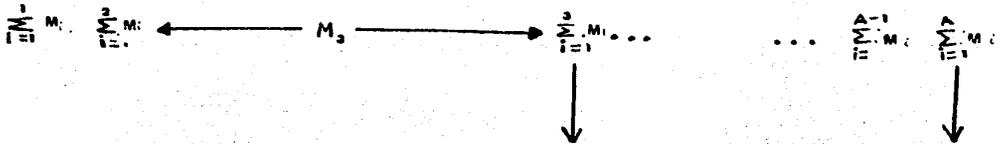
- AGEB - 1  
1,
- AGEB - 5  
2,
- AGEB - 7 ...  
3, ...,
- ... AGEB - A  
.... a

CUADRO N° 2  
 REPETICIÓN DE UN AGEB DENTRO DE LA MUESTRA

INTERVALO



PEA NO  
 ACUMULADA



AGEB  
 SELECCIONADA

AGEB (3)....  
 Para  $\alpha_1, \alpha_2, \alpha_3$   
 (Se incluye 3 veces)  
 NOTESE QUE EL TAMAÑO DE  
 $M_3$  ES SUPERIOR A  $K$

el resto. Es necesario mencionar que si se hubieran incluido las AGEB directamente en la muestra como se acaba de mencionar se tendría que reelaborar las sumas acumuladas, dividir a intervalos regulares y extraer la muestra como se ha descrito pero aunado a esto como fueron excluidos aquellas AGEB con un peso suficientemente grande, esto repercute también en alterar los parámetros poblacionales en la PEA-Secundaria pues recuérdese que ésta se encuentra correlacionada con la PEA-No-Agropecuaria, ya que se trata de un subconjunto de esta última, dado que el CV se calculo de la PEA Secundaria, esta última resentiría la exclusión de los AGEB variando en su tamaño original y reduciéndose en su valor.

Por razones prácticas y por restricciones de tiempo este último criterio no pudo seguirse durante la selección de la muestra rural, por lo cual, en varias entidades federativas se dio el caso de repeticiones de uno o varios AGEB dentro de la muestra.

El siguiente ejemplo explica el desarrollo anterior. Suponga que una entidad ficticia de 10 AGEB cuyas PEA-No-Agropecuarias y suma acumulada son como se detalla a continuación.

AGEB	PEA-No-AGROPECUARIA	SUMA ACUMULADA
1	50	50
2	380	430
3	210	640
4	15	655
5	1170	1825
6	290	2115
7	36	2151
8	507	2658
9	28	2686
10	1	2687

suponga además que el tamaño de muestra "a" es igual a 4 entonces:  $k = M(o)/a = 2687/4 = 671.75 \approx 672$

Luego r es tal que  $1 \leq r \leq 672$ . Sea  $r = 402$  (se obtiene como un número aleatorio bajo una distribución uniforme en este caso).

Por lo cual el primer AGEB es (para  $t = 0$ : en  $w(1) = r + tk \Rightarrow w(1) = 402$ ), aquel cuya menor suma acumulada sea mayor a 402, lo cual se cumple para el AGEB (2), pues la suma acumulada del AGEB (2) = 430 > 402 y ningún otro anterior cumple la desigualdad.

El segundo AGEB es (para  $t = 1$ : i.e.  $w(2) = 1074$ ), aquel cuya menor suma acumulada sea mayor o igual a 1074

por lo cual la suma acumulada AGEB (5) = 1825 > 1074.

El tercero es (para  $t = 2$ :  $w(3) = 1746$ ), nuevamente el AGEB (5), pues la suma acumulada del AGEB sigue siendo mayor que 1746.

El cuarto y último es (para  $t = 3$ :  $w(4) = 2418$ ) el AGEB (8), pues la suma acumulada de este AGEB es 2658 > 2418 y ningún otro anterior cumple la desigualdad.

Por lo tanto la muestra obtenida de tamaño 4 proveniente de una Población compuesta por 10 AGEB, está integrada por los siguientes AGEB.

AGEB (2), AGEB (5), AGEB (5), AGEB (8).

Nótese que el AGEB (5) fue seleccionado dos veces, puesto que la probabilidad de selección de un AGEB depende del tamaño de su intervalo.

El siguiente cuadro muestra los AGEB del ejemplo anterior de acuerdo a la probabilidad de selección de cada uno.

AGEB	PROBABILIDAD DE SELECCION SEGUN LONGITUD DEL INTERVALO
AGEB (1)	0%
AGEB (2)	14%
AGEB (3)	8%
AGEB (4)	1%
AGEB (5)	44%
AGEB (6)	11%
AGEB (7)	1%
AGEB (8)	13%
AGEB (9)	1%
AGEB (10)	0%
	-----
	100%

Del cuadro anterior se observa que las AGEB con mayor probabilidad de selecccion son:

AGEB (5)	con	44%
AGEB (8)	con	19%
AGEB (2)	con	14%

Las cuales fueron seleccionadas mediante el esquema de PRT en este ejemplo.

Cabe señalar que por no contar con los CIT de los estados de Jalisco, Chihuahua, Veracruz, México y Puebla

se empleo como medida de peso, el número de viviendas por AGEB.

Por otra parte, durante la selección de la muestra la única variable que desempeñó un papel importante fue la Población Económicamente Activa no Agropecuaria dentro del AGEB, olvidando otro tipo de factores que posteriormente se vieron involucrados y que afectaron en forma directa los tamaños de muestra. Tales factores fueron, por citar algunos, tamaños del AGEB, accesibilidad al mismo, y muestras enormes, con un alto costo, lo que implicó toparse con dos problemas básicos. Primeramente, asumir el riesgo de un desfazamiento en la calendarización de barrido de AGEB resultado de un retraso de las actividades dentro del operativo. El segundo problema se refiere a la disponibilidad fija de recursos económicos, los cuales no pudieron afrontar el gasto que representa barrer el número tan grande de AGEB que tenían las muestras en ciertas entidades federativas, o el alto costo en tiempo y dinero para barrer algunas, lo cual ocasionaría posiblemente una subcaptación que perturbaría seriamente la inferencia de los resultados.

De esta forma fue necesario reducir el número de AGEB para algunas entidades, cuyo tamaño de muestra fuera lo suficientemente manejable como para correr los

riesgos anteriormente señalados.

La reducción a los tamaños de muestra se efectuó únicamente en 6 entidades federativas, las cuales muestran en el siguiente cuadro las cifras originales:

CUADRO 1

ENTIDAD FEDERATIVA	No. DE AGEB CON REPETICIÓN	No. DE AGEB EN LA MUESTRA SIN REPETICION
Chiapas	184	151
Chihuahua	280	229
Guerrero	172	160
Sinaloa	152	133
Sonora	269	190
Veracruz	192	190

Como se mencionó anteriormente, en varias entidades federativas se dio el caso de que un AGEB fue seleccionado en más de una ocasión, por lo cual la columna "Número de AGEB en la Muestra" se encuentra dividida en dos subcolumnas, donde la primera se refiere al tamaño de la muestra y al segunda al número de AGEB seleccionados para ser barridos en campo.

Para abatir efectivamente los costos, era necesario reducir directamente el número de AGEB por barrer, considerandose convenientemente para las entidades seleccionadas un factor de reducción de 0.25. Sin embargo, para poder determinar los intervalos de confianza resultantes, era menester el tamaño de la muestra teórico después de los ajustes. Esto planteó la siguiente relación:

Factor de reducción en AGEB por barrer: 0.25

Factor de reducción en el tamaño de la muestra:  $x$

Que representa una proporción cuyo valor se puede conocer al fijarse el número de AGEB por barrer y el tamaño de la muestra, y que puede representarse como el cociente de dos enteros, de tal suerte, que para un listado conteniendo previamente ordenada en forma descendente de acuerdo a su PEA-No-Agropecuaria se descartará en forma sistemática tantas AGEB como lo indique la cifra del numerador por cada número que señale el denominador.

Debe notarse que el objeto de ordenar los listados de acuerdo a la cifra de la PEA-No-Agropecuaria, en forma descendente, fue con el propósito de asegurar la eliminación de AGEB con cantidades pequeñas y grandes de PEA-No-Agropecuaria, ya que si en lugar de esto se

hubiera empleado como estrategia el MIA la condición de descartar chicas y grandes no se hubiera cumplido necesariamente.

De esta forma si  $x = p/q$  ( con "p" y "q" primos entre si ), representa la proporción de AGEB que serán eliminadas de la muestra en cierta entidad, quiere decir que se cancelarán p-AGEB por cada q en forma sistemática a intervalos regulares que dependen de los valores de p y q.

A continuación se detalla para cada entidad afectada los cálculos y reglas de decisión para la eliminación de AGEB en la muestra correspondiente:

i) Chiapas

-Tamaño de la muestra: 184

-Número de AGEB por barrer: 151

entonces:

151 : 0.25

184 : x,  $\Rightarrow x = 0.3046 \approx 0.3 = 3/10$

ii) Chihuahua

-Tamaño de la muestra : 280

-Número de AGEB por barrer : 229

entonces:

229 : 0.25

280 : x,  $\Rightarrow x = 0.30567 \approx 0.3 = 3/10$

En ambos casos se eliminan 3 AGEB de cada 10 bajo la siguiente regla:

Los dos primeros por eliminar : 1 de cada 3

el último : 1 de cada 4

Esto es: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 y todas las AGEB cuyo lugar que ocupen termine en 0, 3 o 6.

iii) Guerrero

-Tamaño de la muestra : 172

-Número de AGEB por barrer : 160

entonces:

160 : 0.25

172 : x,  $\Rightarrow x = 0.268 \approx 0.3 = 3/10$

Por lo cual se eliminan 3 AGEB de cada 10 bajo la misma regla que se aplica a los estados anteriores.

iv) Sinaloa.

-Tamaño de la muestra : 152

-Número de AGEB por barrer : 133

entonces:

133 : 0.25

152 :  $x$ ,  $\Rightarrow x = 0.28571 \approx 0.3 = 3/10$

El criterio es el mismo de los anteriores.

v) Sonora

-Tamaño de la muestra: 269

-Número de AGEB por barrer : 190

entonces:

190 : 0.25

269 :  $x$ ,  $\Rightarrow x = 0.3539 \approx 0.36 = 9/25$

Por lo cual se eliminarán 9 AGEB de cada 25 bajo la siguiente regla:

Los 5 primeros por eliminar: 1 de cada 2

El siguiente: 1 de cada 3

Los 3 últimos por eliminar: 1 de cada 4

Esto es: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,  
11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,  
25, ...

vi) Veracruz

-Tamaño de la muestra : 192

Número de AGEB por barrer : 190

entonces:

$$190 : 0.25$$

$$192 : x \Rightarrow x = 0.2526 \approx 0.25 = 1/4$$

Por lo cual se elimina un AGEB de cada 4.

Esto es: 1, 2, 3, (4), 5, 6, 7, (8), 9, 10, 11, (12), ... y todos los múltiplos de 4.

El siguiente cuadro contiene para las mismas 6 entidades los nuevos tamaños de muestra y las AGEB por barrer después de la depuración.

ENTIDAD FEDERATIVA	NUMERO DE AGEB EN LA MUESTRA	NUMERO DE AGEB EN LA MUESTRA DESPUES DE LA DEPURACION
	TAMARO DE LA MUESTRA	AGEB POR BARRER
Chiapas	146	101
Chihuahua	198	155
Guerrero	121	115
Sinaloa	107	88
Sonora	172	101
Veracruz	146	144

El siguiente punto consiste en analizar como se vio afectada la confiabilidad por los efectos de la disminución de las muestras: en el capítulo II "Obtención del Tamaño de la Muestra", se trabajó con la fórmula:

$$a = DEFF * \left( \frac{[(t + cv)/r]^2}{1 + (1/A) + [(t + cv)/r]^2} \right)^{1/2}$$

Donde "t" es la abscisa correspondiente en la distribución normal al .90 acumulativo.

Entonces para calcular la reducción del tamaño en la confianza para la muestra se despeja la variable t en la fórmula anterior y enseguida se asocia la nueva abscisa que le corresponde bajo la distribución normal.

$$t = (1/10 + cv) * \left[ \frac{A * a' - 1/2}{A + DEFF - a'} \right]$$

Donde a' es el nuevo tamaño de la muestra.

El siguiente cuadro reporta por cada entidad federativa el valor de  $t$  calculado de la fórmula anterior, así como la confianza para el nuevo tamaño de la muestra:

ENTIDAD	TAMAÑO DE MUESTRA	" $t$ "	CONFIANZA
Chiapas	146	1.2173	77.65%
Chihuahua	198	1.2030	77.10%
Guerrero	121	1.0348	70.09%
Sinaloa	107	1.0160	69.04%
Sonora	172	1.0994	72.84%
Veracruz	146	1.1655	75.62%

Finalmente a manera de nota anexa, se menciona que contrario a lo expuesto anteriormente, en 5 entidades federativas el tamaño de muestra se incrementó en un 100% aprovechando que su tamaño inicial era lo suficientemente pequeño de manera que además de no repercutir presupuestalmente resulta ideal, pues la muestra es mucho más eficiente y la confianza aumenta considerablemente.

En la siguiente hoja se presentan los 5 estados involucrados, con sus tamaños de muestra anteriores y

actuales.

ENTIDAD	TAMAÑO ANTERIOR	TAMAÑO CON 100% DE INCREMENTO
Aguascalientes	4	8
Colima	5	10
Morelos	4	8
Querétaro	5	10
Quintana Roo	5	10
Tlaxcala	4	8

Por último, el siguiente listado reporta los tamaños definitivos de muestra para cada entidad federativa del País.

ENTIDAD	TAMAÑO DE LA MUESTRA	No. DE AGEB POR BARRER
Aguascalientes	8	8
Baja California	26	24
Baja California Sur	30	27
Campeche	20	19
Coahuila	203	123
Colima	10	10
Chiapas	146	101

Chihuahua	198	155
Durango	226	172
Guanajuato	22	22
Guerrero	121	115
Hidalgo	18	18
Jalisco	194	183
México	32	31
Michoacán	156	152
Morelos	8	8
Nayarit	23	23
Nuevo León	166	105
Oaxaca	101	100
Puebla	37	37
Querétaro	10	10
Quintana Roo	8	10
San Luis potosi	122	106
Sinaloa	107	88
Sonora	172	101
Tabasco	21	21
Tamaulipas	185	139
Tlaxcala	8	8
Veracruz	146	144
Yucatán	28	27
Zacatecas	138	129

## CAPITULO IV

### FACTORES QUE AFECTARON EL TAMANO DE MUESTRA

A lo largo de la etapa del levantamiento rural, Oficinas Centrales fue notificada de una serie de casos atípicos que se dieron indistintamente en varias entidades del país, de tal suerte que será necesario analizar en forma particular para cada entidad los efectos tanto en la muestra como en la Población de la cual se extrajo. Los casos mencionados se describen a continuación:

- 1.- Localidades pertenecientes a AGEB rurales; que fueron cubiertas anteriormente por el Empadronamiento Urbano.
- 2.- localidades que desaparecieron.
- 3.- Localidades no reportadas ubicadas dentro del AGEB.
- 4.- Localidades que se encontraron fuera del AGEB.
- 5.- AGEB que se fraccionaron.
- 6.- AGEB que fueron barridas durante la etapa urbana bajo la misma clave, o bien, sujetas a

modificaciones del Marco Geoestadístico.

- 7.- AGEB peligrosas; en la mayoría de los casos se trata de zonas dedicadas al cultivo de drogas, o AGEB con un alto grado de dificultad para su acceso.
- 8.- AGEB que no aparecían en uno o más productos de la planeación.
- 9.- AGEB cubiertas adicionalmente.
- 10.- Establecimientos importantes fuera del AGEB seleccionado.
- 11.- Establecimientos importantes dentro del AGEB seleccionado.
- 12.- AGEB que fueron afectadas por crecimiento de las urbanas.

Antes de pasar a detallar cada uno de los casos anteriores, se harán unos comentarios de carácter general.

El orden en que se presentan los puntos anteriores

está jerarquizado de acuerdo a su importancia, ya que algunos afectaran únicamente al tamaño de la muestra, mientras que en otros será necesario calcular de nueva cuenta los parámetros poblacionales, esto es, modificaciones en toda la Población.

Por lo tanto primeramente deberán ser atendidos todos los casos que repercuten en el tamaño de la muestra, para que enseguida se proceda, si es necesario, a calcular los nuevos parámetros poblacionales.

La detección de AGEB con problemas (casos atípicos presentados) se puede obtener de dos fuentes: la primera por información directa de las Coordinaciones Censales Estatales, o bien de las Direcciones Regionales; mientras que la segunda consultando dos tipos de listados elaborados via computadora con los resultados del Empadronamiento Rural. Estos últimos contienen información referente al número de establecimientos y personal ocupado por municipio y AGEB en uno de ellos, y municipio, AGEB y localidad en el otro, la diferencia entre ambos es la siguiente:

El primero será empleado en forma más general, puesto que únicamente contiene información de totales por AGEB lo cual podrá auxiliar en la detección de los

casos 5 al 10 que son los puntos referentes a omisión o inclusión de AGEB en la muestra.

El segundo listado será usado para un análisis más exhaustivo, puesto que contiene la información más detallada, (los totales llegan a nivel localidad), motivo por el cual servirá de ayuda en la detección de los casos 1 al 4.

Cabe señalar que la información recibida por las Coordinaciones Censales Estatales y/o por las Direcciones Regionales, o bien mediante la consulta de los listados deberá retroalimentarse una a la otra a fin de garantizar una búsqueda completa de estos casos.

El conjunto de casos atípicos se puede dividir en tres grupos. Aquellos que se refieren a efectos en localidades, los que tratan con AGEB y los referentes a establecimientos importantes.

Los efectos en localidades destinadas para ser levantadas repercutieron directamente en los valores de la PEA secundaria y peso, lo cual incide en los parámetros poblacionales y que por tanto deberán calcularse nuevamente, mientras que los efectos en AGEB causaron problemas en la mayoría de los casos al tamaño

de la muestra y solo en unos casos se vieron afectados los parámetros poblacionales y el tamaño de la población.

Enseguida se detallan cada uno de los casos mencionados, su relación con los demás y el tratamiento asignado.

Caso No. 1: Localidades pertenecientes a AGEB rurales que fueron barridas durante la etapa urbana.

El barrido de dichas localidades obedeció a diversas causas, las más frecuentes fueron:

- Localidades que fueron cubiertas por equivocación.
- Localidades que se encontraban conurbadas con otras que efectivamente eran objeto del Empadronamiento.
- Localidades que sin contar con AGEB urbana, su población es superior o igual a 2,500 habitantes, motivo por lo cual fueron consideradas en la etapa urbana.

Este caso puede verse relacionado con 2, 5 y 6, por lo cual es necesario averiguar primeramente, si se barrieron únicamente localidades y no todo el AGEB pues en este último caso deberá excluirse de la Población y de la muestra (casos 2 y 6).

El tratamiento a seguir para este caso es el siguiente: se deberá restar del AGEB afectado la población de las localidades que fueron barridas y en seguida ajustar la PEA secundaria y la PEA-No-Agropecuaria del AGEB siguiendo los lineamientos establecidos en el Capítulo I "Delimitación de la Población", esto es, en forma proporcional de acuerdo a la población total.

Por lo tanto, todas las modificaciones que se hagan a la PEA secundaria y a la PEA-No-Agropecuaria, así como las fórmulas que de aquí en adelante se empleen, se referirán siguiendo la nomenclatura que se indica en el siguiente cuadro, el cual se aplica en forma general para cualquier entidad del País.

AGEB	POBLACION	PEA-SECUNDARIA	PEA-NO-AGROPECUARIA
1	X(11)	X(21)	X(31)
2	X(12)	X(22)	X(32)
.	.	.	.
.	.	.	.
a	X(1a)	X(2a)	X(3a)

Las siguientes fórmulas se ocupan para las modificaciones en suma ( $\sum X_{ij}$ ) y suma de cuadrados ( $\sum X_{ij}^2$ ).

Para la PEA secundaria y PEA-No-Agropecuria ( $X_{ij}$  se refiere al valor original en el j-ésimo renglón de la i-ésima columna).

Entonces para el cálculo de las modificaciones a  $\sum X_{ij}$ ,  $\sum X_{ij}^2$  se tiene:

$$1) \left( \sum_{k+1}^A X_{ij} \right) = \left( \sum_k^A X_{ij} \right) - X_{ij} + Y_{ij}$$

$$2) \left( \sum_{k+1}^A X_{ij}^2 \right) = \left( \sum_k^A X_{ij}^2 \right) - X_{ij}^2 + Y_{ij}^2$$

donde  $Y_{ij}$  es el valor modificado del j-ésimo AGEB en la i-ésima columna de PEA, esto es  $Y_{ij}$  sustituye a  $X_{ij}$ .

Caso No. 2: Localidades que desaparecieron.

Este caso se debe en general a una de las dos siguientes razones:

- La Localidad fue absorbida por la mancha urbana,  
o bien

- La Localidad dejó de existir por fenómenos de  
carácter migratorio.

Nótese que puede relacionarse con los casos 5 y 6  
por lo cual deben hacerse las averiguaciones  
convenientes.

Este caso no implica ningún problema, por lo cual  
deberá obtenerse la información restante dentro del  
AGEB, es decir, aquella que proviene de las localidades  
que si se encuentran, y no deberán modificarse los pesos  
del AGEB ni los parámetros poblacionales.

Caso No. 3: Localidades no reportadas ubicadas  
dentro del AGEB

Este caso se trata de localidades que no se  
encuentran reportadas dentro de los listados de  
cobertura, y en ocasiones se detectan consultando  
Cartografía, o en último caso encontradas durante el  
operativo de campo.

El tratamiento a seguir es similar al anterior,  
pues al no estar reportadas no es posible obtener  
información referente a su PEA-No-Agropecuaria, por lo

cual, suponemos que no existe sesgo y se deja el AGEB sin modificar pesos ni parámetros.

#### Caso No. 4: Localidades ubicadas fuera del AGEB seleccionado

Este caso se refiere a localidades que se encuentran reportadas dentro de un AGEB pero que en realidad su ubicación está fuera de los límites de éste; lo cual obedece en el mejor de los casos a interpretaciones cartográficas.

El tratamiento a seguir es ajustar los valores de PEA secundaria y PEA-No-Agropecuaria, a fin de modificar el peso del AGEB para la inferencia de los resultados.

Concluidos los casos referentes a localidades, el siguiente grupo comprende aquellos que se refieren a problemas con AGEB en lugar de localidades.

#### Caso No. 5: AGEB que se fraccionaron.

Este es un caso bastante especial y un tanto delicado pues deben considerarse tres subcasos a saber:

- a) El AGEB rural se fraccionó en uno o más

AGEB urbanas debido a crecimiento de localidades en el empadronamiento urbano, pero sin que el original desapareciera por completo. (figura 1).

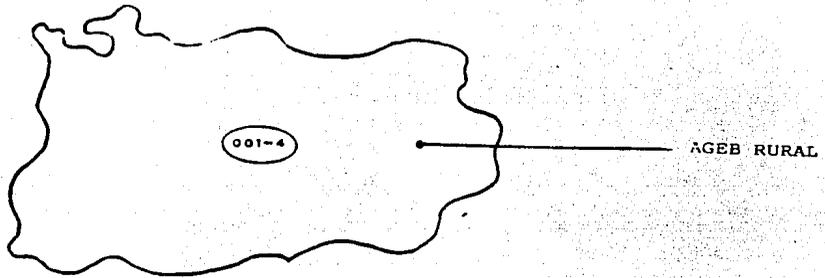
En este caso se deberá esperar la información de cuales son las localidades que se ubican dentro de AGEB rural, a fin de hacer los ajustes debidos en la PEA-secundaria y PEA-No-Agropecuaria en el AGEB y después los cambios en las sumas de cuadrados empleando las fórmulas (1) y (2).

b) El AGEB rural se fracciona en uno o más AGEB urbanos debido a crecimiento de localidades en el Empadronamiento Urbano, desapareciendo por completo la parte rural. (figura 2).

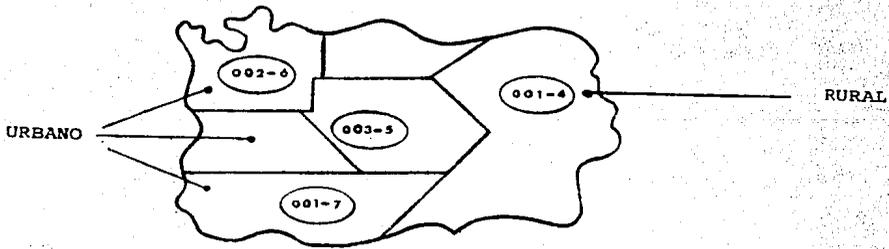
En esta caso deberá excluirse dicho AGEB, tanto de la muestra como de la Población de estudio, pues recuérdese que en el capítulo 1 excluyeron las AGEB urbanas de localidades barridas en el Empadronamiento Urbano, por lo cual para las AGEB que caigan en este caso se modificarán las sumas de A y de a.

Para las sumas se aplicaran las fórmulas (1) y (2) con  $Y_{ij}$  = PEA secundaria y PEA-No-Agropecuaria.

FIGURA NO. 1



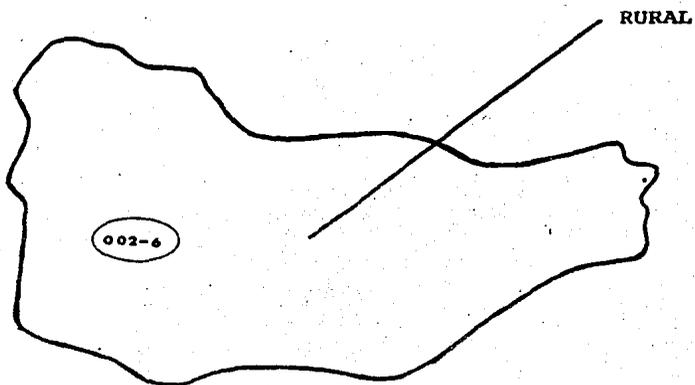
SITUACION ORIGINAL



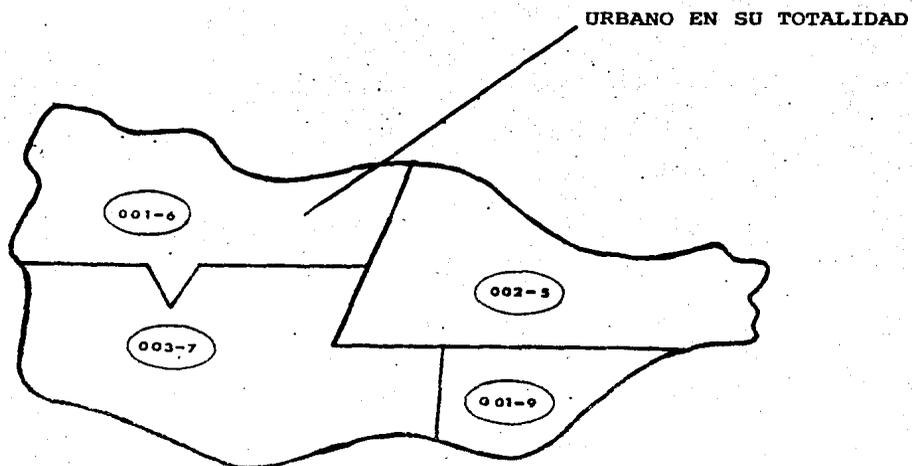
DESAPARICION PARCIAL DE LA PARTE RURAL

FIGURA 2

DESAPARICION DE AGEB RURAL



SITUACION ORIGINAL



DESAPARICION TOTAL DE PARTE RURAL

Los valores de "A" y "a", se modifican como sigue:

Para cada AGEB en este caso:

$$A_{(k+1)} = A_{(k)} - 1$$

$$a_{(k+1)} = a_{(k)} - 1$$

c) El AGEB rural fue dividido por el crecimiento de una localidad urbana dando por resultado varias AGEB rurales (figura 3).

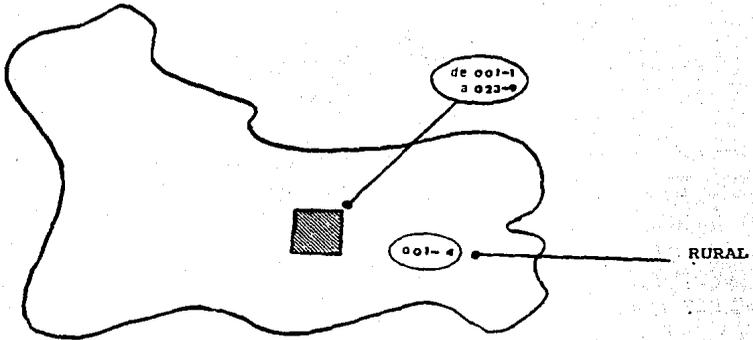
El tratamiento para este caso, es similar al del inciso (a); esto es, se considerarán todas las AGEB rurales como una sola y enseguida se procederá a descontar la población de la parte urbana, es decir, el objeto es realizar los ajustes correspondientes a la PEA secundaria y a la PEA-No-Agropecuaria en la AGEB primitiva, y después los cambios en las sumas de cuadrados, empleando para esto las fórmulas (1) y (2).

caso No. 6: AGEB barridas en la etapa urbana.

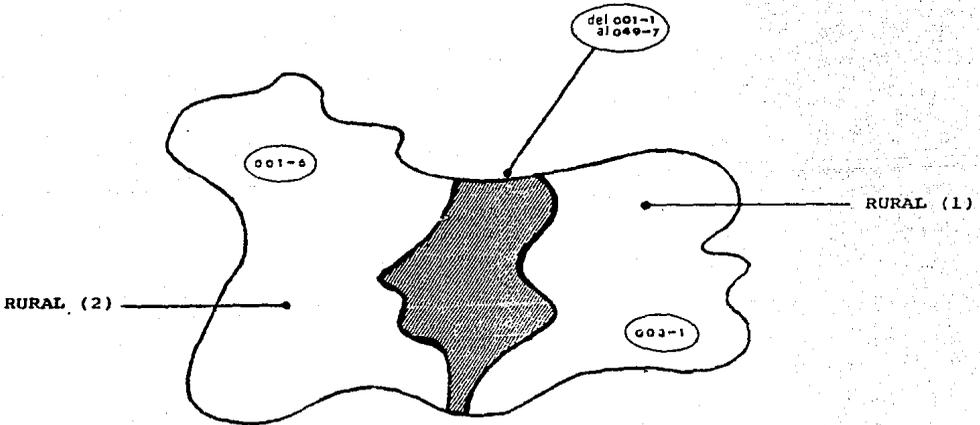
Este caso se refiere a aquellas AGEB que por crecimiento de la mancha urbana y/o de localidades cuya

FIGURA 3

DIVISION DEL AGEB RURAL



SITUACION ORIGINAL



FORMACION DE 2 AGEB RURALES

población fue mayor o igual a los 2,500 habitantes.

fueron barrida por completo durante la etapa del Empadronamiento Urbano; sin embargo, cabe aclarar que en algunos casos tales AGEB se barrieron por equivocación y no por urbanización del mismo. A continuación se detalla el tratamiento a seguir:

El caso puede estar relacionado con el 5, aunque se supone siempre un barrido total del AGEB y además se consideró como urbano, motivo por el cual deberá ser tratado de igual modo que el subcaso B.

Caso No. 7: "AGEB" peligrosas o bien de difícil acceso.

Este caso se refiere a las zonas que por sus "actividades" resultan altamente peligrosas para todo aquel que se introduzca en éstas, por lo cual el AGEB fue excluido del levantamiento y se tratara como no respuesta, su tratamiento consiste únicamente en calcular la nueva "a" y la reducción de la confianza para la muestra (esto se indica más adelante) con los parámetros originales a excepción de "a".

También se verá afectado el peso total (Mo), sin embargo, esto será tratado en el capítulo siguiente.

Caso No 8: AGEB que no aparecieron en por lo menos un producto de la Planeación.

Oficinas Centrales elaboró un listado de AGEB que fueron omitidas en por lo menos un producto de la Planeación Rural, por tanto el tratamiento consiste en averiguar en los reportes de avance si fue o no barrido el AGEB, y en caso de omisión averiguar las causas para canalizarlo posteriormente a cualquiera de los casos restantes.

Caso No 9: AGEB cubiertas adicionalmente.

En caso de detectar que uno o varios AGEB no contemplados dentro de la muestra fueron barridos, éstos deberán excluirse pues se contraponen al proceso de selección de la muestra y por consiguiente, se podría acarrear cierto grado de sesgo en la información.

Caso No. 10: Establecimientos importantes fuera de las AGEB seleccionadas.

este caso es completamente ajeno a los propósitos de la muestra rural y requerirá un tratamiento especial dentro de los lineamientos para el sector urbano, por lo cual debe excluirse de los datos de la muestra.

Caso No 11: Establecimientos importantes dentro del AGEB seleccionado.

Este caso es bastante especial pues por estar dentro del AGEB seleccionado no puede excluirse como en el caso anterior, por lo cual, deberá considerarse en la inferencia y también en el Censo.

Caso No 12: Otras AGEB que fueron afectadas por el crecimiento de las urbanas.

Este caso se refiere a aquellos AGEB que fueron invadidos por el crecimiento de las localidades urbanas interna o externamente (esto es localidades dentro o fuera del AGEB respectivamente), por lo cual su tratamiento será primeramente ajustar de acuerdo a los lineamientos del Capítulo I y en seguida ajustar los parámetros poblacionales.

En seguida se calculan los valores correspondientes a los parámetros poblacionales: media, desviación estandar y coeficiente de variación de la PEA secundaria y la suma total de la PEA-No-Agropecuaria, así como la confianza para la muestra, si ésta se modificó.

Las fórmulas son las siguientes:

$$\text{media: } \bar{x} = \frac{\sum x(i)}{A}$$

$$\text{desviación estandar: } = \left( \frac{\sum x(i)^2 - A \cdot \bar{x}^2}{A} \right)^{1/2}$$

coeficiente de variación:  $cv = (\text{desv. est.}) / \text{media}$

La suma total de la PEA-No-Agropecuaria (el peso en este caso), es:

$$M_0 = \sum_{i=1}^{i=A} x(i,3)$$

Por último, la confianza para la muestra será el valor que reporte "t" que es la abscisa bajo una distribución normal donde:

$$t = \frac{1}{10 \cdot cv} * \left( \frac{A \cdot a}{A \cdot DEFF - a} \right)^{1/2}$$

El DEFF se emplea como se indico en el capitulo II.

A continuacion se presenta un ejemplo para el tratamiento de estos casos en el estado de Aguascalientes.

Los casos presentados en Aguascalientes fueron: exclusion de AGEB y supresion de localidades por haberse barrido durante el Operativo Urbano; estos se detallan a continuacion:

MUN	AGEB	OBSERVACIONES	TRATAMIENTO	PEA-SEC	PEA-N-A
001	108-7	barrida en el Empadronamiento Urbano.	Excluir de la Poblac. y de la muestra.	173	319
002	005-2	Localidad rural que se barrió en urbano, nombre de la localidad "Villa Juarez"	Ajuste	203	338
003	006-4	Mismo caso que el Mpo. 002.		156	434

Los parámetros de Aguascalientes son:

Media PEA Secundaria = 65.05

Desviación PEA Secundaria = 75.59

Poblacion= 65 AGEB

Muestra = 3 AGEB

cv = 1.16

Como primer paso se ajustará la PEA secundaria y la PEA-No-Agropecuaria para los municipios 002 y 003.

Para el primero, su población total es de 5,781. Entonces de acuerdo al procedimiento para ajuste descrito en el Capítulo I, tenemos la localidad por eliminar: Villa Juarez con 2,161 habitantes.

Entonces:

$$\text{PEA secundaria ajustada} = \frac{5781-2161}{5781} * 203 = 127$$

$$\text{PEA-No-Agro ajustada} = \frac{5781-2161}{5781} * 338 = 212$$

Para el municipio 003 su población es de 6,899 y la localidad por eliminar es Mal Paso, con 2,562 habitantes.

Entonces PEA secundaria ajstada = 121

PEA No Agro ajustada = 336

El siguiente paso consiste en modificar los parámetros poblacionales de la entidad, considerando los tres casos.

Para los cálculos de media y desviación se requiere conocer  $\sum X(i)$ ,  $\sum X(i)^2$  de PEA secundaria y PEA- No-agropecuaria, entonces se calcularán a partir de los datos iniciales respectivamente, y eliminando el valor correspondiente al AGEB que se excluya tanto de la muestra como de la Población.

Entonces :

$$[\sum X(i)] = A\bar{X} + (127 - 203) + (121 - 156) - 173 = 5245.25$$

Como únicamente se excluye un AGEB, entonces:

$$A = 83 \text{ y media} = 63.2$$

Luego:

$$(\sum X(i)^2) = A[\text{desv-estandar} + \text{media}^2] + (127*127 - 203*203) + (121*121 - 156*156) - 173*173 = 780.650.801$$

y la desviación estandar = 73.56

Por lo que CV= 1.164

Nótese que el cv es el mismo que el inicial, esto es,  $cv' = cv$  por tanto, el único valor que cambia para el cálculo del tamaño de la muestra es la medida de la Población que en este caso es 83; sin embargo este cambio resulta insignificante, como para afectar el tamaño de la muestra, puesto que la aplicación de la fórmula da como resultado nuevamente un tamaño igual a 4 además, debe recordarse (Capítulo III) que la muestra en Aguascalientes se incrementó en un 100%. Por lo cual, la confianza de la muestra no se ve afectada en ningún momento.

En el siguiente cuadro se resumen los casos presentados y las modificaciones hechas.

MODIFICACIONES DURANTE EL OPERATIVO RURAL A NIVEL  
AGEB, PARA LA ENTIDAD: AGUASCALIENTES

ANTES DEL TRATAMIENTO

No. TOTAL DE AGEB (N) 84

No. DE AGEB EN LA MUESTRA (n) 8

DESPUES DEL TRATAMIENTO

No. TOTAL DE AGEB (N) 83

No. DE AGEB EN LA MUESTRA (n) 7

MPIO.	AGEB	OBSERVACIONES	TRATAMIENTO	DATOS ORIGINALES			DATOS CORREGIDOS		
				Pob. Total	PEA Sec.	PEA no Agrop.	Pob. Total	PEA Sec.	PEA no Agrop.
001	108-7	Barrida en el Empadronamiento Urbano.	Excluir de Universo y Muestra.	----	173	319	----	----	----
002	005-2	Localidad Rural que se Barró en el Empadronamiento Urbano.	Ajuste	5,781	203	338	3,620	127	212
003	006-4	Mismo caso que el Municipio 002.	Ajuste	6,899	156	434	4,337	121	336
	Media			65.05				63.2	
	Desviación			75.59				73.56	
	CV			1.16				1.16	

## CAPITULO V

### INFERENCIA DE LOS RESULTADOS

Con los capitulos anteriores se consiguió disponer del soporte necesario para tratar la información recabada durante el levantamiento rural, de tal suerte que esta pudiera servir como un insumo dispuesto para extrapolar adecuadamente a nivel estatal un conjunto de variables representativas; estas variables son: número de establecimientos en total, establecimientos mayores y menores, Personal Ocupado, remunerado, no remunerado, cantidad de personal masculino y femenino remunerado y no remunerado, remuneraciones en miles de pesos y por último gastos e ingresos.

La información sobre estas variables se obtuvo de los cuestionarios para zonas rurales, los cuales fueron el instrumento empleado durante el operativo para captar la información obtenida de cada establecimiento. El cuestionario para zonas rurales se encuentra dividido en 4 secciones, estas son:

- I Datos Generales
- II Actividad Económica
- III Datos Complementarios
- IV Gastos e Ingresos

Son de interés para los fines de este capítulo las secciones II, III y IV ya que de ellas obtenemos la siguiente información:

#### Sección II Actividad Económica

- i) Personal Ocupado
- ii) Clase de Actividad

#### Sección III Datos Complementarios

- i) Categoría Jurídica
- ii) Tipo de Propiedad

#### Sección IV Gastos e Ingresos

Cada sección se encuentra referenciada por medio de una letra y un subíndice los que serán usados a lo largo del capítulo para hacer alusión a ello.

La información captada sobre estas variables servirá para inferir resultados por Entidad Federativa para un conjunto de secciones, los cuales pueden agruparse en las siguientes tres divisiones:

- i) Por Clase de Actividad .- Con las siguientes clases: Por sector y rama de acuerdo a lo establecido por la CMAP .  
\*

\* Sobre esta se hablará posteriormente.

Por Categoría Jurídica.- Con las siguientes clases: Propiedad Individual, Sociedad Cooperativa, Ejido, Dependencia del Gobierno y Otro Tipo.

iii) Por tipo de Propiedad.- Sus clases son: Privada, Pública Federal, Pública Estatal, Pública Municipal, Social y Otras.

El proceso de inferencia se hará partiendo de los resultados a nivel entidad y posteriormente contemplará las divisiones que se definieron. Para lograr esto, se requiere de la elaboración de un material de apoyo, el cual consiste en una serie de listados en forma tabular, que contiene información obtenida durante el operativo rural, así como datos cuyo cálculo se efectuó en Oficinas Centrales, las cuales son indispensables para el control y procesamiento de la información.

Uno de estos listados consiste en la elaboración de un reporte a nivel entidad en el cual se vaciará por Municipio y AGEB la cantidad correspondiente al peso de este último, que se le asignó de acuerdo al valor de la PEA-No-Agropecuaria,. Al final del listado se anotará el total de la suma de la columna que reporta el peso de

cada AGEB que conforma la entidad y el tamaño de la muestra.

Con los insumos preparados se procede a describir la metodología para inferir los resultados de acuerdo a los siguientes puntos.

Número de Establecimientos, Personal Ocupado, Remuneraciones, Gastos e Ingresos por:

- 1.- Entidad Federativa.
- 2.- Clase de Actividad y Entidad Federativa.
- 3.- Entidad Federativa y Categoría Jurídica.
- 4.- Entidad Federativa y Tipo de Propiedad.
- 5.- Rama de Actividad y Entidad Federativa.

El procedimiento es el siguiente:

Para el número de establecimientos y Personal Ocupado se empleará la siguiente fórmula:

$$\hat{Y}_{est} = (1/a) * \sum Y(i) / Z(i)$$

Donde:

$\hat{Y}$  est. representa el valor estimado de establecimientos o bien de personal ocupado en la Entidad.

a: número de AGEB que conforman la muestra. Nótese que cada AGEB será considerada en la muestra tantas veces como aparezca en la muestra.

$Y(i)$ : representa para el primer caso, el número estimado de establecimientos en el  $i$ -ésimo AGEB; o bien la cantidad de personal ocupado por entidad para cualquiera de las variables P01, P02, ..., P09 correspondientes a la pregunta número 4 de la sección II del cuestionario para zonas rurales las cuales son:

- i) P01 total remunerado hombres y mujeres.
- ii) P02 total remunerado hombres.
- iii) P03 total remunerado mujeres.
- iv) P04 total no remunerado hombres y mujeres.
- v) P05 total no remunerado hombres.
- vi) P06 total no remunerado mujeres.
- vii) P07 total de remunerados y no remunerados incluyendo hombres y mujeres.
- viii) P08 total remunerado y no remunerado hombres.

ix) P09 total remunerados y no remunerados mujeres.

$Z(i)$ : Cociente que resulta de dividir el peso del  $i$ -ésimo AGEB entre la suma total de peso de la entidad, el cual se localiza en los listados elaborados centralmente, es decir  $Z(i) = M(i) / M(o)$ .

La varianza viene dada por:

$$V(\hat{Y}) = \left( Y(i)/Z(i) - (1/a) * \sum Y(i)/Z(i) \right)^2 / a + (a-1)$$

donde  $Y(i)$ ,  $Z(i)$ , y "a" son como se definieron en el párrafo anterior.

Para las variables: Remuneraciones, Gastos e Ingresos, debe considerarse que éstas se obtuvieron de cada uno de los establecimientos que se ubicaron dentro del AGEB, de tal suerte que esto describe un conglomerado donde las unidades son los AGEB y los elementos los establecimientos; por lo que podemos considerar el número de AGEB como una variable auxiliar. Luego entonces, este tipo de variables se extrapolarán para el monto anual a nivel entidad de acuerdo a la variable G12 de la pregunta 14 Sección IV Gastos e Ingresos usando el esquema para muestreo para proporciones dentro de un conglomerado con distinto tamaño.

La fórmula es:  $\hat{Y}_{\text{razón}} = M_0 * \frac{\sum Y(i)}{\sum M(i)}$

donde:

$\hat{Y}$  : representa el monto estimado para cada una razón

variables: Remuneraciones, Gastos e Ingresos de acuerdo a la variable G12 de la pregunta 14 sección IV Gastos e Ingresos.

$M_0$ : representa la suma total de unidades en la entidad, es decir, el número total de establecimientos en la entidad, sin embargo, esta cantidad no se conoce exactamente, por lo cual, el valor obtenido en la estimación del número de establecimientos en la entidad será empleado como un estimador de  $M_0$ ; esto, desde luego implica acarrear cierto sesgo aunque no peligroso.

$Y(i,j)$ : = G12 para el j-ésimo establecimiento dentro del i-ésimo AGEB.

Entonces:  $Y(i) = \sum_j Y(i,j)$  = Cantidad de remuneraciones, Gastos o bien Ingresos en el año 1985

dentro del i-ésimo AGEB.

$Y(i)$ : representa el monto para cualquiera de Remuneraciones, Gastos o Ingresos en un año para la muestra.

$a$ : representa el tamaño de la muestra.

Asimismo, una estimación de la varianza viene dada por la fórmula:

$$v(\hat{Y}_R) = \frac{A^2 \left( \frac{A-a}{A} \right)}{a} \sum M_i \frac{\left( \frac{Y_i}{M_i} - \hat{Y} \right)^2}{A-1}$$

$$\hat{Y} = \frac{\sum Y_i}{\sum M_i}$$

$A$ : representa el total de AGEB en la entidad.

El siguiente cuadro muestra la nomenclatura con la que se hará referencia a los resultados obtenidos:

VARIABLE	TOTAL	VARIANZA
Establecimientos en la entidad.	$\hat{Y}$ -estab	$v(\hat{Y}$ -estab)
Personal ocupado en cualquiera de		

las variables: Po1, po2, ... Po9.	^ Y-Po1	^ v(Y-Po1)
Remuneraciones en el año 1985 a nivel entidad.	^ Y-Rem	^ v(Y-Rem)
Gastos en el año 1985 a nivel en- tidad.	^ Y-Gas	^ v(Y-Gas)
Ingresos en el a ño 1985 a nivel- entidad.	^ Y-Ing	^ v(Y-Ing)

2.- Número de establecimientos, Personal Ocupado, Remuneraciones, Gastos e Ingresos por Entidad Federativa y clase de actividad.

Las variables en este apartado serán tratadas por cada Entidad Federativa de acuerdo a su clase de actividad tal y como se encuentran clasificadas en la "Clasificación Mexicana de Actividades y Productos" (CMAP); la cual funciona como un codificador de las actividades económicas para ser clasificadas, de tal suerte que permite distribuir los datos económicos según categoría de actividades, cuyas características y funcionamiento son análogos.

La CMAP contiene cuatro niveles de agregación:

Sector, Subsector, y Clase de actividad, donde cada nivel es una división de su inmediato anterior.

Para los objetivos de este documento solo se ocuparan los tres primeros niveles de agrupación, es decir, Sector, Subsector y Rama; el esquema a seguir para la estimación sera por proporciones y porcentajes para número de establecimientos y Proporciones y porcentajes dentro de conglomerados de distinto tamaño para el resto de las variables.

La CMAP se compone de 9 sectores:

- 1.- Agricultura, Ganadería, Caza, Pesca, Sivicultura y Servicios para la Agricultura y Ganadería.
- 2.- Minería y Extracción del Petróleo.
- 3.- Industrias Manufactureras.
- 4.- Electricidad, Gas y Agua.
- 5.- Construcción.
- 6.- Comercio Restaurantes y Hoteles.
- 7.- Transporte y Almacenamiento, Comunicaciones y

Agencias de Viaje.

8.- Servicios Financieros, Inmobiliarios, Alquiler de Bienes y Servicios Profesionales.

9.- Servicios Comunales, Sociales y Personales.

Cada uno de los Sectores se Divide en Subsectores y éstos a su vez en Ramas. Por lo tanto, podemos dividir el número total de establecimientos en una entidad (Y-est, en este caso), en 9 clases cada una por sector tal y como se encuentra en la siguiente gráfica:

POBLACION	NUMERO TOTAL DE ESTABLECIMIENTOS (Y-est)
clase sector	C1, C2, ..., ..., C8, C9 Sector1, Sector2, ..., ..., sector9

A su vez cada sector se puede tomar en cuenta como una "población" donde cada subsector es una Subpoblación o Dominio y cada Rama es una clase dentro de cada Dominio.

Esto se muestra gráficamente:

POBLACION	SECTOR (i): i= 1, 2,... ...9			
	D1 Subsector 1		D2 ... Subsector 2	
Dominio (Di)			...	
Clases (Ci)	C1, C2... Cn	C1, C2... Cn		
	Rama <sub>1</sub> ... Rama <sub>n1</sub>	Rama <sub>1</sub> ... Rama <sub>n2</sub>		

De las dos gráficas anteriores se desprende que, primeramente se procederá a estimar los valores correspondientes a sectores y, posteriormente efectuar la ponderación a nivel subsector y Rama, apoyándose para esto en los valores obtenidos a nivel sector, la estrategia se describe a continuación.

## 2.1) SECTORES

### 2.1.1) Establecimientos por Sector

La nomenclatura es la siguiente:

sea:

$N = Y\text{-est}$

$C(i) = \text{sector } (i); i = 1, 2, \dots, \dots 9$

$A(i)$  = número de unidades de  $N$  que caen dentro de  $C(i)$ .

$a(i)$  = número de establecimientos de la muestra que caen dentro de  $C(i)$ .

$n$  = número de establecimientos en la muestra.

$P(i)$  = Proporción de unidades en  $C(i)$

donde:  $P(i) = A(i)/N$

$p(i) = P(i)$ ; proporción de unidades procedentes de la muestra que caen en  $C(i)$ .

donde:  $p(i) = a(i)/n$

además:  $N = \sum A(i)$  y  $n = \sum a(i)$

El procedimiento para estimar  $A(i)$  se logra por medio de la siguiente fórmula:

$$\hat{A}(i)\text{-est} = N * p(i)$$

y un estimador para la varianza es:

$$V(A(i)) = \frac{N-n}{N-1} * \frac{N * p * (1-p)}{n}$$

Los cálculos se harán tomando en cuenta el primer elemento del campo asignado a la variable Ao3 de la pregunta 8 de la sección II, que puede tener como valor 1, 2, ... o 9 dependiendo del sector que se trata, el cual aumentará el valor de su correspondiente a(i).

La nomenclatura para representar estos resultados, es la siguiente:

SECTOR	TOTAL DE ESTAB. EN EL SECTOR	VARIANZA ESTIMADA DEL TOTAL
1	$\hat{A}(1)-est$	$V[\hat{A}(1)-est]$
2	$\hat{A}(2)-est$	$V[\hat{A}(2)-est]$
.		
.		
9	$\hat{A}(9)-est$	$V[\hat{A}(9)-est]$

2.1.2 Personal ocupado, Remuneraciones, Gastos e Ingresos por Sector.

a) Personal ocupado

sea:

$N(p_{oj})$ : toma cualquiera de los valores  $\hat{Y}(p_{oj})$ .

$A(i)$  : número de elementos de  $N$  que caen dentro  
de  $C(i)$ .

$a(i)$  : cantidad de personal ocupado referente  
 $(p_{oj})_k$  a  $P_{oj}$ , del  $k$ -ésimo AGEB (conglomerado)  
de la muestra que cae en  $C(i)$ .

$m$  : cantidad de personal ocupado dentro del  
 $(p_{oj})_k$   $k$ -ésimo AGEB ( $k$ -ésimo conglomerado).

$n$ : número de AGEB en la muestra.

$n$  : cantidad de personal ocupado correspondien-  
 $p_{oj}$  te a  $P_{oj}$  en la muestra.

La proporción de unidades perteneciente a la muestra es como se muestra en la siguiente página.

$$P_{(poj)} = \frac{\sum_{k=1}^m a_{(poj)}^{(i)}}{\sum_{k=1}^m m_{(poj)k}}$$

y un estimador para el total de personal ocupado para la variable  $P_{oj}$  ( $j=1, 2, \dots, 9$ ) dentro del  $i$ -ésimo sector es:

$$\hat{A}_{(poj)}^{(i)} = N_{poj} \cdot P_{(poj)}$$

y un estimador para la varianza es:

$$V(\hat{A}_{(poj)}^{(i)}) = \frac{\hat{N}_{poj}^2 \left( \frac{\hat{N}_{poj} - n_{poj}}{\hat{N}_{poj}} \right)}{n \cdot \bar{m}_{poj}^2} \cdot S$$

$$S = \frac{\sum (a_{(poj)k}^{(i)})^2 - 2 P_{poj} \sum a_{(poj)k}^{(i)} m_{(poj)k} + P_{poj}^2 \sum m_{(poj)k}^2}{n - 1}$$

donde:  $\bar{m} = m / n$   
 (poj)k

b) Remuneraciones, Gastos e Ingresos, en miles de pesos correspondientes a 1985.

Las fórmulas anteriores se volverán a usar en este caso, donde:

$$N = \hat{Y}\text{-rem}, \hat{Y}\text{-gas}, \hat{Y}\text{-ing}$$

para inferir sobre Remuneraciones, Gastos e Ingresos respectivamente.

luego:

$$A(i) \text{ para } A(i) \quad A(i) \quad \text{o} \quad A(i)$$

Rem                      Gas                      Ing

$a(i)_k$ : cantidad de remuneraciones (Rem), Gastos (Gas), e Ingresos (Ing) del k-ésimo AGEB (conglomerado), de la muestra que cae dentro del sector i (Cil).

$$m_k \text{ para } m(\text{rem})_k \quad m(\text{gas})_k \quad \text{o} \quad m(\text{ing})_k$$

n: número de AGEB en la muestra.

m(rem), m(gas), m(ing): cantidad total de la variable por extrapolar en la muestra.

Los resultados obtenidos se agrupan como sigue:

SECTOR	PERSONAL OCUPADO	VARIANZA
1	$\hat{A}(1)$ poj	$V[A(1)]$ ] poj
.	.	.
.	.	.
9	$\hat{A}(9)$ poj	$V[A(9)]$ ] poj

SECTOR	TOTAL REMUNERACION EN MILES	VARIANZA
1	$\hat{A}(1)$ rem	$V[A(1)]$ ] rem
.	.	.
.	.	.
9	$\hat{A}(9)$ rem	$V[A(9)]$ ] rem

SECTOR	TOTAL GASTOS EN MILES	VARIANZA
1	$\hat{A}(1)$ gas	$V(\hat{A}(1))$ gas
:		
:		
9	$\hat{A}(9)$ gas	$V(\hat{A}(9))$ gas

SECTOR	TOTAL INGRESOS EN MILES	VARIANZA
1	$\hat{A}(1)$ ing	$V(\hat{A}(1))$ ing
:		
:		
9	$\hat{A}(9)$ ing	$V(\hat{A}(9))$ ing

Una vez concluidas las estimaciones correspondientes a sectores, el siguiente paso consiste en efectuar con los totales obtenidos las ponderaciones para Subsector y Rama.

El procedimiento es considerar como una población a cada  $\hat{A}(i)$ :  $i = 1, 2, \dots, \dots, 9$ , dentro de éstas existirán tantos Dominios como Subsectores contenga el Sector y a su vez éste último se divide en clases de

acuerdo al número de Subsectores.

Como se indicó anteriormente, podemos representar gráficamente este caso como sigue:

POBLACION I	A(i) SECTOR i = 1,2,... 9			
DOMINIO	D1		D2 . . .	
	SUBSECTOR 1		SUBSECTOR 2 . . .	
CLASES	C1, . . . . Cr	C1, . . . . Cr		
		1		2
	Ramal . . . Rama	Ramal	Rama	
		r1		r2
No. de Unidades	a1, . . . a	a1, . . . a		
		r1		r2

donde el primer paso consiste en hacer las estimaciones correspondientes a Subsector (dominios en este caso) y posteriormente ponderar para Rama (clases que conforman cada dominio).

Para estimar Subsectores el proceso es similar al que se uso para subsectores, esto es, por cada sector se estimará la proporción de unidades que caen dentro de cada subsector empleando los esquemas de muestreo para proporciones para la variable establecimientos y

proporciones por conglomerados de distinto tamaño para el resto de las variables, tal y como se hizo para sectores.

A continuación se describe el procedimiento para establecimientos y para el resto de las variables se discutirá en forma general.

Para el sector (i);  $i = 1, 2, \dots, 9$

$$\hat{N} = A(i)_{est}$$

$C(j)$  = subsector (j)

$A(j)$ : número de establecimientos (unidades), de  $N$  que caen dentro de  $C(j)$

$a(j)$ : número de establecimientos de la muestra que caen dentro de  $C(j)$ .

$n$ : número de establecimientos en la muestra.

$p(j)$ : proporción de unidades procedentes de la muestra que caen en  $C(j)$ .

donde:  $p(j) = a(j)/n$

entonces:

$$\hat{A}(j) = N * p(j)$$

con varianza estimada:

$$V[\hat{A}(j)] = \frac{N - n}{N - 1} * \left\{ \frac{N^2 * p(j) * [1 - p(j)]}{n} \right\}$$

Para el resto de las variables, suponga que z representa cualquiera de las siguientes variables:

Poj, Rem, Gas, o Ing

entonces:

para cada uno de los nueve sectores:

$$N_z = \hat{A}(z)$$

C(j): es como se definió.

A(j): cantidad del tipo "z" de N que cae dentro de

C(j)

$a(j)$ : número o cantidad de "z" procedentes de la muestra que caen dentro de  $C(j)$  del k-ésimo AGEB.

$m(z)_k$ : número o cantidad de "z" dentro del k-ésimo AGEB.

$n$ : número de AGEB en la muestra.

$n_z$ : número o cantidad de "z" en la muestra.

Con la información anterior se procede a efectuar las ponderaciones correspondientes, para lo cual, se emplearán las fórmulas que fueron ocupadas en la sección (2.1.2), inciso "a" denotando a cada valor estimado como se muestra a continuación:

POBLACION	A(i); SECTOR i= 1, 2, ... .. 9		
SUBSECTOR	1	2	...
ESTABLECIMIENTOS:	$\hat{A}'(1)$ est	$\hat{A}'(2)$ est	...
PERSONAL OCUPADO: Poj; j=1..9	$\hat{A}'(1)$ Poj	$\hat{A}'(2)$ Poj	...
REMUNERACIONES (MILES)	$\hat{A}'(1)$ Rem	$\hat{A}'(2)$ Rem	...

GASTOS (MILES):             $\hat{A}'(1)$              $\hat{A}'(2)$             . . .  
    Gas                            Gas

INGRESOS (MILES):         $\hat{A}'(1)$              $\hat{A}'(2)$             . . .  
    Ing                            Ing

además, para cualquier  $\hat{A}'(j)$  del cuadro anterior,  $v(A(i): z)$  representa la varianza estimada para el j-ésimo subsector dentro del i-ésimo sector.

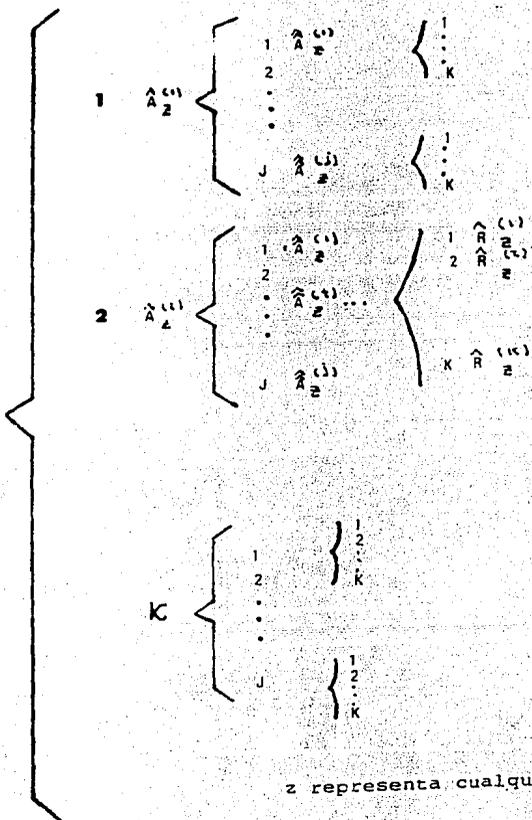
Por último se obtendrán las ponderaciones a nivel Rama dentro de cada subsector.

Para cuestiones de cálculo no hay nada nuevo al respecto con este último nivel, de tal suerte que se emplearán las mismas fórmulas y esquemas seguidos de los casos anteriores, donde la estimación del total para la k-ésima Rama del j-ésimo subsector está dada por:

$$\hat{R}(k, j) = \hat{A}'(j) * p(k)$$

Los totales  $\hat{R}(k, j)$  estimados por cada rama se presentarán con el mismo formato empleado en el cuadro anterior referente a subclases.

De igual forma se aplicará la fórmula para la varianza, ajustandose a la nueva nomenclatura.



z representa cualquiera de las siguientes variables

$P_{0j}, G_{12}, G_{30}, Y_{2j}$

3.- Numero de establecimientos. Personal ocupado. Remuneraciones. Gastos e Ingresos, por Entidad y Categoría Juridica.

Este caso, se refiere a la variable E03 de la pregunta 11 correspondiente a la sección III datos complementarios, donde la variable cae en cada una de las siguientes categorías.

- i) Propiedad
- ii) Sociedad Cooperativa.
- iii) Ejido.
- iv) Dependencia del Gobierno.
- v) Otro Tipo.

El procedimiento a seguir es el mismo que se ha venido empleando: los totales correspondientes a la variable establecimientos bajo el esquema de proporciones y porcentajes mientras, que para el resto de las variables se utilizara el muestreo por proporciones para unidades con conglomerados de distinto tamaño.

La población en este caso se encuentra dividida en cinco clases: C(i), C(ii), C(iii), C(iv) y C(v) para: propiedad individual, Sociedad Cooperativa, Ejido,

Dependencia del Gobierno y Otro Tipo respectivamente.

entonces para:

3.1) Número de establecimientos en cada clase sea:

$$N = Y:est$$

$A_{E03}$  = cantidad de establecimientos de la población que caen dentro de la  $i$ -ésima clave de la variable E03.

$a_{E03}$  = cantidad de establecimientos de la muestra que caen dentro de la  $i$ -ésima clase de la variable E03.

$n$  = número de establecimientos dentro de la muestra.

$p = a(E03)_i/n:est$  = Proporción de unidades de la clase  $C(i)$  que caen dentro de la muestra.

entonces un estimador para  $A(E03)$  es:

$$\hat{A}_{E03}(i):est = N * p$$

con varianza estimada:

$$v[\hat{A}(i)] = \frac{N - n}{N - 1} * \frac{N^2 * p * (1-p)}{n}$$

Para los casos referentes a Personal Ocupado, Gastos e Ingresos, considérese a "z" como se define en la sección anterior, y sea:

$$p = \frac{\sum_{a(i):z} E03(j)}{\sum_{m(j):z} E03}$$

donde el numerador representa la cantidad de elementos en el j-ésimo AGEB (conglomerado) de la muestra para la clase i. Y el denominador representa la cantidad de elementos del j-ésimo conglomerado.

Entonces sea:

$$N = \hat{Y}(z) \Rightarrow A(z)_i = N * p$$

donde  $\hat{Y}(z)$ , representa el total estimado referente a z que cae dentro de la i-ésima clase de acuerdo a la categoría jurídica (E03).

Dichos totales podrán representarse como se muestra en el cuadro de la siguiente página.

**CATEGORIA JURIDICA**

Clase(i)

Establecimientos: A(i):est  
Eo3

Personal Ocupado: A(i):poj  
Eo3

Remuneraciones: A(i):ram  
Eo3

Gastos: A(i):gas  
Eo3

Ingresos: A(i):ing  
Eo3

La siguiente división por inferir, es de acuerdo a Entidad Federativa, y tipo de propiedad (cuadro 4), para las siguientes clases:

- i) Privada
- ii) Pública Federal
- iii) Pública estatal
- iv) Pública Municipal
- v) Social
- vi) Otra

El procedimiento a seguir, es exactamente el mismo que se describió en la división anterior (por entidad y categoría jurídica), de acuerdo a lo señalado en la pregunta 12 y 13 sección III "Datos complementarios".

De esta forma queda descrita la estrategia para la inferencia de los resultados correspondientes a la fase rural dentro del programa Censos Económicos 1986.

## CONCLUSIONES

A lo largo del presente trabajo se ha expuesto toda la estrategia para llevar a cabo un proyecto de inferencia estadística, partiendo desde la Delimitación de la Población hasta el tratamiento adecuado de los datos que sirvieron como materia prima, todo esto, para concluir finalmente en la extrapolación de variables de interés dentro del presente estudio.

Cabe señalar que este documento se muestra únicamente en forma teórica, esto es, se trata de los lineamientos a seguir para la inferencia de los resultados en la muestra rural de los Censos Económicos 1986 y que éstos se consideran óptimos dentro de todas las restricciones físicas, humanas y económicas enfrentadas por el proyecto.

El objeto de este apartado es revisar y evaluar cada una de las partes involucradas en este documento.

La Delimitación de la Población contemplada en el Capítulo I, se concreta en definir la unidad de observación, el AGEB en este caso- así como la característica de estudio que es el personal ocupado dentro del AGEB, y enseguida se procede a establecer la estrategia para la definición de la Población ésta se basó en la depuración de los CIT de Población. Sin

embargo, el desconocimiento de la cantidad de personal ocupado por AGEB obligo a sustituir esta cantidad por la PEA-No-Agropecuaria dentro del AGEB misma que se calculó mediante proporciones de lo urbano contra lo rural, sin embargo este hecho no es preocupante, pues debe tenerse en cuenta que la PEA-No-Agropecuaria es una variable altamente correlacionada con personal ocupado, ya que este último puede considerarse como un subconjunto de la anterior y su utilidad radica en asignar el peso que el AGEB ocupa durante la selección, de tal suerte que éste será directamente proporcional a la cantidad de personal que trabaja dentro del AGEB rural; luego entonces los AGEB con escasa o nula actividad económica quedarían relegados a segundo plano, mientras que aquellos con cierta importancia tendrían más posibilidades de selección de acuerdo al esquema de Probabilidad Proporcional al Tamaño.

Como el método de selección se efectuó en forma sistemática, cabría la pregunta sobre la efectividad de este procedimiento en el sentido de que el interés primitivo es el personal ocupado pero la selección se efectuó sobre la PEA-No-Agropecuaria; para validar este esquema, conviene posteriormente con los datos de la muestra emplear una prueba No-Paramétrica (debido al desconocimiento de la función de distribución), dirigida a observar la correlación entre ambas variables y la consistencia de los datos, tal prueba bien podría ser el

## Coeficiente de Correlación de Rango de SPEARMAN.

En esta fase uno de los máximos problemas enfrentados fue con la PEA-No-Agropecuaria, en el sentido de que no hay distinción entre lo urbano y lo rural. Para enfrentar con efectividad esta situación fue necesario realizar toda una serie de ajustes de dicha cifra para todas aquellas AGEB que bien fueran envolventes y contenían localidades con carácter urbano o se trató de AGEB cuyos límites coincidían con localidades o AGEB urbanas puesto que, en ambos casos, era muy posible que se registrara una invasión a lo rural por parte de lo urbano, resultado del crecimiento de la mancha urbana; esto implicó realizar todos los cálculos descritos en Capítulo I, lo cual se tradujo un trabajo bastante arduo y monótono y que por sus características fue imposible realizarlo con la ayuda de un computador.

La obtención del tamaño de la muestra fue tratada a lo largo del Capítulo II, aquí los esquemas de muestreo MIA y PPT jugaron un papel importante dentro de un modelo de simulación donde se minimizaron los tamaños de muestra pero se mantuvo la eficacia,. El desconocimiento de la cantidad de personal ocupado dentro de cada AGEB y su alta correlación con la PEA-No-Agropecuaria dió como resultado emplear la técnica de

selección de "HANSEN y HURWITZ" .

Cabe señalar, sin embargo, que por restricciones de tiempo sólo se ocuparon 9 estados de 32 en el modelo de simulación; sin embargo el número de AGEB en cada estado era diferente y solo se trató de decidir el esquema de muestreo a elegir que resultara más apropiado, por lo cual el hecho de no haber empleado los 32 es irrelevante.

Otra situación que debe comentarse es que, para calcular los coeficientes de variación que son necesarios para poder obtener el tamaño de muestra, en 5 estados fue necesario elaborar un modelo de regresión por falta de información, dicho modelo confrontó número de AGEB contra coeficiente de variación.

Las pruebas F, así como la estrechez de los intervalos de confianza para los valores extrapolados, dieron plena confianza del empleo de los coeficientes de variación estimados.

Con respecto a la obtención del tamaño de la muestra, debe aclararse que la metodología planeada fue: selección sistemática proporcional al tamaño y sin reemplazo, sin embargo, el volumen de trabajo aunado a las restricciones de tiempo y personal obligaron a realizar la selección sin excluir aquellos AGEB con un

peso suficientemente grande, dando como resultado que, en algunos casos fuera imposible evitar el reemplazo, lo que ocasionó que el número de AGEB seleccionados fuera menor al tamaño de la muestra, sin embargo esto se soluciona tomándolos en las fórmulas para la inferencia del tamaño de la población tantas veces como fueron seleccionados.

Otro serio problema que se presentó, fue la reducción a los tamaños de muestra por motivos económicos-presupuestales; esto se trata en el capítulo III, en el cual se desarrolla una estrategia para tal efecto; sin embargo, el punto crítico radica en la reducción en la confianza para la muestra en las entidades afectadas.

El problema mayor se presenta en el Estado de Sinaloa, donde la confianza de la muestra, se redujo de un 90% a un 69.04%, le sigue Guerrero con un 70.09% y Sonora con 72.84%; para los estados restantes la confianza fue mayor estricta a 75% .

Esto debe tenerse en cuenta, pues cualquier descuido podrá repercutir considerablemente en la inferencia de los resultados; cabe recordar que la muestra se obtuvo en forma independiente para cada entidad del país. Luego entonces, la reducción en los

Peso suficientemente grande, dando como resultado que, en algunos casos fuera imposible evitar el reemplazo, lo que ocasionó que el número de AGEB seleccionados fuera menor al tamaño de la muestra. Sin embargo esto se soluciona tomándolos en las formulas para la inferencia del tamaño de la población tantas veces como fueron seleccionados.

Otro serio problema que se presentó, fue la reducción a los tamaños de muestra por motivos económicos-presupuestales: esto se trata en el capítulo III, en el cual se desarrolla una estrategia para tal efecto; sin embargo, el punto crítico radica en la reducción en la confianza para la muestra en las entidades afectadas.

El problema mayor se presenta en el Estado de Sinaloa, donde la confianza de la muestra, se redujo de un 90% a un 69.04%, le sigue Guerrero con un 70.09% y Sonora con 72.84%; para los estados restantes la confianza fue mayor estricta a 75% .

Esto debe tenerse en cuenta, pues cualquier descuido podrá repercutir considerablemente en la inferencia de los resultados; cabe recordar que la muestra se obtuvo en forma independiente para cada entidad del país. Luego entonces, la reducción en los

tamaños de muestra y por otra parte, el incremento a los tamaños de ésta en otras entidades puede dar pie a falta de homogeneidad en los resultados finales a nivel Nacional, puesto que por una parte la muestra se reduce, afectando la confianza de la misma, mientras que para otro grupo se aumenta el tamaño de la muestra, favoreciendo considerablemente la calidad de los resultados que dichas muestras arrojen y el resto se mantiene tal y como se planeó; por lo tanto, se puede detectar de este procedimiento la heterogeneidad en cuanto a los resultados, entonces solo se garantiza una confianza a nivel Nacional igual a la del menor estado.

Sin embargo el capítulo III realiza un esfuerzo en dos sentidos para cumplir con las restricciones presupuestales y simultáneamente optimizar hasta donde sea posible la confianza de la muestra con objeto de no afectar la calidad de la información obtenida.

El Capítulo IV "Factores que afectaron el Tamaño de la Muestra" se presenta como reporte de una serie de datos atípicos, presentados durante el operativo de campo y a la vez funciona también como un filtro para la información obtenida, pues es en este capítulo donde se reajustaron parámetros y pesos, se reducen en ciertos casos tamaños de muestra y se calcula la confianza de esta última a modo de que la información se encuentre

lista para para alimentar las formulas para la inferencia estadística contempladas en el Capitulo V "Inferencia de los Resultados".

Aunque el Capitulo V define una solución a cada caso, la falta de material cartográfico actualizado, así como los reportes actualizados de población, influyeron en cierta medida en unos cuantos casos en los cuales no fue posible ajustar por completo pesos y parámetros, por lo cual la única salida es suponer, aunque es posible, que no hay sesgo en la inferencia.

Esta conclusión, sin embargo no es del todo infundada, si se toma en cuenta que dadas las características de las zonas rurales cualquier posible desarrollo o crecimiento observado en éstas es lo suficientemente pequeño como para esperar incrementar considerablemente las cifras originales, de tal suerte que los ajustes que se efectúan de acuerdo a lo lineamientos establecidos en éste Capitulo servirán eficazmente en la etapa de la inferencia.

El total de establecimientos por entidad fue la primera variable a extrapolar, empleando la fórmula para totales bajo muestras con reemplazo; para el resto de las variables (excepto personal ocupado), como esta son captadas a su vez dentro de cada AGEB, obliga a emplear

como técnica de inferencia: proporciones y porcentajes dentro de conglomerados de distinto tamaño, sin embargo, es necesario tomar en cuenta que la fórmula para inferir totales mediante este sistema, requiere del conocimiento de  $M_0$ , total de unidades dentro de la población, es decir, el total de establecimientos en este caso, y esto significa que se extrapola una variable empleando como variable auxiliar otra variable previamente estimada y que además para el resto de las secciones acontece la misma situación.

Para la sección 2 las variables de interés se extrapolan de acuerdo a la clasificación Mexicana de Actividades y Productos (CMAP), por los métodos descritos anteriormente, de tal suerte que los datos obtenidos en la sección anterior son ahora empleados a manera de totales. Como la CMAP está compuesta por varios niveles de agregación entonces se ponderará el  $i$ -ésimo nivel ( $i=1$ ) tomando las ponderaciones obtenidas del Nivel ( $i-1$ ) a manera de totales.

Para el resto de las secciones se obtienen resultados apoyándose en la sección inicial tal y como se ha descrito.

La cantidad de variables por extrapolar y la relación de estas (esto es, unas son subconjunto de

otras), obliga a concebir como estrategia lo descrito anteriormente, pues tratar de extrapolar de manera aislada cada variable significaría una carga de trabajo que aparte de excesiva queda por completo fuera de contexto.

Por lo demás solo resta discutir algunas actividades relacionadas con la elaboración del trabajo delineado en cada Capítulo y con el Operativo de Campo.

Por principio debe tenerse presente en todo momento que se está trabajando en un estudio a nivel Nacional lo que desde luego implica grandes cantidades de información y enormes gastos, que además se trata de Areas Rurales donde la información es escasa y en ocasiones poco confiable, que en muchos casos el obtenerla implica bastante peligro, llegándose a registrar algunas veces graves consecuencias en la seguridad de los encuestadores con resultados muy desagradables. Y por último, es claro que la información solicitada es delicada y puede presentar sesgo.

Por otra parte referente al trabajo a nivel central cabe comentar lo siguiente:

Los ajustes mencionados en el primer Capítulo, dadas sus características resulta ser una labor que es

imposible realizarla via cómputo, lo cual se traduce en un proceso excesivamente lento y laborioso pues significa cotejar material cartográfico contra información proveniente de una serie de listados.

Los capítulos II y III en su primera parte no ofrecen tanto problema, pues éstos pueden llevarse a cabo via cómputo lo cual facilita ampliamente su desarrollo.

El Capítulo III en su segunda parte, no implicó grandes cargas de trabajo y se limita a unas cuantas operaciones aritméticas para la reducción y enseguida a seleccionar sistemáticamente los AGEB que se excluyen, esto puede realizarse sin apoyos de cómputo pues solo se efectúa en unas cuantas entidades, finalmente para el cálculo de una nueva confianza basta contar con una tabla de valores para la función Normal.

Quizá el trabajo más agobiante resulte en el Capítulo IV pues no solo engloba el efectuado en el Capítulo I. Sino que previo a ello, debe hacerse la detección de cada caso, auxiliándose para esto en todos los listados elaborados via cómputo y en algunos casos por información directa de las Direcciones Regionales y/o Coordinaciones Censales Estatales, y finalmente concluir con los ajustes correspondientes.

Finalmente, para realizar las inferencias propuestas en el Capitulo VI resulta inconcebible no contar con el apoyo de cómputo, para lo cual, es indispensable desarrollar los sistemas correspondientes que serán destinados para los fine que este capitulo persiguen e incluso cualquier posible modificación.

## APENDICE

El objetivo de este apéndice es presentar el desarrollo algebraico elaborado para la obtención de ciertos resultados a lo largo del documento.

No se profundizó en cada tópico que se aborda y sólo se concreta en desarrollar lo necesario para satisfacer cualquier justificación que sobre el tema en cuestión se pretenda.

1.- Obtención del modelo de regresión empleado para la estimación del coeficiente de variación en los estados de Chihuahua, Jalisco, México, Puebla y Veracruz.

El total de datos fue 24

donde:  $Y$  = Coeficiente de Variación.

$X$  = Número de AREB por entidad.

$$\begin{aligned}\sum Y(i) &= 56.64 & \sum Y(i)^2 &= 158.18 & \bar{Y} &= 2.36 \\ \sigma_Y &= 1.01\end{aligned}$$

$$\begin{aligned}\sum X(i) &= 10,843 & \sum X(i)^2 &= 7,665,709 & \bar{X} &= 451.79 \\ \sigma_X &= 339.54\end{aligned}$$

$$\begin{aligned}\sum X(i)*Y(i) &= 31,418.23 & r &= 0.708\end{aligned}$$

El modelo diseñado es de la forma:

$$Y = \beta_0 + \beta_1 X$$

donde:  $\hat{\beta}_0 = \frac{\sum X^2 \sum Y - \sum X \sum XY}{N \sum X^2 - (\sum X)^2}$

$$\hat{\beta}_1 = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2}$$

de donde:  $Y = 1.4083 + 0.0021 * X(i)$

Con los datos de la ANOVA se efectuó la prueba:

$$H_0: \hat{\beta}_1 = 0 \quad \text{v.s.} \quad H_a: \hat{\beta}_1 < 0$$

Con la regla de decisión, rechazar  $H_0: \hat{\beta}_1 = 0$  a un 95% de confianza si  $F$ -calculada es mayor que:

$$F(24; 1; .05) = 2.9270$$

dado que  $F_0 = 21.94 \Rightarrow F_0 > F(24; 1; .05)$ , por lo cual se rechaza  $H_0: \hat{\beta}_1 = 0$ .

En seguida se realizan los cálculos de las varianzas para  $\hat{\beta}_0$  y  $\hat{\beta}_1$ , así como sus intervalos de confianza.

-Para  $\hat{\beta}_0$  su varianza es:

$$\hat{\sigma}_{\hat{\beta}_0}^2 = \frac{\hat{\sigma}_e^2 \sum X_i^2}{N (\sum X_i^2 - N \bar{X}^2)} = 0.064$$

-Para  $\hat{\beta}_1$  su varianza es:

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{\hat{\sigma}_e^2}{\sum X_i^2 - N \bar{X}^2} = 2.015 \times 10^{-7}$$

donde :

$$\hat{\sigma}_e = \frac{SCY - \hat{\beta}_1 \sum XY}{n - 2} = 0.5577$$

Por lo tanto los intervalos de confianza al 95 %

son:

$$L = \hat{\beta}_i \pm \hat{\sigma}_{\hat{\beta}_i} t_{(\alpha/2)}^{(n-2)} ; t_{(0.025)}^{(22)} = 2.0739$$

$$\beta_0 \in (1.2756, 1.5410)$$

$$\beta_1 \in (-0.0021, 21.004 \text{ E } -4)$$

Para los CV extrapolados la fórmula para el error estandar de  $\hat{Y}(i)$  es:

$$\hat{\sigma}_{\hat{Y}_i} = \hat{\sigma}_e \left( \frac{1}{N} + 1 + \frac{X_i - \bar{X}^2}{\sum X_i^2 - N \bar{X}^2} \right)$$

De donde los intervalos de confianza se obtienen aplicando la siguiente fórmula:

$$L = \hat{\psi}_i \pm \int_{(\alpha)}^{(m-2)} \hat{\sigma}_{\hat{\psi}_i}$$

## BIBLIOGRAFIA

- Calero Vinedo, Aristides. Teoría de la estimación.  
Ministerio de Educación Superior. Cuba, 1981.
- Cochram, William G. Técnicas de muestreo. C.E.C.S.A.  
México, 1975.
- Des Raj. Teoría del muestreo. Fondo de Cultura  
Económica. México, 1980.
- INEGI. Lineamiento para la adecuación del  
marco geoestadístico. Manual. México  
1984.
- INEGI. Metodología de trabajo para la delimitación  
de áreas geoestadísticas. Manual. México 1978.
- INEGI. Clasificación mexicana de actividades y pro-  
ductos. México, 1985.
- Leslie Kish. Muestreo de encuestas. Trillas.  
México, 1982.
- Pandurang V. Sukhatme. Teoría de encuestas por muestreo.  
Fondo de Cultura Económica. México 1956.

SPP. Catálogos de población total por entidad federativa municipio y localidad. México, 1980.

SPP. Cobertura de cabeceras municipales y localidades de 2500 habitantes y más por entidad federativa. Mimeo. México, 1980.