

050 15
4
29

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

**UNIDAD ACADÉMICA DE LOS CICLOS PROFESIONALES
Y DE POSGRADO DEL C.C.H.**

INSTITUTO DE INVESTIGACION EN MATEMÁTICAS APLICADAS Y SISTEMAS

**"SISTEMA COMPUTACIONAL DE ACCESO Y PREPARACION DE LOS
DATOS DEL DISCO COMPACTO CODICE 90 PARA SU ANALISIS"**

**OPCION TERMINAL MODALIDAD CONFERENCIA PARA OBTENER EL DIPLOMA DE
ESPECIALIZACION EN ESTADISTICA APLICADA**

PRESENTA: IRENE FETTER NATHANSKY

MEXICO 1993

**TESIS CON
FALLA DE ORIGEN**



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 esta 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.

SISTEMA COMPUTACIONAL DE ACCESO Y PREPARACION DE LOS DATOS
DEL DISCO COMPACTO CODICE 90 PARA SU ANALISIS

RESUMEN

El código 90 es un disco compacto elaborado por INEGI que contiene la información recabada en el censo de población y vivienda de 1990 en la República Mexicana.

La información se presenta a manera de tablas y se puede consultar a nivel nacional, por estado, por municipio (delegación en el caso del Distrito Federal) o por área geoestadística básica (AGEB)

El presente trabajo se refiere al Distrito Federal exclusivamente.

El sistema que se elaboró permite obtener archivos de datos por AGEB.

El interés principal de la creación de estos archivos es la posibilidad de emplearlos en diversos ejercicios estadísticos (tales como análisis observacionales, técnicas de muestreo, etc.), una vez que estos se han validado, es decir, verificar la concordancia entre los datos presentados a diferentes niveles de agregación.

El presente trabajo cubrió dos fases. La primera consistió en la elaboración de bases de datos a partir de las tablas por AGEB presentadas en el código 90, de manera que:

a) la clave de identificación por AGEB entre localidades de delegación fuera diferenciada (se agregó la identificación por localidad y delegación por AGEB)

b) no hubiese elementos extraños, tales como títulos, subtítulos, totales, subtotales.

La segunda fase de este trabajo consistió en validar la información aportada por el código 90; pasando de un nivel de agregación mayor a un nivel menor de agregación, se intentó reconstruir la información de los tabulados básicos, que están a nivel delegacional, a partir de la información que se presenta a nivel AGEB.

Por estar más familiarizado con su manejo, el problema se trabajó con el lenguaje de programación de Clipper 5.01 y Dbase IV.

Se realizaron 32 programas. Uno de diagnóstico (DUPLI), para mostrar las claves de AGEB repetidas, uno de identificación (BORRA), para crear un código específico por delegación y localidad y AGEB y borrar los elementos extraños, otro de unificación (UNINOA3), para reunir todas los grupos de variables de población(8) y vivienda(4) en un solo archivo,

uno de concentración de información en datos por AGEB (TOT), para globalizar la información por delegación, y veintiocho de transformación de datos, para combinar y reagrupar las variables de los tabulados básicos para hacerlas coincidir con la información por AGEB.

El producto final es una base de datos de 1972 registros y 71 variables, en la que cada registro representa un AGEB del Distrito Federal, con su respectivo código para identificar la localidad y delegación a la que pertenece.

Cabe destacar que en el procedimiento de la validación se encontraron dos problemas.

1. No toda la información que se presenta en los tabulados básicos puede obtenerse a partir de los datos por AGEB, existen múltiples variables que no se encuentran registradas a nivel AGEB. Además la forma de clasificación y el número de niveles para las variables es diferente, observándose mayor detalle en los tabulados básicos.

2. Al totalizar los valores que se obtienen a nivel AGEB por delegación no se obtienen los mismos resultados que los presentados en los tabulados básicos. Esto se debe a la presencia de '*' y '-' en los tabulados originales, que al exportarlos se transforman en '-1' y '0' respectivamente. Al consultar la documentación que acompaña al código 90, se deduce que los guiones son equivalentes a ceros y

posiblemente los asteriscos se refieran a la información confidencial, es decir, en aquellos casos en los que había menos de 3 viviendas, para que el usuario no pueda identificar a quien se refiere la información recopilada se sustituyó por alguno de esos símbolos, aunque no se indica cual de ellos. Para los demás estados de la República se proporciona una cifra global para todas aquellas localidades que tengan menos de 3 viviendas para cada una de las variables, para el Distrito Federal, no.

Para constituir el archivo maestro de trabajo es necesario sustituir esos valores. Consultando a personal especializado de INEGI, tal vez se podrá recuperar dicha información, de otra manera habría que hacer una especie de prorrateo, para distribuir la cantidad faltante para cada variable entre los registros que presenten un valor de -1.

El procedimiento descrito pudiera utilizarse para todos los estados de la República Mexicana y finalmente obtener un archivo maestro para México.

ELABORACION DE UN ARCHIVO MAESTRO A NIVEL NACIONAL

ENTIDAD	NIVEL DE DESAGREGACION		
D.F.	AGEB	LOCALIDAD	DELEGACION
ESTADO 2	AGEB	LOCALIDAD	MUNICIPIO
ESTADO 3	AGEB	LOCALIDAD	MUNICIPIO
.			
.			
.			
ESTADO 32	AGEB	LOCALIDAD	MUNICIPIO

Con el archivo maestro, que ya se mencionó anteriormente se pretende producir material relativo al ensayo de esquemas de muestreo y otros análisis estadísticos para hacer inferencias a nivel nacional.

Vo. Bo.



M. en C. Rafael Madrid Rios

SISTEMA COMPUTACIONAL DE ACCESO Y PREPARACION DE LOS DATOS
DEL DISCO COMPACTO CODICE 90 PARA SU ANALISIS

OBJETIVO GENERAL: Elaborar un instrumento que permita acceder y preparar la información del censo de 1990 de población y vivienda que se encuentra en el disco compacto Códice 90 para su posterior análisis.

INTRODUCCION: La información que se recopiló en 1990 durante el censo de población y vivienda para la República Mexicana se encuentra en el disco compacto Códice 90 elaborado por el INEGI. El Códice 90 es un sistema de consulta que opera a través de diferentes ventanillas de selección. Estas incluyen información general sobre el levantamiento del censo, la metodología empleada, el cuestionario censal, la codificación de las variables, así como breves explicaciones sobre el propio manejo del sistema. Los datos arrojados por el censo se pueden consultar de dos formas: una, a través de la cartografía, localizando geográficamente la entidad que interesa; o bien de manera directa, identificando el Estado de interés y el nivel de desagregación de la información que se desea.

Para fines de este trabajo se decidió usar esta segunda posibilidad y trabajar con los datos del Distrito Federal

exclusivamente. El equipo que se empleó fue una computadora PC (Olivetti) 386 de 33 Mhz, con memoria RAM de 6 Megabytes, y se utilizaron aproximadamente 20 Megabytes de espacio en disco duro para el desarrollo del trabajo global. La información para el D.F. se presenta a dos niveles de agregación : a nivel de AGEB (Area Geoestadística Básica) y a nivel delegacional.

Este sistema ofrece la posibilidad de exportar archivos con el fin de manipularlos posteriormente de acuerdo a las necesidades del usuario.

Se manejan dos formatos de exportación:

DIF: Data Interchange Format, en el que cada registro es una fila de una hoja de cálculo y cada campo representa una columna;

SDF: System Data Format, es un archivo de texto de longitud fija con un retorno de carro manual al final de cada registro y que emplea un delimitador de campos.

Para crear archivos de trabajo que permitan la posterior aplicación de paquetes estadísticos se pensó en la utilización de bases de datos. Por razones de familiaridad se eligió el lenguaje de programación Clipper, que es un manejador de bases de datos, que además posee la característica de poder trabajar con archivos SDF.

La idea principal que se encuentra detrás de este trabajo es explorar las posibilidades para utilizar los datos del censo de 1990 como marco muestral. El censo contiene diversos tipos de variables, categóricas y numéricas, y ofrece la oportunidad de ensayar diversos tipos de muestreo para la estimación de parámetros y compararlos con los valores poblacionales reales.

METODOLOGIA

El trabajo consta de dos partes:

La primera se refiere a obtener una base de datos a nivel AGEB que se pueda utilizar (después de validarla) para realizar diversos ejercicios estadísticos.

La segunda tiene como finalidad validar la base de datos anterior corroborando si es posible obtener con esta información los mismos datos que aparecen en los tabulados por delegación que se incluyen en el Códice 90 y se denominan como tabulados básicos.

Primera parte:

Se inició con la información por AGEB. Los datos están divididos en dos grandes rubros: Población y Vivienda. Los datos de Población abarcan 8 grupos de variables:

DISTRIBUCION ESPACIAL

(8 VARIABLES)

MIGRACION

(7 VARIABLES)*

RELIGION Y ALFABETISMO

(8 VARIABLES)*

NIVEL DE INSTRUCCION

(8 VARIABLES)*

ESTADO CIVIL Y FECUNDIDAD

(6 VARIABLES)*

CONDICION DE ACTIVIDAD

(5 VARIABLES)*

SITUACION EN EL TRABAJO

(7 VARIABLES)*

RANGOS DE INGRESO.

(6 VARIABLES)*

Los datos de vivienda se refieren a 4 grupos de variables:

TECHOS, PAREDES Y PISOS

(7 VARIABLES)

NUMERO DE CUARTOS Y DORMITORIOS

(7 VARIABLES)**

GAS, DRENAJE Y ELECTRICIDAD

(6 VARIABLES)**

DISPONIBILIDAD DE AGUA Y TIPO DE TENENCIA

(6 VARIABLES)**

TOTAL

(81 VARIABLES)

* La primera de las variables de estos cuadros es repetitiva, es la misma que la correspondiente a la variable 1 de la distribución espacial.

** La primera de las variables de estos cuadros es la misma que la correspondiente a la variable 1 del cuadro de techos, paredes y pisos. Por lo tanto el número total de variables se reducé a 71.

Cada uno de estos grupos de variables se encuentra representado en un cuadro. La información se presenta de la manera siguiente:

A lo largo de la tabla:

- primero aparece un total por el Distrito Federal
- en segundo término un total de las AGEB urbanas del D.F.
- en tercer lugar un total de las AGEB rurales del D.F.
- en cuarto sitio se encuentra un total por Delegación
- en quinto lugar un total por AGEB urbanas de la delegación respectiva
- en sexta posición viene el total de la primera localidad
- el séptimo nivel de desagregación es ya por cada AGEB
- si hay más localidades, aparece primero un total por localidad y luego su desglose para cada AGEB que le corresponde
- en último sitio se proporciona un total por AGEB rurales de la delegación respectiva.

A lo ancho de la tabla se presentan las variables de interés correspondientes.

Este esquema se repite a partir de la posición cuatro para cada delegación.

De acuerdo a las indicaciones del sistema para exportar tablas de datos, primero se posiciona uno dentro de la tabla, luego con "Control P" aparece una pantalla, que pregunta si se desea copiar un bloque o toda la tabla, se señala toda la tabla. En seguida pregunta por el tipo de formato a emplear para su exportación: SDF o DIF. Se marca el SDF. Después pregunta por la unidad de disco: puede ser

al disco duro a bien a los discos flexibles. Finalmente se le asigna un nombre al archivo de salida. El sistema automáticamente le asigna la terminación SDF. De esta manera se exportaron los 12 cuadros de la información por AGEB. Para facilitar su posterior manejo se les asignaron nombres nemotécnicos. Así por ejemplo al archivo de distribución espacial se le identificó como APD (A porque proviene de información de los cuadros por AGEB, P porque corresponde al grupo de cuadros de POBLACION, y D por DISTRIBUCION ESPACIAL).

APD AGEB POBLACION DISTRIBUCION ESPACIAL

APM AGEB POBLACION MIGRACION

APR AGEB POBLACION RELIGION Y ALFABETISMO

APN AGEB POBLACION NIVEL INSTRUCCION

APE AGEB POBLACION ESTADO CIVIL Y FECUNDACION

APA AGEB POBLACION ACTIVIDAD

APT AGEB POBLACION TRABAJO

API AGEB POBLACION INGRESO

AVT AGEB VIVIENDA TECHOS, PAREDES Y PISOS

AVC AGEB VIVIENDA CUARTOS Y DORMITORIOS

AVG AGEB VIVIENDA GAS, DRENAJE Y ELECTRICIDAD

AVD AGEB VIVIENDA DISPONIBILIDAD AGUA Y TIPO DE TENENCIA

Los archivos con terminación SDF pueden verse con cualquier editor de textos, y esto permite observar la cantidad de variables que componen cada tabla, así como su tipo

(numérica o de carácter) y su longitud. El Códice 90 incluye un apartado de documentación de variables, que se puede mandar imprimir o bien guardar en un archivo.

A partir de esta información se elaboraron una serie de tablas en el paquete PMAKER (CENSO.PM4) (ANEXO 1) para documentar de manera detallada las variables, indicando el nombre del cuadro a que pertenecen, su identificación nemotécnica y su descripción.

Esta información se concentró en un cuadro resumen denominado (RELAC.DBF).

El siguiente paso consistió en crear las estructuras de las bases de datos a partir de la información contenida en RELAC.DBF. Para favorecer su manejo, las variables se nombraron con las mismas iniciales que la base original y numerándolas de manera sucesiva. Así, por ejemplo, a partir del archivo APD.SDF se creó una base de datos llamada APD.DBF, y las ocho variables que la conforman son APD1, APD2, APD3, APD4, APD5, APD6, APD7 y APD8.

Una vez creada la estructura de la base de datos, se procede a introducirle los datos.

Para vaciar la información de los archivos SDF a las bases de datos correspondientes, lo único que se requiere hacer es

renombrarlas con la terminación TXT, así el archivo APD.SDF, se renombró como APD.TXT.

DBASE permite vaciar la información del archivo APD.TXT, indicándole nada más que los diferentes campos están delimitados con comas, con la orden: APPEND FROM APD DELIMITED.

Al exportar esta información se presentan dos problemas:

- el primer problema radica en que no se incluye una identificación de delegación y localidad para cada registro; estos solamente aparecen al inicio de la información de cada delegación; es decir, en la sección donde vienen enumerados los datos de cada AGEB, no se incluye la identificación de la delegación ni la localidad a que pertenece. La clave de identificación de cada AGEB no es única, se repite a lo largo del archivo. (002-6: Milpa Alta y Benito Juárez; 002-7: Gustavo A. Madero y Alvaro Obregón, etc.) Esto se descubrió aplicando un programa llamado DUPLI.PRG, que es un programa que identifica registros duplicados, y al comparar la bases de datos original con la base de datos que se obtiene después de correr este programa, el tamaño es diferente. Esto indica que la clave de AGEB no se puede usar de manera exclusiva como identificador de un registro, necesita agregarse la identificación de la delegación y la localidad.

Se creó una base de datos, (DELEGACIONES.DBF) para identificar el número de delegación, las localidades con que cuenta cada una, el número de AGEB y si cuentan o no con AGEB rurales. Cabe señalar que los datos de las AGEB rurales vienen de manera global, no desglosada.

- el segundo problema se refiere a que los archivos creados incluyen letreros del título de las tablas, los marcos de los cuadros, los totales y subtotales, lo cual dificulta su manejo, ya que hay información duplicada y elementos extraños (no numéricos)

Para resolver el primer problema, se creó una base de datos temporal (AAPD.DBF), que contiene la información tal como el archivo original, y posteriormente se transcribió a la base definitiva (APD.DBF) que incluye campos adicionales que permiten identificar la delegación y la localidad. Las delegaciones se codificaron con una variable numérica progresiva del 2 al 17, respetando el orden en que se presentan en los diversos archivos, ya que siempre es el

mismo. Así los identificadores que se usaron son los siguientes:

AZCAPOTZALCO	2
COAYOACAN	3
CUAJIMALPA	4
GUSTAVO A. MADERO	5
IZTACALCO	6
IZTAPALAPA	7
MAGDALENA CONTRERAS	8
MILPA ALTA	9
ALVARO OBREGON	10
TLAHUAC	11
TLALPAN	12
XOCHIMILCO	13
BENITO JUAREZ	14
CUAUHTEMOC	15
MIGUEL HIDALGO	16
VENUSTIANO CARRANZA	17

En relación al segundo problema se creó un programa (BORRA.PRG) que a partir de la observación de la regularidad en la aparición de líneas, letreros, totales y subtotales, borra esta información superflua de los archivos y le asigna a cada registro la identificación de la delegación y

localidad a que corresponde, conservando el código de cada AGEB.

Este procedimiento se siguió para cada un de los cuadros por AGEB (12).

Cabe señalar que, desde el momento en que la información se exporta en formato SDF, se pierden los títulos de los cuadros originales.

El siguiente paso consistió en crear una base de datos que reuniera las variables de todos los subgrupos en un solo archivo.

Para realizar esto se elaboró un programa (UNINOA3.PRG) que permitiera unir los diferentes archivos, tomando como campo llave la delegación y la clave de la AGEB. Se tomó como archivo de referencia APD, por contener lo datos de población y no tener información faltante. El programa consiste en, localizar una AGEB específica de determinada delegación en este archivo y buscar para esta misma AGEB, en cada uno de los archivos, (APM, APR, APN, APE, APA, APT, API, AVT, AVC, AVG, AVD) el valor correspondiente a cada una de las 71 variables y sustituirlo en la base de datos UNIAGEB. No se presentaron problemas en esta fase, todos los archivos contenían el mismo número de registros y la clave llave coincidió para todos. El resultado es una base llamada

UNIAGEB.DBF que consta de 71 variables y 1972 registros (1956 AGEB urbanas y 16 registros corresponden a la información de las AGEB rurales). Esta base, de validarse, puede considerarse como la base de datos maestra, a partir de la cual se pueden ensayar estudios para muestras de diferente tamaño aplicando diferentes esquemas de muestreo.

Posteriormente, a través del programa TOT.PRG se concentró esta información, sumando los valores de cada variable por delegación, obteniéndose la base TOTAGEB.DBF, que consta de 16 registros y 71 variables.

SEGUNDA PARTE:

En segundo término se trabajó con los datos agrupados por delegación, que se denominan tabulados básicos. En estos cuadros se presenta la información de manera concentrada, a nivel delegacional y con un total estatal. El Códice 90 presenta 49 tabulados básicos, divididos en 10 grandes rubros o subgrupos de variables:

DISTRIBUCION ESPACIAL	3 TABLAS
MIGRACION	2 TABLAS
LENGUA INDIGENA	4 TABLAS
RELIGION	1 TABLA
NIVEL DE INSTRUCCION	8 TABLAS
ESTADO CIVIL	1 TABLA
FECUNDIDAD Y MORTALIDAD	6 TABLAS
CARACTERISTICAS ECONOMICAS	11 TABLAS
CARACTERISTICAS DE LA VIVIENDA	12 TABLAS
HOGARES	1 TABLA

La finalidad de trabajar con estas bases de datos es conformar una base que contenga la información poblacional, para delegación y Distrito Federal, es decir, contendría la información del universo de trabajo. Por ejemplo, en ejercicios posteriores, a partir de diferentes técnicas de muestreo, se puede seleccionar una parte de este universo, empleando el archivo UNIAGEB.DBF y se estima un parámetro, y posteriormente se compara con los datos incluidos en los tabulados básicos.

La forma de trabajar las bases de datos respectivas fue muy similar a la descrita anteriormente. Los archivos se exportaron en formato SDF y se renombraron a TXT.

El nombre de los archivos consta de: P o V para la primera letra, según si es variable de población o de vivienda,

seguido de la primera letra del nombre del archivo original, por ejemplo D para Distribución Espacial; y posteriormente el número del tabulado básico, desde el 1 al 49. Las variables de cada base de datos se identificaron desde V1 hasta Vn, donde n corresponde al número de columnas que contiene cada cuadro.

PD1-PD3	BASICO DISTRIBUCION ESPACIAL
PM4-PM5	BASICO MIGRACION
PL6-PL9	BASICO LENGUA INDIGENA
PR10	BASICO RELIGION
PN11-PN18	BASICO NIVEL INSTRUCCION
PE19	BASICO ESTADO CIVIL
PF20-PF25	BASICO FECUNDIDAD Y MORTALIDAD
PC26-PC36	BASICO CARACTERISTICAS ECONOMICAS
VV37-VV48	BASICO VIVIENDA
VH49	BASICO HOGARES

Para el procedimiento de borrado de elementos superfluos no se pudo elaborar un programa único, ya que cada cuadro es de diferente tipo y longitud sin poder generalizarse el procedimiento; por lo que se elaboraron 28 programas. (PD2.PRG, PM4.PRG, PM5.PRG, PR10.PRG, PN11.PRG, PN12.PRG, PN13.PRG, PN14.PRG, PN16.PRG, PN17.PRG, PN18.PRG, PE19.PRG, PF20.PRG, PC26.PRG, PC28.PRG, PC30.PRG, PC32.PRG, PC32B.PRG, PC33.PRG, PC36.PRG, VV37.PRG, VV39.PRG, VV41.PRG, VV42.PRG, VV43.PRG, VV45.PRG, VV46.PRG Y VV48.PRG). Para validar los datos que se proporcionan a nivel AGEb con la información

contenida en los tabulados básicos, se procedió a comprobar si realmente se podía llegar a estos tabulados básicos sumando los datos a nivel AGEB. Para esto fue necesario identificar la correspondencia entre cada una de las variables de los tabulados básicos y las variables que se describen en las tablas por AGEB. (RELACION.DBF).

En este punto se encontró un problema:

- el problema consiste en que no todas las variables que se incluyen en los tabulados básicos, aparecen en las tablas por AGEB. Por ejemplo, los tabulados básicos presentan información acerca de lengua indígena, y en los cuadros por AGEB esta variable no se encuentra. Por otro lado el nivel de desagregación de las variables de los tabulados básicos es mayor que la indicada en los cuadros por AGEB; es decir en los tabulados básicos la información se desglosa por sexo, edad, etc., mientras que las tablas por AGEB se presenta una sola cifra global.

Por ejemplo, mientras que en los tabulados básico se proporciona información del estado civil por grupos quinquenales de edad, para hombres y mujeres, detallando si son solteros, casados por el civil, por lo religioso, o por ambos, "en unión libre separados, divorciados y viudos; a nivel AGEB solo se proporciona el número de personas de la población mayor de 12 años casada y soltera.

Por estas razones se decidió, proceder de manera contraria, es decir a partir de los datos de los cuadros por AGEB, tratar de identificar su ubicación en los tabulados básicos.

El procedimiento que se siguió fue elaborar una relación para cada variable de los cuadros por AGEB, identificando en que tabulado básico se presenta, registrando el renglón y columna que le corresponden. En algunos casos, la correspondencia no es tan directa, es necesario realizar sumas o restas de diversas columnas o renglones; por ejemplo la variable APR3, o sea, la tercera variable de la tabla de AGEB, de Población, del tema de Religión, que corresponde a : Población de 5 años y más no católica, necesita obtenerse a partir del tabulado básico PR10, que se refiere a Población, Religión, cuadro No. 10, y se requiere sumar los datos de las columnas 3, 4, 5 y 7, que se refieren a protestante, judaica, otras y no especificada.

La obtención de estos valores se realizó a partir de los programas ya mencionadas, a través de instrucciones sencillas, indicando en cada caso la operación necesaria. En estos programas también se incluyen las instrucciones para eliminar los elementos extraños. De esta manera se crearon 28 bases de datos.

Posteriormente esta información se concentró en una sola base de datos denominada TOTBAS.DBF

A continuación se comparó la información de esta base de datos, con la obtenida en UNIAGEB.DBF, en donde TOTBAS.DBF representa la información contenida en los tabulados básicos, y UNIAGEB. DBF la aportada a partir de la suma de los datos por cada AGEB, de cada localidad de cada delegación.

RESULTADOS

En la mayoría de los casos se encontraron discrepancias al comparar los datos provenientes de la suma de las AGEB, y los que aparecen en los tabulados básicos.

Se hicieron varias comparaciones. Primero se compararon las sumas de los datos por AGEB (TOTAGEB.DBF) con la información de totales por delegación que se incluye en las mismas tablas por AGEB. No coincidieron.

Posteriormente se compararon los datos obtenidos por suma de las AGEB con los contenidos en los tabulados básicos. No coincidieron.

Cabe comentar sin embargo, que los datos que se derivan de los tabulados básicos son idénticos a los que aparecen como totales en los cuadros por AGEB.

De aquí se concluye que la información que se presenta en los tabulados básicos no se obtuvo a partir de la suma de la información por AGEB. De igual manera, los totales que aparecen en los cuadros por AGEB parecen haber sido obtenidos de una fuente diferente, pues no corresponden a la suma de los valores contenidos en dicha tabla. Esto se debe a que en los archivos por AGEB aparecen en todas las variables una serie de valores -1 y 0, que al integrarse a las sumas modifican el resultado. Al consultar los archivos originales se observa que estos valores corresponden a '*' y '-' respectivamente. En ninguna parte del Códice 90 está documentada explícitamente el significado de estos símbolos. Se infiere que el guión representa al valor 0, ya que en ninguna parte se utilizan ceros; en las delegaciones que no tienen AGEB rurales, por ejemplo, se indica con un guión. El asterisco podría referirse a los casos de confidencialidad, es decir a los casos en que por tratarse de dos o una vivienda nada más, no se da a conocer el dato. En la documentación del sistema se hace referencia a este problema y se indica que en otra parte del sistema se encontrará una cifra global que agrupa todos los casos confidenciales para cada variable. Sin embargo, esto solo pudo ser comprobado

para los demás estados de la República, pero para el Distrito Federal no se encontró dicha información.

CONCLUSIONES

El código, 90 como sistema de consulta es de fácil manejo para diferentes usuarios, como estudiantes, profesionistas o investigadores. Es muy sencillo pasar de una ventanilla a otra, y localizar la información específica que uno necesita para un Estado en particular.

Hay algunos cuadros en los que la información incluida no corresponde estrictamente al título del cuadro, es decir, por ejemplo, que en la base de datos migración (APM), se incluyen datos de población,; en la tabla de rangos de ingreso⁴ (API), aparecen variables de situación del trabajo, etc.

El código 90 ofrece como una de sus posibilidades la exportación de los datos contenidos en él, para que el usuario busque sus propias aplicaciones a los datos del censo. A partir de lo realizado en este trabajo se concluye que difícilmente, una persona sin conocimientos a nivel intermedio de computación, será capaz de manejar los datos provenientes del código 90 para otros usos. Es necesario invertir mucho tiempo y trabajo para la correcta identificación de las variables. Hace falta documentar apropiadamente el significado de los '-1' y '*' para poderlos sustituir adecuadamente.

No todas las variables que se presentan en los tabulados básicos se encuentran en los cuadros por AGEB, y los niveles de agregación de las variables que se manejan en ambos son diferentes. Esto limita la selección de variables que se pueden tomar como referencia para los ejercicios estadísticos.

Para poder utilizar la base de datos por AGEB (UNIAGEB) como archivo de trabajo, hace falta sustituir los valores -1.

Una sugerencia para sustituir los valores de -1 es, suponer que los totales que se presentan en los tabulados básicos son correctos y prorratear la diferencia existente de los totales por AGEB entre el número de municipios con valor de -1.

Es importante señalar que todas las variables se manejan como conteos, es decir número de sujetos que cumplen con cierta condición; esto a su vez limita el tipo de ejercicios estadísticos que se pueden realizar con los datos del código 90.

La única variable del archivo que no presentó problemas de '*' y '- ', y en la que si coincidieron las sumas por AGEB con las presentadas en los tabulados básicos fue APD1, o sea, la referente a la población total.

Para poder obtener un archivo de trabajo con datos a nivel nacional, es necesario contar con la información de los demás estados que integran la República Mexicana. Para esto hace falta repetir este ejercicio para los demás estados, para comprobar si los procedimientos seguidos para el Distrito Federal son generalizables a otros estados o no. Quizás sea posible omitir algunos pasos y simplificar otros, debido a que en esos archivos si se incluye información de los casos confidenciales.