

21 AGO.
IC/M
Roub.

1972
Nº 5



ASAMBLEA GENERAL DE MIEMBROS del

00X-L	0058409
18 11 76	
ORIGINAL de DOCUMENTOS	
Original NO SALE de la oficina	

Sexta Asamblea General
Santiago, Chile
Noviembre 9-13, 1972

6446 Esp. 10/31/72-700

I-536

ALMACENAMIENTO Y RECUPERACION DE INFORMACION ESTADISTICA

Por René Peralta ~~Sac~~*
y Valdecir F. Lopes**

Contenido

	Página
I. LA PRODUCCION DE INFORMACION ESTADISTICA Y SU UTILIZACION POR DIFERENTES USUARIOS	1
II. ALGUNOS PROBLEMAS PARA LA CREACION Y DESARROLLO DE BANCOS DE DATOS	6
1. Funciones u objetivos y campo de aplicación de un banco de datos	6
2. Empleo de paquetes de programas estadísticos ..	13
3. Problemas que derivan de la confidencialidad de la información	17
4. Recursos necesarios para que un banco de datos sea operativo	19
III. DIFERENTES ENFOQUES QUE SE PUEDEN DAR AL ALMACENAMIENTO Y RECUPERACION DE INFORMACION	23
IV. EL USO DE TERMINALES CONVERSACIONALES	35
V. UNA EXPERIENCIA EN AMERICA LATINA: EL BANCO DE DATOS DE CELADE	42

* Jefe, Sector de Computación, Centro Latinoamericano de Demografía (CELADE), Santiago, Chile.

** Demógrafo, Centro Latinoamericano de Demografía (CELADE), Santiago, Chile.

Nota: Este documento fue preparado para la reunión técnica sobre "Almacenamiento y Recuperación de Información Estadística y Elaboración de Tabulaciones Especiales mediante Computadores."

7577



I. LA PRODUCCION DE INFORMACION ESTADISTICA Y SU UTILIZACION POR DIFERENTES USUARIOS

Cuando se trabaja con grandes volúmenes de datos, sea que estos provengan de censos, de encuestas o de estadísticas continuas, se plantean inmediatamente varias interrogantes. Algunas de ellas podrían ser:

1. Seremos capaces de prever todo lo que se puede obtener de los datos básicos, para fines de estudios y análisis por parte de los usuarios de la información?
2. Es realmente posible producir, en una sola etapa, toda la información que se necesita?
3. Atenderán los cuadros que se han diseñado a las necesidades de los diferentes tipos de usuarios?

No quisieramos seguir planteando dudas, sino que al contrario, tratar de desarrollar estos puntos, con miras a discutir algunas de las varias posibilidades de decisión que se puede considerar para la solución de estos problemas. Hay que tener presente, inicialmente, que aún cuando la elaboración o procesamiento de los datos estadísticos sea sólo una etapa relativamente pequeña de todo el proceso de obtención de información, representando muchas veces un pequeño porcentaje del costo global de la investigación, es a través de esta etapa que se llega al producto final deseado. Equivocaciones o empleo de procedimientos incorrectos o no adecuados,

durante la elaboración de la información, pueden conducir a desperdiciar el esfuerzo anterior, echando por tierra todo el trabajo realizado en las etapas precedentes.

Pasando a las interrogantes planteadas: seremos capaces de prever todo lo que se puede obtener de los datos básicos, para fines de estudios y análisis por parte de los principales usuarios de la información? Seguramente que no. A medida que aumenta la posibilidad de uso de la información estadística, para fines de planificación global o sectorial y para propósitos de análisis en los diferentes campos (económico, social, cultural, etc.), se hace más difícil determinar a priori cuales son los usos que se pueden hacer de los datos obtenidos a través de un censo o de una encuesta. Las mismas tabulaciones censales, en el campo demográfico, para sólo citar un ejemplo, se han revelado insuficientes para el estudio de ciertas características de la población, cuando se pretende ^{de} ~~abundar en estudio~~ mediante el empleo de técnicas o métodos más sofisticados. Lo mismo pasaría seguramente en relación a otras áreas de conocimiento, y es muy probable que los economistas, los sociólogos y los planificadores de la salud, al igual que los demógrafos, no se satisfagan con las tabulaciones que tradicionalmente se han producido para los censos de población.

Podría decirse que sería necesario contar con un equipo interdisciplinario de mucha experiencia para determinar todos los

usos que se puede o se debe hacer de la información producida. Y aunque eso fuese prácticamente posible, sería difícil prever todo lo que se puede objetivamente obtener de un censo o una encuesta, antes de contar con algunas informaciones básicas o preliminares. Al tabular las respuestas a una pregunta se llega muchas veces a la conclusión de que la calidad de los datos básicos no permite la obtención de cierto tipo de información, por diferentes motivos: el número de personas que contestó a la pregunta; el número elevado de "no responde"; etc. Por todo ello, muy frecuentemente es necesario producir un conjunto de cuadros básicos, o tabulaciones marginales, para poder establecer las posibilidades que estos mismos datos ofrecen.

En cuanto a la segunda pregunta - Es realmente posible producir, de una sola etapa, la información que se necesita? También diríamos que no. Aun en el caso que se pudiera establecer todas las necesidades de los distintos tipos de usuario y las posibilidades que la información ofrece, y se dispusiera de un gran sistema de computación, eso no sería factible, en la mayoría de los casos. La experiencia ha revelado que la demanda de datos primarios o básicos va creando otra demanda, de datos más refinados, y que esta última no puede ser atendida sino después de la primera.

Finalmente, se puede decir que la tercera interrogante ya está contestada, al contestarse a las dos primeras. Por más amplio

que sea un programa de tabulación, difícilmente podrá atender a los diferentes usos que los diversos usuarios darán a la información.

Las respuestas a estos tres puntos conducen obligadamente a la necesidad de hacer un sistema de almacenamiento y recuperación de la información de tal forma estructurado que permita sacar el máximo de provecho a la información básica disponible, a un costo razonable y con la debida oportunidad.

Dada la ayuda que para ese tipo de funciones ofrecen las computadoras, se han creado bancos de datos, sean sectoriales, nacionales, regionales o internacionales. Los objetivos y las funciones de estos bancos son muy diversas y, en muchos casos, están condicionados al tipo de información almacenada.

En los capítulos siguientes se tratará de especificar algunos de los enfoques que se puede dar a un banco de datos y algunos de los problemas más importantes en relación con la manera de operarlos.

Entre las muchas instituciones que se han dedicado a almacenar información estadística de interés para la región y a ofrecer sus servicios, se puede mencionar:

**"International Data Library and Reference Service",
University of California, Berkeley**

"Latin American Data Bank"
Center for Latin American Studies, University of
Florida. Gainesville, Florida

"Banco de Datos de CELADE"
Centro Latinoamericano de Demografía

"Data Bank and Information Systems for National Statistics"
Bureau of Labor Statistics, U.S. Department of Labor

"NEER Data Bank for Project Economics"
National Bureau of Economic Research

Se debe mencionar, además, aunque no se le ha dado el nombre de banco de datos, al "Bureau of the Census", del Departamento de Comercio de los Estados Unidos, institución que recopila y es depositaria del mayor volumen de datos numéricos en aquel país.

II. ALGUNOS PROBLEMAS PARA LA CREACION Y DESARROLLO DE BANCOS DE DATOS

"Un Banco de Datos y sistema de información para estadísticas nacionales será más útil cuando el usuario pueda acudir a este sistema con pocos requisitos adicionales a su propia disciplina profesional. Esto implica que los especialistas en procesamiento de datos deben entender este planteamiento y dar a los usuarios un idioma de computación basado en su vocabulario de uso corriente. Debemos ser como un técnico en mantención y reparación de teléfonos, reparar los instrumentos, mantener las líneas, reemplazar las plantas manuales por automáticas y aumentar la nitidez y fuerza de las señales, etc., pero nunca inmiscuirnos en la conversación". ^{1/}

Este punto de vista, que podríamos considerar como la meta a la que se pretende llegar con un banco de datos y el sistema de información que se desarrolla para hacerlo operativo, servirá como introducción para plantear una serie de problemas que se presentan cuando se desea crear y desarrollar un banco de datos.

1. Funciones u objetivos y campo de aplicación de un banco de datos

Se puede decir que el objetivo general de un banco de datos es permitir la recuperación y uso de la información acumulada, en

^{1/} "Data Banks and Information Systems for National Statistics".
R.C. Mendelsohn, Invited and Submitted Papers, IPIP Congress
71 Ljubljana.

forma expedita, para atender a las necesidades de diferentes clases de usuarios.

Dentro de este planteamiento general, se pueden considerar como funciones de un banco de datos:

- a) Atender simultáneamente a un cierto número de usuarios de la misma institución o de instituciones diferentes;
- b) Asegurar una actualización permanente de los archivos, a medida que se dispongan de nuevos datos;
- c) Permitir la pronta obtención de informaciones, reduciendo el tiempo entre el planteamiento de la necesidad y la recuperación de las mismas;
- d) Permitir el fácil acceso a los datos que se desean procesar o recuperar;
- e) Permitir el acceso de los usuarios al sistema, eliminando la necesidad de recurrir cada vez a un especialista en procesamiento de datos;
- f) Permitir la recuperación de información de diferentes niveles (microdato o datos básicos y macrodato o datos elaborados);
- g) Reducir los costos de operación para la recuperación de información;
- h) Facilitar la integración e interrelación de informaciones de diferentes fuentes o de distintas épocas.

Algunas de estas funciones pueden llegar a ser contrapuestas; por ejemplo, los puntos d) y g), ya que por un lado se impone prácticamente lo que se llama un banco de datos en línea, y por otra, se exige que el sistema opere a un costo razonable. Estas dos posiciones pueden ser incompatibles en la medida que el volumen de información disponible sea muy grande. El punto d) con el f) y c) también están en esta condición, ya que cuando el volumen de datos sea muy grande y se desee minimizar el tiempo de respuesta del sistema, es probable que sólo se pueda operar a nivel de macrodatos. Podría pensarse en otros objetivos que en cierta forma son incompatibles, no en la generalidad de los casos, sino que bajo ciertas condiciones muy específicas.

Vamos ahora el campo de aplicación de un banco de datos.

Distinguiremos cuatro casos, a saber:

- a) Sistemas de estadísticas continuas (o series en el tiempo);
- b) Sistemas de observación de larga periodicidad, con gran volumen de datos (censo, por ejemplo), para tratar en forma distinta y siempre en base a microdatos;
- c) Sistemas de consultas en base a macrodatos;
- d) Sistemas mixtos.

Para el primer caso (estadísticas continuas), es fundamental la función b), vale decir, su capacidad de actualización y/o definición de subarchivos, de tal forma que se pueda procesar independientemente uno o más subarchivos o el archivo total, si es necesario. Además, cobra mucha importancia el objetivo expresado en el ítem f), cual es poder definir niveles de datos en función del tipo de usuario al cual está destinado el resultado de la investigación. Este es un punto de mucha importancia y dice relación con el poder de síntesis que hay que buscar al entregar la información a distintos niveles de decisión.

Por otro lado, dependiendo de la actualidad que tenga este tipo de información podría cobrar importancia la posibilidad de atender simultáneamente a diferentes usuarios, esto es, tener un sistema de diseminación en línea. Ejemplos de este tipo son la obtención de informaciones de interés para la toma de decisión en relación a los recursos humanos disponibles en una zona, cuando en otra hay escasez de estos recursos. En la medida que la información no es oportuna, el costo real o social que esta falta de información puede acarrear pasa a ser mucho más importante que el costo de mantener un sistema en línea para entregar la información.

Si examinamos el segundo de los campos de aplicación enumerados, podemos dar como ejemplo los censos nacionales o regionales. Para estos casos no tienen prácticamente validez los objetivos

a), b), y d), por lo menos a nivel de microdatos. En cambio, puede tener cierta importancia el objetivo c), de minimizar el tiempo entre el planteamiento de necesidad de datos y la recuperación de los mismos. Bajo las mismas circunstancias, el punto e) no podrá llegar a cumplirse, pues se trata de un problema muy complejo como para eliminar la necesidad de que los técnicos participen en el procesamiento de datos. Sin embargo, si se piensa en el nivel de macrodato, cobra mucha importancia este objetivo. Ello implica, por otro lado, cumplir el objetivo f), cual sea el de operar con micro y macro datos.

Al trabajar con un censo, por ejemplo, se hace imprescindible cumplir el objetivo g), dado el volumen de las cifras involucradas en el procesamiento de los censos.

Un aspecto muy importante, en relación con bancos de datos, es el que se refiere a la privacidad y confidencialidad de los datos. Dentro de un mismo banco podrá haber datos que se puede utilizar sin restricciones y otros que son de uso restringido.

Finalmente, en cuanto a la posibilidad de interrelacionar o integrar diferentes datos, si se piensa en un censo en forma aislada, no cobraría mucha importancia. Pero basta que se incorpore como meta algún tipo de análisis entre resultados de dos o más censos sucesivos para una misma localidad, o la integración de dos o más encuestas para ir actualizando los macrodatos para que este

objetivo pase a tener una gran importancia. Tal como se explicará en mayor detalle en el capítulo V, este ha sido uno de los principales objetivos del Banco de Datos de CELADZ, ya que la reunión de información de diferentes países crea la necesidad de interrelacionar diferentes fuentes, e incluso uniformar criterios de codificación, a fin de buscar la necesaria comparabilidad de los datos.

Con referencia al tercero de los campos de aplicación mencionados para un banco de datos, esto es, sistemas de consultas sobre macrodatos, se puede indicar que son posibles gracias al desarrollo de sistemas de tiempo compartido, es decir, utilización simultánea de la unidad central de proceso de la computadora por varios usuarios, como si cada uno de ellos dispusiera de la máquina en forma total y exclusiva. Estos sistemas atienden a los objetivos a) y g), esto es, capacidad para soportar simultáneamente un cierto número de usuarios que se encuentran ubicados remotamente, operando a un costo razonable. Indirectamente se estarían cumpliendo los objetivos c), d) y e) ya que por la forma de trabajar usando terminales de consulta se logra minimizar el tiempo de respuesta sin que esto signifique necesariamente un aumento en el costo. Por otro lado, existen técnicas de operación en la instalación de la computadora que equivalen prácticamente a disponer de los archivos de macrodatos en línea, y además existen ya implementados sistemas de programación a través de terminales, cada vez más orientados hacia el usuario. Los objetivos b) y h), capacidad para actualizar los archivos

e interrelacionar información de diferentes fuentes, tendrían mayor o menor importancia, dependiendo del caso concreto de que se trate. En todo caso se puede decir que, en general, al trabajar con terminales, estas operaciones no simplifican enormemente en razón de las condiciones básicas que se han impuesto a los lenguajes de programación con que operan.

Con referencia al cuarto y último campo de aplicación considerado, no es nuestra intención entrar a detallar todas las posibles combinaciones que se pueden formar entre los tres campos ya descritos, sino que por el contrario, se los mencionamos con la finalidad de llamar la atención para el problema. Un banco de datos, salvo que su campo de acción sea muy restringido, debe contemplar la posibilidad de cubrir necesidades muy diversas, en las cuales pueden coexistir informaciones que cubran campos de naturaleza muy diversa. De igual modo, puede que el sistema de recuperación de información no sea uno solo, existiendo una filosofía general y varios subsistemas para permitir que se resuelvan en forma más eficiente y económica algunos problemas concretos. Esto sin perder la visión de conjunto y la posibilidad de relacionar los subsistemas. Por otro lado, parece que el plantear un supersistema que pudiera cumplir con todos los objetivos enumerados, y cubrir todos los posibles campos de aplicación, sería algo enteramente inoperante y antieconómico con el nivel actual de conocimiento de técnicas especiales de procesamiento de datos. Aun más, según la experiencia ha

demostrado, estos superistemas sólo funcionan en el papel.

Con respecto a este último punto, parece interesante insertar algunas experiencias del "Defense of the Caucasus" de los Estados Unidos (DUCZ), reflejadas en algunos de sus líneas de conducta generales, que son:

- El DUCZ no tiene la intención de depender de una sola computadora, cualquiera que sea la capacidad de esta para hacer sus trabajos.
- Los sistemas mayores de computación se diseñan para ser utilizados durante 10 años, pero hay una revisión de estos cada

5 años.

- La compatibilidad entre los sistemas llega a ser más y más importante, especialmente a medida que se agrandan los archivos de datos.

2. Empleo de paquetes de programas estadísticos

Como una forma de poder cumplir con algunos de los objetivos planteados como deseables para el sistema de información de un banco de datos, diferentes instituciones han emprendido el trabajo de diseñar sistemas que entrarían en lo que se ha definido como POL ("Programas Oriented Languages"), los que están orientados más hacia el usuario que a la máquina.

La mayoría de estos POL está orientada a cierto tipo de aplicaciones. Dependiendo de su estructura y compatibilidad, pueden ser considerados en forma general, como subsistemas de un sistema mayor que serviría de enlace a todos ellos.

En este capítulo sólo se mencionan algunos POL o subsistemas, y en especial a aquellos que son usados con mayor frecuencia para análisis estadísticos, y que hemos tomado la oportunidad de conocer y usar. $\frac{2}{-}$

Ellos son:

- SPSS "Statistical Package for the Social Sciences", desarrollado por la Universidad de Chicago.
- OGDINS/40 "Organized Set of Integrated Routines for Investigation with Statistics", desarrollado por la Universidad de Michigan.

Amos subistemas tienen características muy similares, a

seber:

a) La descripción del archivo que se quiere crear se

hace sólo una vez y queda incorporado al archivo propiamente tal,

en forma de diccionario;

b) Se ha estructurado el subsistema de tal modo que

el usuario, una vez que esté creado el archivo, pueda hacer análisis

2/

"Programas de Computación usados en el CEIADIE". Centro Latinoamericano de Demografía, agosto 1972, Serie J N° 1.

alta (90 Kbytes de memoria principal disponible) ;

f) Por el momento, estos subsistemas no aceptan consolas a través de terminales ;

g) Estos sistemas cuentan con una gran variedad de programas de análisis, como por ejemplo: frecuencias marginales, tabulaciones cruzadas, correlaciones, regresiones, análisis factorial, etc.

h) Con su empleo, se cumple plenamente la condición de minimizar el tiempo entre el planteamiento de la necesidad de datos y la recuperación de los mismos.

En contraposición a las ventajas y facilidades mencionadas, en general estos sistemas son relativamente menos eficientes que

de los datos, necesitando para ello una o dos instrucciones que son muy simples de codificar;

c) Cuando es pertinente, la salida de un programa sirve automáticamente como entrada para los otros programas de análisis. Esto quiere decir que se ha estructurado la compatibilidad entre programas que pueden usarse consecutivamente;

d) Ambos subsistemas tienen clara capacidad de elaboración de datos que se puede resumir en las siguientes funciones: recodificación; selección de casos; cálculo de medidas; definición de categorías faltantes; incorporación de casos; incorporación de variables;

e) En general necesitan una capacidad de computador

programas especialmente diseñados, lo que encarece su costo de operación. Esto está de acuerdo con lo que se mencionó antes, esto es, que en general es imposible que un subsistema pueda cumplir eficientemente muchos objetivos diferentes. En este caso concreto, la diferencia está en el aumento del costo de operación, pero hay que considerar que, por lo general, para un pequeño grupo de datos, es corriente que se consuma 5 a 10 veces el tiempo de proceso final en pruebas; luego, para este tipo de aplicaciones, podría decirse que resulta más económica la operación con el subsistema que empleando programas especialmente diseñados. Al no usar subsistemas se debe considerar, además, el costo del programador para hacer el programa y efectuar las correcciones.

Otro subsistema bastante más orientado al problema de tabulación de datos de censos y encuestas es CENTS ("Census Tabulation System"), desarrollado por el "Bureau of the Census" de los Estados Unidos.

CENTS es un subsistema que ha sido diseñado para elaborar tabulaciones de datos cruzados con hasta 5 variables de control, pudiendo especificarse niveles de totales, subtotales y algunos indicadores estadísticos como promedios, media, tasas, etc.

Algunas de las ventajas que ofrece ese subsistema son:

a) A pesar de ser un POL, por estar programado en lenguaje ensamblador se obtiene una gran eficiencia de proceso.

Como limitación de este subistema puede decirse que no cuentan con facilidades de elaboración o manipulación de los datos que van a ser tabulados, y es evidente que esta es una fase que no se debe descuidar. Como su aplicación principal es para tabular conjuntos de datos que tienen un volumen mediano o grande, es necesario contemplar en estos casos una etapa de verificación de consistencia y preparación de los datos.

3. Problemas que derivan de la confidencialidad de la información

Dado que este tema en sí puede ser polémico, encontrándose al respecto distintas opiniones, se transcribe parte de los párrafos

Por ello es altamente recomendable para volúmenes medios y grandes de datos;

b) Requiere una relativamente pequeña cantidad de horas-hombre para programarlo y muchos de sus parámetros pueden ser especificados por parámetros que no tengan conocimientos previos de programación;

c) Es posible obtener la presentación final que se desea, incluso directamente para publicación;

d) Existen versiones de CRTIS que son operativas en computadoras pequeñas (con 32 Kbytes de memoria principal), y otras versiones que permiten obtener mayor número de cuadros por página, para equipos de mayor dimensión.

de la publicación de Naciones Unidas, "La aplicación de la tecnología de computadores al desarrollo".^{3/}

"En algunos países existe la creciente preocupación de que las computadoras menoscaban la vida privada y los derechos humanos. Hasta hace poco, los registros que se mantenían sobre un individuo contenían datos acerca de un aspecto de él y se llevaban en un lugar con una responsabilidad funcional bien definida. La tendencia del Gobierno moderno es organizar bancos electrónicos de datos con los registros de un individuo y reunirlos en un solo lugar. La consecuencia es que llega a ser fácilmente accesible un gran volumen de información detallada sobre una persona, información que puede no siempre ser correcta y quizás pueda usarse en perjuicio y sin conocimiento del interesado. Pocos países tienen un concepto jurídico de la vida privada. Ninguno ha aprobado todavía legislación que regule los bancos de datos en cuanto se relacionan con la vida privada, pero en varios se están estudiando propuestas. El problema se complica porque se dan circunstancias en que se necesita legítimamente, por parte de la policía o por motivos de la seguridad nacional, información muy confidencial sobre un individuo."

"Al examinar qué tipo de reglamentación podría adoptarse para los bancos de datos se plantean varias cuestiones evidentes sobre las clases de datos que pueden reunirse, quien las ha de reunir, cuanto tiempo deben conservarse los datos, cómo se comprueba

^{3/} "La aplicación de la tecnología de computadores al desarrollo" Naciones Unidas, Nueva York, 1971. Pp. 80-81

su exactitud, quién debe tener acceso a ellos, y de qué manera debe mantenerse la seguridad. Las respuestas a algunas de estas cuestiones dependen de consideraciones técnicas y quizás haya que diseñar nuevas características en los equipos de computadoras para obtener técnicas satisfactorias. Sin embargo, la mayoría de las cuestiones que se refieren a los derechos generales del individuo, según la definición y costumbres de su país, y hay que preocuparse por qué las computadoras no se empleen como instrumento para limitar tales derechos".

De acuerdo con estos puntos de vista, el problema existe cuando están en juego los derechos individuales o la protección del individuo. No todos los datos estarán en la misma situación y muchas veces se pueden usar sin restricciones informaciones almacenadas en un banco de datos.

4. Recursos necesarios para que un banco de datos sea operativo

Un banco de datos lo entendemos como un sistema dinámico.

Eso implica ciertas acciones, como por ejemplo:

- a) Actualización de los microarchivos y macroarchivos;
- b) Difusión del tipo de información accesible y de los servicios que el banco pueda prestar;
- c) Recuperación de la información que entregue en un mínimo de tiempo y a un costo razonable.

Es evidente que los recursos necesarios para que un banco de datos opere de manera eficiente dependerán fundamentalmente del tipo de servicios que preste y de sus objetivos básicos. Así, por ejemplo, en un comienzo, digamos hace 5 años, se hacía mucho énfasis en la función de almacenar grandes cantidades de datos, lo que exigía equipos de computación muy avanzados, pero no un esfuerzo encaminado a analizar los requerimientos básicos de esos datos, o su actualización y recuperación.

Suponiendo que las acciones que se desean realizar son las tres enunciadas al comienzo de este capítulo, es necesario cumplir una serie de requisitos que se reflejan en la existencia de personal calificado y equipos de computación, sistemas de recuperación de información operables y recursos financieros.

El primero de los requerimientos es probablemente el más importante, y entendemos que este personal debe incluir tres categorías:

- a) Especialistas en computación propiamente tal;
- b) Analistas de sistemas que puedan hacer de puente entre el usuario de la información y el especialista en computación; y
- c) Especialistas en los temas sustantivos a que se refiere la información.

Algunos usuarios de información estadística, ciertamente desearían que toda la información que necesiten en un momento dado

esté accesible en una misma fuente. Se puede imaginar las dificultades que acarrearía contar con especialistas en diferentes campos, como estadísticas de transportes, de comercio exterior, de empleo, de migración, etc.

Un banco de datos y sistema de información que pueda cubrir todos estos campos requeriría una gran cantidad de personal especializado y parecería lógico que se tendiera a esto, pero en la práctica no conocemos aún ningún sistema de este tipo que esté en operación. En algunos casos no es que no se haya intentado, sino que por otras razones se ha preferido contar con varios sistemas sectoriales que cubran tan sólo parte del problema global. En estos casos se restringe la especialidad de los profesionales que deben intervenir. En todo caso, si se pretende que el material disponible sea realmente aprovechable, es fundamental contar con profesionales que conozcan del tema sustantivo, como apoyo a los especialistas en computación.

Con respecto a las necesidades de equipos de computación, ellas dependen fundamentalmente del objetivo del banco, de los volúmenes de datos en juego, de los sistemas de recuperación desarrollados y de la forma como sean implementados los sistemas de actualización. Por ejemplo, algunos sistemas requieren dispositivos en línea para recuperar y actualizar los archivos, mientras en otros casos se desea una frecuencia de proceso mensual, trimestral, semestral o anual. En todo caso, se puede decir que en general un banco

de datos no está asociado necesariamente a un centro de computación de su uso exclusivo. Se puede decir, además, que en general será más económico, hasta que el volumen de datos lo justifique, hacer uso de equipos ya instalados. Ello queda justificado por el problema de economía de escalas, que en el campo de la computación pasa a tener cada vez más vigencia. Este punto no siempre ha sido tomado suficientemente en cuenta, y es así como en América Latina se ve en general una gran cantidad de pequeños computadores, no todos suficientemente ocupados y cada uno de ellos con costos fijos muy altos, y sin contar, a veces, con personal del nivel técnico que sería deseable. Ojalá en un futuro cercano se busquen los mecanismos para llegar a la integración de los recursos existentes, lo cual redundará en mejor uso de los equipos y en un costo más razonable.

III. DIFERENTES ENFOQUES QUE SE PUEDE DAR AL ALMACENAMIENTO Y RECUPERACION DE INFORMACION.

En este capítulo se analizarán algunas características inherentes a niveles de datos. Al respecto, se puede indicar que entre los especialistas que manejan sistemas de datos estadísticos surge la necesidad de distinguir entre dos niveles de datos: un micro nivel, al cual llamamos microdato, que corresponde a las respuestas individuales a un cuestionario de encuesta, censo o formulario que indica una acción determinada. Un ejemplo de microdato podría ser el precio de un determinado producto en un almacén, un determinado día. Ese microdato será recogido y eventualmente incorporado al índice nacional de precios al consumidor. A este tipo de microdato se acostumbra llamar datos básicos o primarios.

El segundo, sería el macro nivel, en el cual se manejan macrodatos. Podríamos definirlos como sumarios y/o estimaciones preparadas en base a los microdatos. El índice de precios al consumidor es un macrodato, como también lo son la estructura por edad de la población de un país. También los llaman datos elaborados o estadísticas derivadas.

Se hace esta distinción entre niveles de datos por el efecto que tienen sobre los procedimientos de computación. Algo de ello ya fue mencionado en el capítulo II; al trabajar con microdatos, la confidencialidad pasa a ser un problema muy importante. Además,

debido a la variedad y al volumen de los mismos, se requieren procedimientos especiales de almacenamiento y recuperación. Su característica fundamental, y que dice relación con elaboración de datos, es que por su volumen, en la mayoría de los casos un archivo de microdatos es de difícil manejo, lo que se refleja en la lentitud para atender a las solicitudes de los usuarios y en el elevado costo de operación.

Es por estas razones que, por lo general, además de tener un microarchivo del total de los casos disponibles, se generan subarchivos de microdatos que sirven para estudios muy concretos. Así, por ejemplo, en un censo o muestra de censo, podríamos tener, además del archivo total, algunos subarchivos de todos aquellos individuos que cumplen con ciertas características, como por ejemplo:

- a) Total de extranjeros;
- b) Personas de 10 y más años, económicamente activas;
- c) Mujeres de 15 a 49 años;
- d) Migrantes entre división administrativa mayor, etc.

Dependiendo de la frecuencia con que son consultados y estudiados los subconjuntos del microarchivo, podría justificarse su almacenamiento en archivos físicos bien diferenciados. Por otra parte, existen técnicas de programación que permiten estudiar selectivamente parte de un microarchivo mediante el uso de programas como SPSS, OSIRIS y CENTS. Sin embargo, ello implica en cada oportunidad leer

en la computadora el microarchivo completo para encontrar y procesar solo aquellos casos de interés. Así planteado el problema, la decisión debe ser tomada en base a consideraciones de tiempo y de tipo económico. Además de otras razones, este tipo de dificultad justifica el empleo de muestra para facilitar el uso de la información que proviene de los censos, por parte de diferentes usuarios. En un país de 20 millones de habitantes se pueden hacer tabulaciones especiales en base a una muestra de 400 o 500 mil personas, pero no con la población total.

Los microdatos presentan problemas diferentes, ya que son los datos más frecuentemente usados por los investigadores y en general los más solicitados por el público. Por estos motivos, se

debe preparar la documentación necesaria para explicar los datos a este nivel, y se deben contemplar las condiciones especiales para su acceso. Así se puede producir, a través de una variedad de formas de almacenamiento, tabulaciones especiales, informes para publicación, microfilm, tarjetas perforadas, cintas magnéticas, despliegue de datos a través de pantallas de televisión (terminales CRT), despliegue de datos o cuadros a través de terminales de consola, etc.

El "Bureau of the Census" de Estados Unidos ha incurrido en nuevas formas de presentar macrodatos, teniendo en cuenta que muchos usuarios encuentran dificultades para leer o interpretar tablas estadísticas. Para tratar de subsanar estos inconvenientes, se

están produciendo informes explicativos, gráficos e informaciones periódicas, todas ellas elaboradas mediante programas de computadora previamente establecidos. Esta información llega a las agencias de prensa locales, para uso a través de diarios y revistas, estaciones de radio y canales de televisión.

Es posible que al trabajar con macrodatos se pueda llegar a salvar en cierta forma las dificultades que existen para contar con bancos de datos que cubran diferentes campos. Al tener bancos de datos especializados en temas específicos, los que hemos llamado sectoriales, y que cuenten con ambos niveles de dato, se puedan organizar archivos de macrodato sobre temas que sin ser parte sustantiva del sector puedan ser de utilidad para los usuarios. Es lo que llamaríamos sistemas de almacenamiento y recuperación mixtos. Se estima que cada vez se darán más las condiciones para que esto sea realidad, ya que basta con que existan ciertas reglas de estandarización para los bancos de datos sectoriales, de modo que ellos puedan intercambiar archivos de macrodato.

Existen aún casos muy especiales, que no se ajustan a las situaciones ya mencionadas. Un ejemplo de ellos podría ser las estadísticas de servicio en salud maternal y planificación familiar. En este caso concreto, la principal unidad de estudio, para ciertos análisis, es la mujer, y se hace el seguimiento de diferentes acontecimientos. Se registran como acciones las actividades más

significantes, desarrolladas por el programa. No tendría sentido hacer un resumen de estas acciones, ya que cada una de ellas representa una unidad de registro y deba incluirse en el sistema. Luego, cada vez que se desea hacer un estudio de eficacia de uso de un método anticonceptivo, se usa como base el archivo de los microdatos, actualizado hasta el momento de cierre del método. Sin embargo, tal como se desprende del diagrama de bloques que se muestra en la publicación de referencia ("Elaboración de datos de estadísticas de servicio en Salud y Planificación Familiar", CELADE, 1972), hay otros estudios o tabulaciones que se van haciendo en función de períodos de tiempo, o sea, existen subarchivos que van siendo resumidos en macrodatos para atender a los informes mensuales y semestrales de actividades.

Finalmente, se presenta un ejemplo de obtención de un cuadro que podría considerarse como un macrodato que estaba almacenado en una cinta magnética, para obtener una presentación a través de un programa graficador. Esto demuestra que se puede pensar en hacer correlaciones entre dos macrodatos, o cualquier otro tipo de análisis.

El ejemplo mencionado, que es parte de uno de los proyectos de CELADE (Investigación de Migración Internacional en Latinoamérica - DIMIA), tiene la siguiente secuencia de operaciones:

a) Del censo de Panamá de 1970 se seleccionó el total de extranjeros, en un subarchivo que llamaremos "EXPAN" (archivo de microdatos);

b) Este subarchivo EXPAN fue procesado a través de un programa escrito en lenguaje GENS, para obtener ocho cuadros. De acuerdo a un plan de tabulaciones preestablecido, estos resultados son almacenados adicionalmente en un archivo de macrodatos que llamaremos CUEXPAN;

c) Posteriormente se creyó conveniente, además de entregar el Cuadro 1 en forma de tabulación tradicional, obtener un gráfico que presentara la estructura de la población por edad y sexo.

En un diagrama de bloques estos pasos quedarían representados como se muestra en la Figura 1.

En las figuras 2 a 5 se indican las salidas correspondientes al programa de tabulación y su graficación, mostrando su sistema, que permite almacenar y recuperar macrodatos.

Figura 1

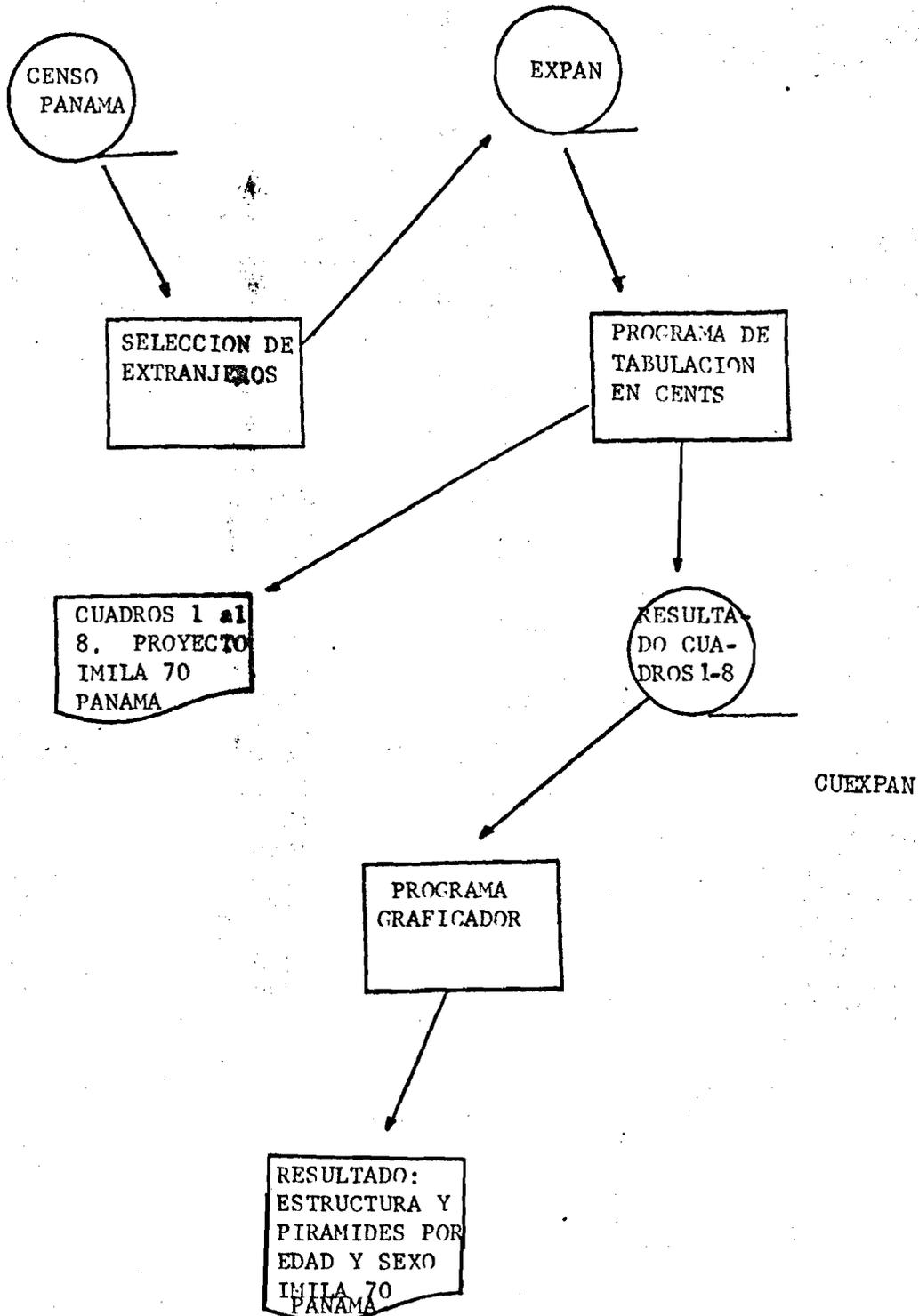


FIGURA 2

CELADE BANCO DE DATOS
PROGRAMA IMILA 1970
INVESTIGACION MIGRACION INTERNACIONAL
DE LATINOAMERICA ** PANAMA **

POBLACION NACIDA EN EL EXTRANJERO, SEGUN SEXO,
POR GRUPOS DE EDADES

GRUPOS QUINQUENALES DE EDADES	P O B L A C I O N			
	TOTAL	S E X O		Indice de masculinidad
		HOMERES	MUJERES	

** RESUMEN - TOTAL EXTRANJEROS

T O T A L	57275	31799	25476	124.8
00-04 AÑOS	2825	1382	1443	95.8
05-09 AÑOS	3189	1624	1565	103.8
10-14 AÑOS	2719	1402	1317	106.5
15-19 AÑOS	2767	1268	1499	84.6
20-24 AÑOS	5040	2674	2366	113.0
25-29 AÑOS	4417	2315	2102	110.1
30-34 AÑOS	4082	2175	1907	114.1
35-39 AÑOS	3849	2138	1711	125.0
40-44 AÑOS	3781	2194	1587	138.2
45-49 AÑOS	3837	2381	1456	163.5
50-54 AÑOS	3872	2436	1436	169.6
55-59 AÑOS	3972	2402	1570	153.0
60-64 AÑOS	3685	2187	1498	146.0
65-69 AÑOS	3093	1759	1334	131.9
70-74 AÑOS	2251	1247	1004	124.2
75-79 AÑOS	1682	955	727	131.4
80-84 AÑOS	1354	795	559	142.2
85 AÑOS Y MAS..	860	465	395	117.7

FIGURA 3

ESTRUCTURA POR EDAD Y SEXO

PANAMA. TOTAL DE EXTRANJEROS

E D A D	P O B L A C I O N			
	H O M B R E S		M U J E R E S	
	Números Absolutos	%	Números Absolutos	%
85 y Más	465	0.81	395	0.69
80-84	795	1.39	559	0.98
75-79	955	1.67	727	1.27
70-74	1247	2.18	1004	1.75
65-69	1759	3.07	1334	2.33
60-64	2187	3.82	1498	2.62
55-59	2402	4.19	1570	2.74
50-54	2436	4.25	1436	2.51
45-49	2381	4.16	1456	2.54
40-44	2194	3.83	1587	2.77
35-39	2138	3.73	1711	2.99
30-34	2175	3.80	1907	3.33
25-29	2315	4.04	2102	3.67
20-24	2674	4.67	2366	4.13
15-19	1268	2.21	1499	2.62
10-14	1402	2.45	1317	2.30
5- 9	1624	2.84	1565	2.73
0- 4	1382	2.41	1443	2.52
TOTAL	31799	55.52	25476	44.48

FIGURA 3b

PIRAMIDE POR EDAD Y SEXO

HOMBRES

**

MUJERES



FIGURA 4

CELADE BANCO DE DATOS
PROGRAMA IMILA 1970
INVESTIGACION MIGRACION INTERNACIONAL
DE LATINOAMERICA ** PANAMA **

POBLACION NACIDA EN EL EXTRANJERO, SEGUN SEXO,
POR GRUPOS DE EDADES

GRUPOS QUINQUENALES DE EDADES	P O B L A C I O N			INDICE DE MASCULINIDAD
	TOTAL	S E X O		
		HOMBRES	MUJERES	

TOTAL PAISES DEL CARIBE - TOTAL

T O T A L	7750	3740	4010	93.3
00-04 AÑOS	40	19	21	90.5
05-09 AÑOS	78	39	39	100.0
10-14 AÑOS	125	73	52	140.4
15-19 AÑOS	109	53	56	94.6
20-24 AÑOS	151	70	81	86.4
25-29 AÑOS	165	72	93	77.4
30-34 AÑOS	166	75	91	82.4
35-39 AÑOS	190	86	104	82.7
40-44 AÑOS	237	117	120	97.5
45-49 AÑOS	246	144	102	141.2
50-54 AÑOS	364	180	184	97.8
55-59 AÑOS	549	230	319	72.1
60-64 AÑOS	900	391	509	76.8
65-69 AÑOS	1065	436	629	69.3
70-74 AÑOS	1014	457	557	82.0
75-79 AÑOS	957	513	444	115.5
80-84 AÑOS	870	515	355	145.1
85 AÑOS Y MAS...	524	270	254	106.3

FIGURA 5

ESTRUCTURA POR EDAD Y SEXO

PANAMA. TOTAL DE PAISES DEL CARIBE

E D A D	P O B L A C I O N			
	H O M B R E S		M U J E R E S	
	Números Absolutos	%	Números Absolutos	%
85 y MAS	270	3.48	254	3.28
80-84	515	6.65	355	4.58
75-79	513	6.62	444	5.73
70-74	457	5.90	557	7.19
65-69	436	5.63	629	8.12
60-64	391	5.05	509	6.57
55-59	230	2.97	319	4.12
50-54	180	2.32	184	2.37
45-49	144	1.86	102	1.32
40-44	117	1.51	120	1.55
35-39	86	1.11	104	1.34
30-34	75	0.97	91	1.17
25-29	72	0.93	93	1.20
20-24	70	0.90	81	1.05
15-19	53	0.68	56	0.72
10-14	73	0.94	52	0.67
5- 9	39	0.50	39	0.50
0- 4	19	0.25	21	0.27
TOTAL	3740	48.26	4010	51.74

FIGURA 5b

PIRAMIDE POR EDAD Y SEXO

HOMBRES

**

MUJERES



IV. EL USO DE TERMINALES CONVERSACIONALES

Existen sistemas que permiten el uso simultáneo de una computadora por una gran cantidad de usuarios. Esto se hace a través de "terminales", que son máquinas de escribir convenientemente acondicionadas para comunicarse en forma directa con el computador central, utilizando un cierto "lenguaje de terminales"

La utilización simultánea de un computador por muchos terminales, o sea, por muchos usuarios diferentes, es posible gracias a los "sistemas de tiempo compartido" que aprovechan al máximo el tiempo de uso de la unidad central de proceso de la computadora (CPU), atendiendo a cada uno de los distintos terminales por fracciones de segundo. Esto da la impresión de que cada terminal está utilizando una computadora en forma privada, y en la práctica es como si así fuera.

Tal como se ha planteado en los capítulos anteriores, esta es una nueva forma de enfocar el problema de almacenamiento y recuperación de información estadística. Su campo de aplicación abarca los siguientes temas:

- a) Actualización de archivos;
- b) Consulta de archivos;
- c) Recuperación de macrodatos;
- d) Análisis estadísticos con macrodatos y archivos de microdato;

e) Almacenamiento y recuperación de información no numérica;

f) Ensayo de líneas de investigación con subconjuntos pequeños de datos, a fin de demostrar ciertas hipótesis antes de procesarlos en forma masiva, en la computadora.

Existe una gran variedad de terminales. Algunos son sólo pequeñas máquinas de escribir; a otros se les agrega una pantalla de televisión (terminales CRT CATODE-RAY-TUBE). Pueden tener entrada y salida en cinta de papel perforada o tarjetas perforadas; por otro lado, en terminales que trabajan a alta velocidad, podría incluirse una impresora más veloz que una de teclado. Por último, existen terminales que en el momento que están conectadas a la computadora pasan a ser de tipo conversacional, pero que tienen cierta capacidad de procesamiento fuera de línea. Es lo que se conoce como "terminales inteligentes".

En general, por su forma de trabajar a través de "sistema de tiempo compartido", sus costos de operación para trabajos del tipo almacenamiento y recuperación de información son relativamente bajos. Daremos algunos costos unitarios a esta fecha:

a) Arriendo mensual de un terminal tipo máquina de escribir, con su modulador: del orden de US\$ 150 por mes.

b) Arriendo de una línea dedicada, en el area urbana (en función de número de plantas por que pase): puede variar entre

US\$ 15 y US\$ 30 por mes.^{4/}

c) En cuanto a la computadora, generalmente se considera por separado el tiempo de conexión, y por otra parte el tiempo de CPU (Unidad Central de Proceso). Esto último es lo que hace atractivo el empleo de este tipo de equipo para almacenamiento y recuperación de datos. Sólo para dar una idea, se acotan algunos valores promedio:

- i) costo de una hora de conexión a un equipo como un 360/40-H: US\$ 10
- ii) costo de un minuto de CPU a un equipo como un 360/40-H: US\$ 6.

Estos sistemas de tiempo compartido son soportados, en forma general, por equipos de medianas o grandes dimensiones, ya que necesitan como mínimo 100 KBytes de memoria principal. Esto en realidad depende del fabricante, pues también existen ejemplos concretos de equipos que tienen esta posibilidad y sólo exigen 16 KBytes de memoria principal, pero en estos casos se presentan algunas limitaciones de trabajo.

^{4/} Se ha considerado línea dedicada porque, en casos como Chile, no se acepta transmisión de datos a través de líneas conmutadas. En todo caso, se ha supuesto la condición más desfavorable, cual sea tener necesidad de una línea privada. Como información adicional, y debido al gran impacto que se prevea van a tener los terminales en el futuro, queremos mostrar cual sería un costo de corrección, vía satélite, entre dos lugares distantes. Así, por ejemplo, entre Santiago de Chile y Panamá, el valor del minuto de conexión es de US\$ 3. Esto abre grandes expectativas y refuerza aún más la idea de crear bancos de datos sectoriales, incluso a nivel regional e internacional, con la posibilidad de enlaces entre ellos para resolver problemas concretos.

Otro factor importante, es el lenguaje con el que se puede trabajar a través del terminal, el que está condicionado por el soporte de programación que ofrece el equipo central que se usa. Sin pretender mencionar los múltiples lenguajes desarrollados e implementados en muchas universidades, mencionaremos solo aquellos de alto nivel y de uso más frecuente.

- APL ("A Program Language")
- PL/I (Program Language # 1)
- RAX ("Remote Access Computing System")
- BASIC
- COBOL ("Common Business Oriented Language")

Un aspecto importante que debe tomarse en cuenta es el campo de aplicación de cada uno de estos lenguajes, pues algunos de ellos sirven especialmente para casos específicos, y otros cumplen eficientemente otras funciones; dependiendo de la aplicación que se desea hacer a través de terminales es que se debe elegir el lenguaje a emplear. En algunos casos concretos, como es el de Chile, no se ofrecen servicios de tiempo compartido sino que con un lenguaje, lo que restringe el campo de aplicación.

En las figuras 6 y 7 se presentan dos aplicaciones realizadas con lenguaje APL, y que tienen relación con problemas o análisis estadísticos.

FIGURA 6

15 9 p GA1

785	746	756	770	748	795	814	753	794
805	759	750	737	752	773	809	777	784
760	779	774	746	758	761	791	832	776
782	823	778	792	807	833	769	766	834
831	797	760	817	776	823	860	848	795
764	831	850	847	836	769	818	846	796
798	790	849	830	814	754	853	781	770
806	834	738	783	854	751	832	829	812
798	774	783	803	873	860	740	849	826
789	790	854	796	743	821	758	876	832
865	810	760	749	830	766	881	879	825
776	841	884	839	878	850	877	781	800
837	771	804	754	853	739	846	831	788
862	869	811	799	759	785	840	767	859
833	824	806	874	775	875	769	872	802

DSTAT GA1

SAMPLE SIZE	135
MAXIMUN	884
MINIMUN	737
RANGE	147
MEAN	804.963
VARIANCE	1585.8
STANDARD DESVIATION	39.8221
MEAN DESVIATION	34.0944
MEDIAN	800
MODE	760 769 776 831 832

FIGURA 7

MH120

0.0163 0.00205 0.00102 0.00102 0.00173 0.00246 0.00328 0.00444 0.00662 0.0105
 0.01724 0.02891 0.04785 0.07882 0.12664 0.20252 0.30774 0.44793 1

TVQX1 MH120

ESCRIBA TITULO DE ESTA TABLA

TABLAS MODELO DE MORTALIDAD : EXTRAPOLACION *** NIVEL 120 *** HOMBRES

TABLAS MODELO DE MORTALIDAD : EXTRAPOLACION *** NIVEL 120 *** HOMBRES

X	Q(X,N)	L(X)	D(X,N)	L(X,N)	T(X)	E(X)	P(X,X+5)	P(X Y +)
0	16300	100000	1630	98778	7335223	7335	997473	986534
1	2050	98370	202	393057	7236445	7356	0	945684
5	1020	98168	100	490591	6843389	6971	998980	928312
10	1020	98068	100	490091	6352797	6478	998625	922854
15	1730	97968	169	489417	5862706	5984	997905	916520
20	2460	97799	241	488392	5373289	5494	997131	909107
25	3280	97558	320	486991	4884897	5007	996141	900307
30	4440	97238	432	485111	4397907	4523	994472	889695
35	6620	96806	641	482430	3912795	4042	991446	876705
40	10500	96166	1010	478303	3430366	3567	986148	860568
45	17240	95156	1640	471678	2952062	3102	976976	840221
50	28910	93515	2704	460818	2480385	2652	961759	814215
55	47850	90812	4345	443196	2019567	2224	937045	780549
60	78820	86466	6815	415294	1576371	1823	898251	736551
65	126640	79651	10087	373038	1161077	1458	837985	678714
70	202520	69564	14088	312600	788039	1133	750798	603319
75	307740	55476	17072	234700	475439	857	634912	506352
80	447930	38404	17202	149014	240739	627	0	381017
85	1000000	21202	21202	91726	91726	433	0	0

PB = 0.983668

V. UNA EXPERIENCIA EN AMERICA LATINA: EL BANCO DE DATOS DE CELADE

El Banco de Datos de CELADE comenzó a operar a raíz de la recopilación de muestras de los censos de población de 1960, y se fue ampliando con la anexión de materiales de otras fuentes, tales como encuestas de fecundidad, encuestas sobre aborto, encuestas de migración a áreas metropolitanas, censos experimentales y otros. Para mayor información ver "Boletín del Banco de Datos , N° 5, editado por CELADE.

El Banco de Datos cuenta actualmente con un Servicio de Computación que hace uso principalmente del computador IBM 360/40G, de la Universidad de Chile. Además, está instalado en CELADE un terminal conversacional IBM 2741, conectado a un computador IBM 360/40H, en la modalidad de tiempo compartido. Es usado un promedio de 40 horas mensuales, empleando para ello el lenguaje disponible APL. El servicio de computación cuenta con 9 personas: 2 calculistas y 7 especialistas en procesamiento de datos.

En ese Banco de Datos se trabaja indistintamente con archivos de microdatos y macrodatos y operativamente se cuentan con los siguientes subsistemas o paquetes de programas:

- CENTS ("Census Tabulation System")
- SPSS ("Statistical Package for the Social Sciences")
- OSIRIS/40 ("Organized Set of Integrated Routines for Investigation with Statistics")

- SYMAP ("Synagraphic Computer Mapping Program")
- KWIC ("Key Word in Context")
- SIDES (Sistema de Estadísticas de Servicio)
- Programas Estadísticos en APL.

Además de muchos programas específicos, escritos en ASSEMBLER/
360 y FORTRAN IV.

Para la mayoría de las tabulaciones cruzadas que son solicitadas se opera con el sistema CENTS. Se emplean los sub-sistemas SPSS y OSIRIS para análisis estadísticos especiales y se utiliza el subsistema SYMAP para la representación gráfica, usando una impresora de computadora. El subsistema KWIC ha sido usado para almacenar y recuperar el inventario de cintas magnéticas, tal como se muestra en el anexo del Boletín del Banco de Datos N° 5. El sistema SIDES ha sido usado como plan piloto para hacer una evaluación de los programas de planificación familiar y materno infantil del SERMENA - Servicio Médico Nacional de Empleados, en Valparaíso.

Trataremos de explicar un poco en qué consiste el Proyecto OMUECE (Operación Muestra de Censo), que es el que aporta el mayor número de datos al Banco. Nos referiremos tan sólo a la parte correspondiente a aspectos generales, toda vez que el Boletín del Banco de Datos N° 5 presenta una información exhaustiva sobre el tema.

El proyecto se constituye de muestras de censos nacionales de población, con énfasis en los países de la región latinoamericana. Hasta la fecha se ha logrado reunir muestras de 17 países que levantaron censos en la década del 60. Están incorporadas al Banco de Datos, de igual modo, las muestras de los censos de población realizados en la década del 70 de los siguientes países: Chile, Nicaragua, Panamá, República Dominicana, y Trinidad y Tobago.

El proyecto es de duración indefinida. Se están transfiriendo los datos a cintas magnéticas, después de uniformar los códigos y criterios de clasificación. Se mantiene, además, la información en su forma original.

Independientemente de las tabulaciones que prepara CELADE para sus programas de estudios comparativos y análisis de la información demográfica a nivel de país, otras instituciones, tanto nacionales como internacionales, tienen acceso a la información concentrada en Santiago, solicitando la preparación de tabulaciones especiales, según sus propias necesidades.

Siempre que no haya restricciones impuestas por los donantes, el Centro podrá suministrar copia de los datos, en cintas magnéticas, acompañadas de los respectivos códigos.

OMUECE 60

Cuadro 1. Estado actual de las muestras disponibles en CELADE

País	Fecha del Censo	Población censada (miles)	Trasfido de la muestra		Situación actual
			Porcentaje	Número de casos	
Argentina	30-09-60	20 009	2.50	499 264	T
Brasil	01-09-60	70 967	1.29	913 598	T
Colombia	15-07-64	17 585	2.00	349 563	T
Costa Rica	01-04-63	1 336	6.04	80 723	T
Chile	29-11-60	7 340	1.20	87 933	T
Ecuador	25-11-62	4 476	3.04	135 999	T
El Salvador	02-05-61	2 511	1.03	25 814	T
Guatemala	18-04-64	4 210	4.98	209 556	T
Honduras	17-04-61	1 885	1.00	18 818	T
México	08-06-60	34 923	1.43	500 267	T
Panamá	11-12-60	1 013	5.27	53 298	T
Paraguay	14-10-62	1 819	4.91	89 360	T
Puerto Rico	01-04-60	2 350	10.21	240 000	NC
R. Dominicana	07-08-60	3 047	6.61	201 556	T
Uruguay	16-10-63	2 593	4.93	127 925	T
U.S.A.	01-04-60	179 562	0.10	179 562	NC
Venezuela	26-02-61	7 524	1.76	132 224	NC

Nota: Las letras empleadas en la columna "situación actual" significan:
 NC: No convertida a la codificación OMUECE;
 CC: Convertida a la codificación OMUECE
 T: Tabulados los cuadros. (La indicación de uno de estos símbolos supone que fueron cumplidas las etapas anteriores)

OMUECE 70

Cuadro 2. Datos de las muestras censales disponibles en CELADE

País	Fecha del Censo	Población censada (miles)	Tamaño de la muestra		Situación Actual
			Porcentaje	Número de casos	
Chile	04-1970	8 835	5.0%	442 657	NC TE ^{1/}
Nicaragua	05-1971	1.911	10.0%	189 469	NC MTE ^{2/}
Panamá	05-1970	1.428	20.0%	286 188	NC MTE
Trinidad y Tobago	04-1970	945	7.3%	69 396	NC --
R.Dominicana	01-1970	4 006	9.8%	390 151	NC MTE

Nota: Las letras empleadas en la columna "Situación Actual" significan:
 NC: No convertida a la codificación OMUECE 70
 M: Marginal de Frecuencias por Variable
 TE: Tabulaciones especiales solicitadas por el país a CELADE

^{1/} Publicada por el Instituto Nacional de Estadística de Chile

^{2/} Publicada por la Oficina Ejecutiva de los Censos de Nicaragua

BIBLIOGRAFIA

"International Data Library and Reference Service, List of Archive Holdings". Survey Research Center, University of California, Berkeley, California.

"File Inventory, Latin American Data Bank". Center for Latin American Studies, University of Florida. Diciembre 22, 1971.

"Boletín N° 5 del Banco de Datos" CELADE, agosto de 1972.

"Data Banks and Information Systems for National Statistics". R.C. Mendelsohn, Invited and Submitted Papers, IPIP Congress 71, LJUBLJANA.

"NBER Data Bank for Project Economics". NBER National Bureau of Economic Research, Inc. Agosto 1971.

"Procesamiento de Datos y Análisis de Sistemas". Leo J.A. Musseuma. Ed. F. Trillas, S.A. Enero 1970.

"Programas de Computación usados en el CELADE". Centro Latinoamericano de Demografía, Agosto 1972. Serie J. N° 1

S.P.S.S. "Statistical Package for the Social Sciences". Norman H. Nie, Dale H. Bent and C. Nedali Hull. Mc Graw Hill, 1970.

"Manual de Usuarios de OSIRIS/40". Institute for Social Research (ISR) University of Michigan, Ann Arbor, Michigan.

"Census Tabulation System"(CENTS), "An Approach to the Computer Tabulation of Data from a Population and Housing Census". Series ISP 4, N° 1, U.S. Department of Commerce, Bureau of the Census.

"La aplicación de la tecnología de computadoras al desarrollo". Publicación de Naciones Unidas S.71 II.A.1, 1971.

"Computerized Reporting of Census Data for Small Areas: A Promising New Technique for the 70's". Dr. George H. Brown, Director, U.S. Bureau of the Census, Conference on the Role of the Computer in the Economic and Social Research, Mexico, 1971.

"Estadísticas de Servicio en Programas de Salud y Planificación Familiar". Serie A, N°115, Anexo 3. CELADE

"APL-360 Primer". Ed. IBM, Serie H20-0689.

"IBM System/360 OS/DOS, ITF: PL/1 Introduction" Ed. IBM, Serie SC28-6833.

"System/360 Remote Access Computing System (360A-CX-17X) Versión 4 Application Description". Ed. IBM, Serie H20-0545.

"Basic Language, Reference Manual". Ed. Hewlett Packard, Serie 02116-9077.

"Programación COBOL". Daniel D. McCracken, Ed. Limusa Wiley, 1967.

"SYMAP", User's Reference Manual. Elliot E. Dudnik, Report N° 71-1 Department of Architecture, University of Illinois at Chicago Circle.

"Key Word in Context (KWIC) Indexing". Ed. IBM, Serie E20-3090.

"Estadísticas de servicio en programas de salud y planificación familiar" Experiencia en (SERMENA) Servicio Médico Nacional de Empleados, Valparaíso. CELADE, Serie A N° 115, Anexos 2 y 4.