

4. El Generador de Tablas Estadísticas

Antonio Flos Bassols

Los problemas de tabulación de los Censos y el G. T. S.

Tradicionalmente la presentación de los Censos se realiza mediante tablas estadísticas en las que se cruzan varios conceptos para reflejar de manera clásica una determinada característica de un campo de interés. Entre las tablas que se extraen están las básicas, es decir, aquellas mediante las cuales el Organismo encargado de los Censos presenta los resultados de éste y las tablas "a medida", esto es, aquellas tablas particularizadas sobre campos de especial interés para determinadas entidades.

Con la utilización de los ordenadores electrónicos los problemas de tabulación quedan muy reducidos respecto a los antiguos procedimientos manuales; sin embargo, se plantea un nuevo problema: la realización de los programas necesarios para obtener las tablas. En principio cada tabla o cada familia de sub-tablas —más adelante se definen estos conceptos— representaría un programa. Esto llevaría, en la tabulación de los Censos, a un tiempo muy elevado de programación y, por lo tanto, a un elevado coste. Además, cualquier explotación marginal —la confección de las tablas "a medida"— representaría también un tiempo y unos costes elevados, que en la mayoría de los casos dificultaría enormemente la utilización de la información censal como base de información en determinadas aplicaciones.

Estos razonamientos llevaron al I.N.S.E.E. a plantearse, a partir del Censo de Población de 1962, la necesidad de realizar una serie de programas generales destinados a automatizar las salidas de tablas estadísticas del Censo. Es decir, se trata de una serie de programas, generales para

todos los casos, cuya adaptación a cada caso particular se efectúa mediante la incorporación al programa de un número limitado de parámetros adecuados, fáciles de definir.

El sistema, denominado "Générateur de Tableaux Statistiques"—G.T.S.—, fue utilizado ya para el Censo de Población francés de 1962 sobre un ordenador 7074 IBM y un 1401 IBM. Posteriormente fue puesto a punto y ampliado para su utilización en los ordenadores de la serie IBM 360 de que dispone actualmente el I.N.S.E.E. El G.T.S. requiere un ordenador de la serie 360 con sistema operativo OS, cuya configuración mínima debe permitir una participación de 200 K y debe disponer de los siguientes periféricos:

- una impresora;
- tres o cinco unidades banda magnética (si se desea utilizar todas sus opciones);
- una unidad de discos 2314 y dos unidades 2311;
- un perforador de tarjetas.

Los programas han sido realizados en PL/1 y en *Assembler*.

A partir de 1970 el I.N.S.E.E. realizó un acuerdo con el C.A.P. (Centre d'Analyse et de Programmation), empresa privada de servicios en informática, que tomó a su cargo la comercialización del sistema. El precio del mismo en Francia, sin impuestos, es de 27.000 dólares incluyendo su instalación, formación del cliente, documentación y mantenimiento del sistema durante seis meses. Ha sido adquirido para la explotación de los Censos españoles de 1970 por el Instituto Nacional de Estadística.

La organización del fichero de los Censos

El generador de tablas requiere una especial organización del fichero a procesar. Este debe estar organizado

en niveles jerárquicos, con un máximo de cuatro niveles. En otras palabras, la información adopta forma arborescente lo que permite que la información común a varios registros aparezca solamente una vez como factor común a un nivel superior. Concretamente, el fichero del Censo de Población francés está estructurado del modo siguiente:

- Nivel geográfico (nivel 1) que contiene uno o más.
- Niveles edificio (nivel 2) que contiene uno o más
- Niveles vivienda (nivel 3) que contiene uno o más
- Niveles individuo (nivel 4)

El orden de los registros queda expresado en la figura 1.

N.º de orden	Nivel 1	Nivel 2	Nivel 3	Nivel 4
01	001			
02		001		
03			001	
04				001
05				002
06				003
07			002	
08				001
09				002
10		002		
11			003	
12				001
13				002
14				003
15	002			
16		001		
17			001	
18				001
19				002
20			002	
21				001

Fig. 1.



Definición de tablas

Para la comprensión de los *outputs* que pueden obtenerse mediante el G.T.S. es preciso distinguir entre: tablas, familias de sub-tablas y sub-tablas. Una tabla estadística viene definida por:

- sus criterios de ventilación;
- sus criterios de totalización;
- sus filtros.

Los criterios de ventilación son los criterios que definen las líneas y columnas. Puede haber un máximo de cuatro criterios por línea y cuatro por columna. Todas las unidades estadísticas que toman el mismo valor para todos los criterios que definen una casilla van en aquella casilla.

Por otra parte, en cada casilla pueden definirse hasta un total de 14 tipos de acumulaciones distintas, es decir, de criterios de totalización. Por ejemplo, un criterio sería sumar 1 a la casilla cada vez que aparece un nuevo individuo adjudicado a la misma. Otro criterio sería sumar el valor que toma una determinada variable en cada uno de los individuos de una casilla, etc.

Por filtros entendemos aquellas condiciones que debe tener una unidad estadística para ser tomada en cuenta en la confección de una tabla. Por ejemplo, en una tabla se tienen en cuenta sólo aquellos individuos de edad comprendida entre 14 y 60 años.

Si además de los tres criterios mencionados tenemos en cuenta la escala geográfica, se obtienen las denominadas familias de sub-tablas. Es decir, una familia de sub-tablas viene definida por una tabla y la escala de producción de la misma. Escala es el concepto genérico que designa los subconjuntos de población definidos por una participación del conjunto del fichero. Concretamente la escala puede ser, por ejemplo, la manzana. Así, cuando deseamos obtener la tabla A por manzanas, de hecho pedimos la familia de

sub-tablas A, siendo el *i* el subíndice que indica la escala "manzana". Cada una de las sub-tablas de la familia $A_i (a_{i1}, a_{i2}, \dots, a_{in})$ corresponderá a un suceso particular de la escala, es decir, a una manzana concreta. Desde un punto de vista de tratamiento podemos decir que cada vez que se produce una ruptura del código de manzana (que está a nivel 1) se confecciona una nueva sub-tabla.

Las únicas tablas que el G.T.S. obtiene directamente del fichero de Censos son las familias de sub-tablas que suelen denominarse tablas brutas. A partir de aquí es posible obtener, si se desea, las "tablas hijas", mediante superposición o recapitulación de las sub-tablas que componen una tabla bruta.

Diagrama de fases del G. T. S.

El diagrama de fases se describe gráficamente en la figura 2. Una vez se recibe la demanda de unas determinadas tablas se producen las tarjetas perforadas que contienen los parámetros correspondientes.

Mediante estas tarjetas de parámetros se generan los programas pertinentes. De hecho el G.T.S. no obtiene por sí mismo las tablas, sino que genera un programa (o varios, según la memoria requerida), el cual a su vez realizará las tablas.

En la segunda fase se ejecuta el programa generado, es decir, el programa trata el fichero del Censo y obtiene las tablas brutas.

En la fase siguiente se producen, si así lo indica la demanda, las tablas hijas. Es la fase de recapitulación de tablas.

Finalmente la última fase es la de edición. En ella se calculan los totales parciales, porcentajes, medias, etc., se disponen las tablas por páginas de listado, se añade el título y se imprime. La presentación de resultados puede efectuarse de distintas formas. En la figura 3 se ex-

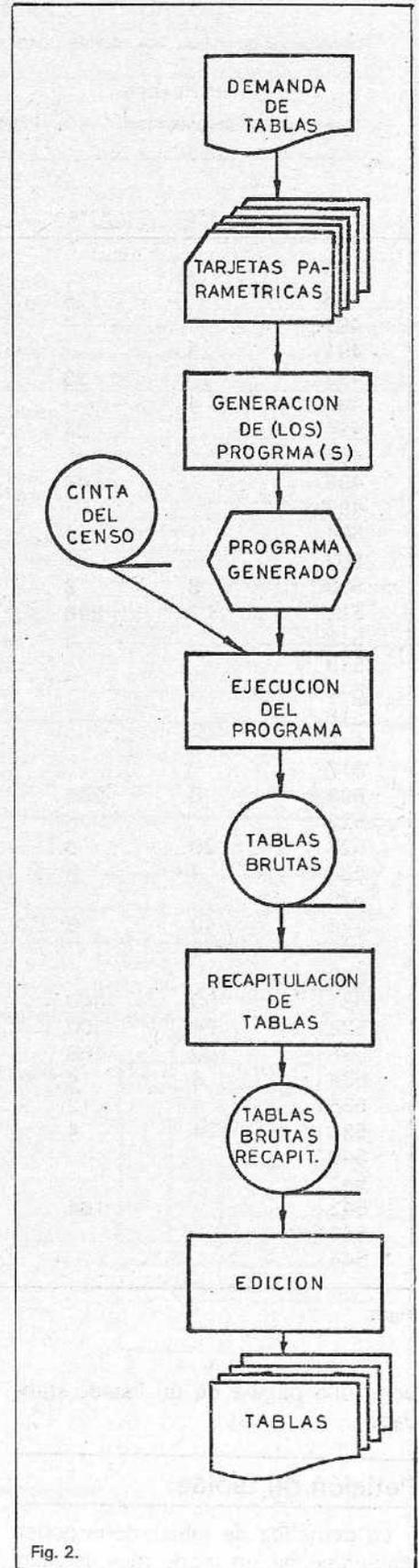


Fig. 2.

HAUTE-LOIRE
 sous ensemble géographique ↑
 (D). GA

RP68Q USP:4

CHAMP: TAI-IOU2 (ACTIFS AYANT UN EMPLOI)

DESSIN: (MCDAE)*(S*ST2K)

1*0 1*4 1* 2*0 2*4 2* * T0 T4

485									
486		176	176		124	124	300		300
487									
491	28		28	52	16	68	96	80	16
492	20	36	56	28	612	640	696	48	649
493	4		4	20	4	24	28	24	4
494		24	24		56	56	80		80
495									
496		44	44		116	116	160		160
497									
501									
502									
503	8	8	16	12	4	16	32	20	12
511		896	896		184	184	1080		1080
512									
513									
514									
515									
516	20		20	4		4	24	24	
517									
521	8	204	212		164	164	376	8	368
522		76	76		60	60	136		136
523	20	8	28	4		4	32	24	8
524	4	8	12		32	32	44	4	40
525									
526	40	8	48		4	4	52	40	12
527	12		12				12	12	
528									
531	100	640	740	24	16	40	780	124	656
532	24	300	324		24	24	348	24	324
533	100	188	288	8	20	23	316	108	208
534	8	4	12				12	3	4
535	4	12	16	4	4	44	60	8	52
536	24	4	28	12		12	40	36	4
540									
541									
542		164	164		52	52	216		216
543									
544									

Fig. 3.

pone una página de un listado *standard*.

Petición de tablas

La demanda de tablas debe poder efectuarse de un modo muy sencillo

para que el G.T.S. sea realmente útil. El demandante sólo tiene que rellenar dos impresos: el de descripción del *tableau père* y del *tableau fils* (figura 4). Para ello basta con conocer la nomenclatura del G. T. S. Por otra parte en esta tarea puede ser

ayudado por un especialista en G. T. S. con lo que desaparece todo obstáculo.

A partir de aquí se prepara el fichero y se obtienen las tarjetas de parámetros. La explotación puede ya empezar.

TABLEAU PERE

N° TABLEAU :

TITRE :

Nbre de FILS :

US de PRODUCTION : ECHELON : RECAPITULATION :

DEFINITION ECHELON - NON - EMBOITE :

CHAMP { GEOGRAPHIQUE :
NON GEOGRAPHIQUE :

DESCRIPTION DU TABLEAU : CRITERES DE TOTALISATIONS

CRITERES LIGNE	CRITERES COLONNE	N°	TITRE	DEFINITION	CHAMP	N°	TITRE	DEFINITION	CHAMP
		01	<input type="text"/>	<input type="text"/>	<input type="text"/>	08	<input type="text"/>	<input type="text"/>	<input type="text"/>
		02	<input type="text"/>	<input type="text"/>	<input type="text"/>	09	<input type="text"/>	<input type="text"/>	<input type="text"/>
		03	<input type="text"/>	<input type="text"/>	<input type="text"/>	10	<input type="text"/>	<input type="text"/>	<input type="text"/>
		04	<input type="text"/>	<input type="text"/>	<input type="text"/>	11	<input type="text"/>	<input type="text"/>	<input type="text"/>
		05	<input type="text"/>	<input type="text"/>	<input type="text"/>	12	<input type="text"/>	<input type="text"/>	<input type="text"/>
		06	<input type="text"/>	<input type="text"/>	<input type="text"/>	13	<input type="text"/>	<input type="text"/>	<input type="text"/>
		07	<input type="text"/>	<input type="text"/>	<input type="text"/>	14	<input type="text"/>	<input type="text"/>	<input type="text"/>

RENSEIGNEMENTS DIVERS : Rayer la mention inutile

Impression à l'échelon de production	OUI - NON
Sortie en 11 colonnes	OUI - NON
Impression du tableau PERE	OUI - NON
Totalisations en clair	OUI - NON
Consultation { Lignes } Type 1	OUI - NON
{ Colonnes }	OUI - NON
Suppression des lignes blanches	OUI - NON
Retour minuscule { Lignes }	OUI - NON
{ Colonne }	OUI - NON

Figura 4a. 1019-6-70 L

TABLEAU FILS

1019-6-70 L

N° TABLEAU PERE : N° FILS :

TITRE :

ou

IDEM PERE

TRAITEMENT LIGNE

4	3	2	1

8 7 6 5 4 3 2 1

TRAITEMENT COLONNE

4	3	2	1

8 7 6 5 4 3 2 1

Indiquer dans chaque colonne du tableau les noms des critères du tableau père et mettre, dans les cases sous-jacentes, une croix sous ceux qui sont à conserver.

Chaque case correspond à 1 caractère de l'indicatif du tableau père ; mettre une croix dans les cases correspondant aux totaux désirés.

1 % horizontaux
2 % verticaux
3 % par rapport à la case total général

Figura 4b. 79