Centro Latinoamericano de Demografía

CELADE

Documentos de Seminarios

MARGINALES

DS/10.
100.
1975.

CURSO LATINOAMERICANO DE
PROCESAMIENTO DE DATOS (PED)
APLICADO A LAS CIENCIAS SOCIALES

I N D I C E

		Página
I.	INTRODUCCION	1
II.	LOS PROGRAMAS MARGINALES DE CELADE	5
III.	PROGRAMAS MARGBCD Y MARBINCO	7
	1. Tarjeta CERO	10
	2. Tarjeta TITULO del trabajo	11
	3. Tarjeta de parámetro principal (DATA)	12
	4. Tarjetas de CONDICION de variable	13
	5. Tarjeta de FORMATO	14
	6. Tarjetas de EXCLUSION (uso opcional)	16
	7. Tarjeta de NOMBRE DE VARIABLE	17
	8. Tarjeta NOMBRE DE CATEGORIA (uso opcional)	17
IV.	PROGRAMA MARGINAL	19
	1. Tarjeta TITULO	19
	2. Tarjeta PRINCIPAL	19
	3. Tarjetas de CONDICION	19
V.	MARGBCD32 Y MARBINCO32	22
	1. Tarjeta de Título	22
	2. Tarjeta de parámetros	22
	3. Tarjeta de formato	23
	4. Tarjeta nombre de variable	24

I. INTRODUCCION

Por marginales de una variable entendemos la distribución de frecuencias de las unidades de estudio de un censo o una encuesta en cada una de las categorías de la variable. Pueden expresarse en frecuencias absolutas, porcentajes o porcentajes acumulados. En la figura 1 los marginales de la variable "religión de la entrevistada" se presentan en las tres formas.

Variable: "Religión de la entrevistada"

Categoría	Código	Frecuencia	Porcentaje	Porcentaje acumulado
Ninguna	1	14	1.05	1.05
Católica	2	1258	94.16	95.21
Protestante	3	23	1.72	96.93
Otra Religión	4	39	2.92	99.85
No Responde	0	2	0.15	100.00
TOTAL		1336		

Figura 1: Ejemplo de marginales de una variable

El cálculo de los marginales de variables es un tópico de suma importancia en el PED de datos censales o de encuestas. La necesidad de contar con los marginales de las variables deriva de su importancia como instrumento de análisis para los investigadores y para los analistas y programadores.

A partir de los marginales se pueden obtener gráficos estadísticos (histograma y polígono de frecuencias) y medidas estadísticas como la media (el valor medio de la variable), la moda (la categoría con mayor frecuencia), la mediana (categoría por debajo de la cual se encuentra la mitad de las unidades de estudio) y la desviación estándar (mide la dispersión de las unidades de estudio en relación a la media), etc.

Para los investigadores sociales los marginales permiten:

1. Decidir sobre la importancia de la variable, para un determinado estudio según su distribución en cada categoría. Por ejemplo, si en una determinada

variable la mayoría de las unidades de estudio se concentra en una sola categoría, la variable no tiene poder explicativo (en el ejemplo de la figura 1, el 94 por ciento de las mujeres son católicas). Una situación similar se presenta cuando un gran porcentaje de las unidades de estudio caen en las llamadas categorías residuales ("no responde", "no sabe", "otros", "no se aplica", etc.). Por ejemplo, si en la variable "sueldo" tenemos el 60 por ciento de los casos con "sueldo ignorado", no tiene sentido efectuar estudios con el 40 por ciento restante.

2. Hacer estudios comparativos. Por ejemplo, se pueden comparar los marginales de una variable entre censos de diferentes países o entre censos de un mismo país, realizados en diferentes épocas.

3. Tomar decisiones respecto a la transformación de variables. Por ejemplo, si se quiere recodificar una variable en un número menor de categorías, los marginales de la variable original permiten al investigador elaborar la recodificación más recomendable.

Durante el procesamiento de un censo o una encuesta, los marginales permiten a los analistas y programadores:

1. Detectar errores de códigos incorrectos y verificar las correcciones en la etapa de limpieza. Es un buen procedimiento, antes de empezar los procesos en un censo o una encuesta, obtener los marginales de las variables de una muestra para tener una idea de la calidad de los datos y de la cantidad de correcciones necesarias.

2. Sirve para chequear los procesos de transformación de variables. En el procesamiento de un censo o una encuesta, es frecuente la necesidad de transformar variables y al contar con los marginales de la variable antes y después de la transformación, es fácil verificar la corrección del programa. Sea, por ejemplo, la recodificación de la variable "número de hijos nacidos vivos" de la siguiente forma: menos de 2 hijos, 2 a 3 hijos, 4 a 7 hijos, 8 y más hijos; la transformación de los códigos puede ser expresada en el siguiente esquema:

Códigos de la variable original		Códigos de la variable recodificada
0, 1	---------->	1
2, 3	---------->	2
4, 7	---------->	3
8, 17	---------->	4

La figura 2 presenta los marginales de la variable antes y después de recodificada. Se puede ver que la frecuencia de casos en los nuevos códigos corresponden a la suma de las frecuencias de las agrupaciones de códigos originales correspondientes.

Variable: Número de hijos nacidos vivos						
Original				Recodificada		
Categoría	Código	Frecuencia		Categoría	Código	Frecuencia
Ningún hijo	0	103	→	Menos de 2 hijos	1	232
	1	129				
	2	169	→	2 a 3 hijos	2	330
	3	161				
	4	162	→	4 a 7 hijos	3	520
	5	137				
	6	114				
	7	107				
	8	75	→	8 y más	4	254
	9	71				
	10	47				
	11	31				
	12	13				
	13	9				
	14	4				
	15	2				
	16	0				
	17	2				
TOTAL		1336				

Figura 2: Marginales de la variable "hijos nacidos vivos" antes y después de recodificada.

3. Sirve para verificar si están correctos los cruces entre variables. En un cruce los totales horizontales y verticales representan los marginales de las respectivas variables (el nombre de marginales se originó por estar al margen del cruce). La figura 3 presenta un cruce entre las variables sexo y edad. Los totales verticales son los marginales de la primera y los horizontales de la segunda en grupos quinquenales.

POBLACION POR SEXO, SEGUN GRUPOS QUINQUENALES DE EDADES

Edad	Población		
	Total	Sexo	
		Hombres	Mujeres
Total	25814	12526	13288
0-4	4543	2266	2277
5-9	4045	1975	2070
10-14	3151	1632	1519
15-19	2421	1144	1277
20-24	2113	967	1146
25-29	1843	839	1004
30-34	1578	781	797
35-39	1420	702	718
40-44	1137	491	646
45-49	858	422	436
50-54	792	391	401
55-59	538	265	273
60-64	564	279	285
65-69	295	137	158
70-74	231	95	136
75-79	156	81	75
80-84	68	40	28
85-89	33	10	23
90-94	15	8	7
95-97	4	1	3
98 y más	6	-	6
No declara	3	-	3

Figura 3: Ejemplo de cruce entre dos variables.

II. LOS PROGRAMAS MARGINALES DE CELADE

Actualmente existen varios sistemas o paquetes de programas destinados al análisis de datos censales o de encuestas, como por ejemplo el SPSS, OSIRIS, etc. En cada uno de estos sistemas el usuario dispone de instrucciones para obtener marginales de variables.

Sin embargo, estos sistemas, al lado de las facilidades que ofrecen a los usuarios, traen consigo una serie de limitaciones que los hacen inaplicables en determinadas situaciones. Las principales limitaciones de estos sistemas son: exigen gran capacidad de almacenamiento y son lentos para el proceso de gran cantidad de datos, como es el caso de los censos.

Para solucionar este tipo de problemas, el Area de Computación de CELADE ha desarrollado un conjunto de programas destinados al cálculo de marginales de variables, atendiendo principalmente a los aspectos de velocidad del proceso y mínimo consumo de almacenamiento principal.

Evidentemente, los programas de CELADE tienen sus limitaciones. El usuario que cuenta con otras posibilidades, debe decidir en cada situación lo que es más conveniente.

Los programas marginales de CELADE son cinco y a continuación se describe la aplicación de cada uno de ellos:

MARGBCD - variables en caracteres (EBCDIC) de 1 o 2 dígitos
MARBINCO - variables que ocupan 1 byte en binario
MARGINAL - variable en carácter de 1 a 99 dígitos o en binario de 1 a 4 bytes
MARGBCD32 - versión chica de MARGBCD para 32K
MARGBINCO32 - versión chica de MARBINCO para 32K

Todos estos programas son dirigidos por <u>tarjetas parámetros</u>, en las cuales el usuario especifica controles que definen la encuesta o censo en estudio. De este modo, salvadas las limitaciones de cada uno de los programas, se puede decir que son de uso generalizado, es decir, sirven para cualquier encuesta o censo.

Todos los programas suponen que los parámetros son suministrados en tarjetas perforadas y los datos en estudio están en cinta o disco magnético. La figura

4 muestra el diagrama de sistema para los programas. Evidentemente, manejando recursos del sistema operativo y cambios pequeños en los programas, se puede ampliar el manejo de los archivos de entrada y salida.

Los programas están escritos para computadores IBM/360 o 370, ya que están escritos o poseen subrutinas en el lenguaje Ensamblador IBM/360 o 370 de estos computadores.

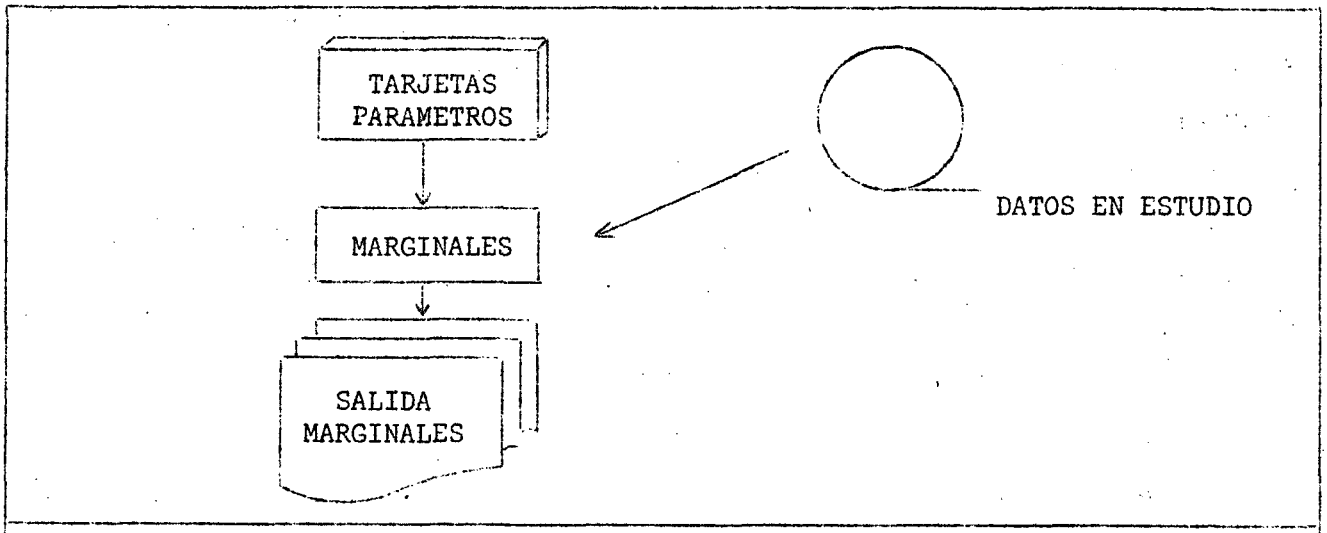


Figura 4: Diagrama de sistema de los programas marginales.

Existen versiones en OS y DOS de los programas MARGBCD, MARBINCO y MARGINAL. Las versiones chicas de los dos primeros son para DOS.

El usuario de cualquiera de los programas debe preocuparse en detalle cuando codifique las tarjetas parámetros, ya que los parámetros deben ocupar las columnas reservadas especialmente y las tarjetas parámetros deben tener un orden determinado.

A continuación estudiaremos cada uno de los programas en particular.

III. PROGRAMAS MARGBCD Y MARBINCO

Los programas MARGBCD y MARBINCO se documentan juntos porque son similares en muchos aspectos. Antes de preparar las tarjetas de control es necesario tener claro algunos conceptos generales, así como las limitaciones de los programas.

Codificación de las tarjetas de control: cada parámetro tiene su campo reservado en una determinada tarjeta; la codificación de valores numéricos debe ser justificada a la derecha y las columnas restantes a la izquierda pueden llenarse con ceros o dejarse en blanco; las tarjetas de control deben ser preparadas en un determinado orden.

Número de clases de tarjeta por unidad de estudio: cuando una unidad de estudio posee más de una clase de tarjeta, antes de calcular los marginales es necesario clasificar los datos por el número de la clase de tarjeta; los marginales son procesados solamente por una clase de tarjeta a la vez, es decir, procesa la primera clase de tarjeta y publica los resultados, a continuación procesa la segunda y publica los resultados, y así sucesivamente; entonces, el usuario debe preparar las tarjetas de control para cada clase de tarjeta en el mismo orden en que están clasificados los datos.

Numeración de las variables: existen dos conceptos de numeración de las variables: uno para designar el orden de las variables para el programa y el otro para los efectos de la impresión de los resultados. Para el programa, se da el número de la variable por su orden en la lectura, es decir, cuando el usuario especifica la lista de las variables de entrada, automáticamente le está asignando un número de orden. Por lo tanto, los números son siempre ascendentes a partir de 1 hasta el número igual al número de variables. Como el programa procesa una clase de tarjeta a la vez, para cada nueva clase de tarjeta la numeración empieza en 1. Para los efectos de la impresión, el usuario puede definir el inicio de la numeración con un valor diferente de 1 para cada clase de tarjeta y a partir de ahí la numeración es ascendente según el orden de lectura.

Resultados de los programas: la figura 5 presenta un ejemplo de salida de los programas MARGBCD y MARBINCO.

El programa trabaja con tres totales básicos llamados TOTAL A, TOTAL B, y TOTAL C:

TOTAL A = total de los casos para cada clase de tarjeta

TOTAL B = total A - (blancos + pers. X + perfs. Y, etc.)

TOTAL C = total B - (categorías excluidas)

Las estadísticas media, mediana y desviación estándar son calculadas siempre en relación al total C.

a) Frecuencia y porcentajes

Para cada categoría, se calculan las frecuencias, los porcentajes y porcentajes acumulados, sobre el Total B y el Total C.

b) Media

La media se calcula para todas las variables. La fórmula de cálculo es:

XB = IS2/IS1, donde: IS2 = ΣfX (X=valor de la categoría)
IS1 = Σf (f=frecuencia de la categoría)

c) Mediana

La mediana se calcula suponiendo siempre datos ordinales, los resultados son el valor de una categoría o un valor promedio entre dos categorías. Por ejemplo, si la mediana es 3, significa que la persona que divida la muestra pertenece a la categoría 3. En esos casos particulares, no tenemos una categoría como valor de la mediana sino un valor entre dos categorías. Esto significa que existe un 50 por ciento de casos en las categorías superiores. Como el promedio, la mediana se calcula considerando los casos que entran en el Total C.

d) Desviación estándar

Se calcula sobre el Total C, según la fórmula:

$$SD = \sqrt{S}; \quad S = (fx-XB)^2 / \Sigma fX$$

Exclusión de códigos de los cálculos estadísticos: los marginales permiten la posibilidad de excluir de los cálculos estadísticos (porcentajes simple y acumulado sobre total C, media, mediana y desviación estándar) hasta cuatro categorías por variable.

Marginales condicionados: el programa MARBINCO permite la posibilidad de seleccionar un subconjunto de la población para calcular marginales de una variable; es decir, se puede definir para cada variable una población particular que será considerada en los marginales.

MARGINALES DE REGISTRO FAMILIA BRASIL 60

URBANIZACION

V A R I A B L E : 57 TOTAL (A) DE CASOS =173620
TARJETA : 1 CASOS EN BLANCO= 0 PERF. X= 0 PERF. Y= 0 OTRAS 0 TOTAL (B) EXCL BL,X,Y=173620
COLUMNA : 57 CATEGORIAS NUMERICAS EXCLUIDAS: 100,100,100,100 TOTAL (C) FINAL =173620

NOMBRE	CATEG.	FRECUENCIA	TOTAL (B) PORCENT	TOTAL (B) PCT ACUM	TOTAL (C) PORCENT	TOTAL (C) PCT ACUM
	0	140155	80.73	80.73	80.73	80.73
	1	19458	11.21	91.93	11.21	91.93
	2	6905	3.98	95.91	3.98	95.91
	4	7102	4.09	100.00	4.09	100.00

ESTADISTICAS SOBRE TOTAL (C):
MEDIANA= 0
MEDIA= 0.36
DESV. STANDAR = 0.89

Fig. 5.- Ejemplo de una salida de MARGBCD o MARBINCO

Limitaciones principales: son las siguientes:

largo registro: 80 bytes.

largo variables: MARGBCD: 1 ó 2 columnas (bytes)

MARBINCO: 1 columna (byte)

número de variables por clase de tarjeta: 80

número de clase de tarjetas: ilimitado

número de casos: ilimitado

exclusión de categorías: 4 por variable

nº máximo de condicionamiento (MARBINCO): 3

para MARGBCD los códigos en blanco, perforaciones en zona 11(X), 12 (Y) y otros no numéricos son automáticamente excluidos de los cálculos estadísticos.

Preparación de las tarjetas de control: para cada uno de los programas, el usuario debe preparar las siguientes tarjetas de control:

MARGBCD	MARBINCO
CERO	CERO
TITULO	TITULO
DATA	DATA
FORMATO	CONDICIONES
	FORMATO
EXCLUSION	EXCLUSION
NOMBRE DE VARIABLES	NOMBRE DE VARIABLES
NOMBRE DE CATEGORIAS	NOMBRE DE CATEGORIAS

Existen dos diferencias entre MARGBCD y MARBINCO:

a) las tarjetas de FORMATO son distintas en la forma de codificarse

b) las tarjetas de CONDICION sólo existen para MARBINCO.

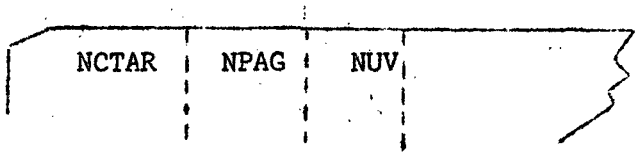
Las siguientes tarjetas de control son opcionales: EXCLUSION, CONDICIONES y NOMBRE DE CATEGORIAS. Las restantes son obligatorias.

A continuación se especifican los campos de cada tarjeta de control. En la figura 6 se esquematiza el ordenamiento de las tarjetas de control para MARGBCD y MARBINCO respectivamente.

1. Tarjeta CERO

Es obligatoria. Existe solamente una para toda la encuesta o para todo el proceso.

Forma General:



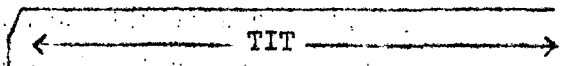
Descripción de la Tarjeta CERO:

Variable	Columna	Descripción
NCTAR	1-5	Número de clases de tarjetas diferentes que contiene la encuesta. Ej. si un cuestionario contiene tres clases de tarjetas, debe codificarse 00003. Aunque sea sólo una clase de tarjeta debe codificarse 00001 o ØØØØ1.
NPAG	6-10	Las hojas impresas con los marginales se numeran. NPAG especifica el valor inicial de la numeración de las páginas, y siempre es una unidad menos que el valor a ser impreso. Así: si NPAG = 0 o blanco imprimirá 1, y si NPAG = 199 imprimirá página 200. Tal parámetro se justifica cuando los impresos representan la continuación de un estudio anterior.
NUV	11-16	Especifica (de la misma forma que NPAG) el valor inicial de la numeración de las variables; Así: si NUV = 0 imprimirá primera variable = 1, y NUV = 500 imprimirá primera variable = 501. NUV representa un control general de numeración de las variables, y se destina a continuación de análisis anteriores. Si se omite, asume 0 y primera variable = 1.

2. Tarjeta TITULO del trabajo

Es obligatoria. Para cada página impresa se imprime el título del trabajo. Existe una tarjeta TT para cada clase de tarjeta de la encuesta.

Forma General:



Descripción de la Tarjeta TITULO del trabajo:

Variable	Columna	Descripción
TIT	1-80	80 caracteres alfanuméricos que puede incluir símbolos especiales. El título encabezará todas las páginas de los marginales con relación a una clase de tarjeta.

3. Tarjeta de parámetro principal (DATA)

Es obligatoria. En esta tarjeta se perforan los datos fundamentales de cada clase de tarjeta. Existe una para cada clase de tarjeta. Los datos codificados siempre se refieren a la clase de tarjeta en estudio.

Forma General:

DATA	NV	NTC	NVE	NCAS	NFM	NCDT	NEXT

Descripción de la tarjeta de parámetro principal (DATA)

Variable	Columna	Descripción
DATA	1-4	Se codifica la palabra DATA
NV	6-7	Número de variable que se marginalizará en la clase de tarjeta en estudio.
NTC	12	Define si se desean o no marginales con nombres de las categorías (que serán codificadas en las tarjetas NOMBRES DE CATEGORIAS), dos posibilidades: blanco - no hay tarjetas NOMBRE DE CATEGORIAS 1 - hay TARJETAS NOMBRE DE CATEGORIAS
NVE	14-15	Número de variables con exclusión de categorías de los cálculos estadísticos (media, mediana, desviación estándar). Las informaciones sobre cuales categorías de cada variable se van a excluir, serán codificadas en las tarjetas de EXCLUSION.
NCAS	17-23	Número de casos, número de tarjetas de datos pertenecientes a una misma clase de tarjeta. Cuando hay más de una clase de tarjeta, NCAS debe estar preciso para evitar que una clase de tarjeta se lea como otra clase. Cuando hay solamente una clase de tarjeta, NCAS debe ser mayor o igual al número de casos reales.
NFM	25	Número de tarjetas de FORMATO. Para cada encuesta, y para cada clase de tarjeta, varía el formato de ubicación de las variables, por lo tanto, es necesario indicar la disposición de las variables que serán analizadas. Los formatos de las variables serán codificados en las tarjetas de FORMATO; en la columna 25 se debe codificar el número de tarjetas de FORMATO necesarias.

Variable	Columna	Descripción
NCDT	27-28	Número de la clase de tarjeta en estudio
NEXT	30	Dos posibilidades:
		blanco - indica que la numeración de las variables continúa en forma creciente a partir de la última variable de la última clase de tarjeta; si se trata de la primera clase, la numeración es a partir de NUV (codificada en DATA)
		1 - indica que la numeración de las variables de esa clase de tarjeta empieza del 1.
NVC	32-33	Número de variables cuyos marginales serán condicionados por otras variables (sólo para MARBINCO).

4. Tarjetas de CONDICION de variable

Deberán colocarse a continuación de la tarjeta DATA, tantas veces como sea el valor indicado en NVG de la tarjeta DATA. Es posible condicionar una variable por hasta 3 variables, indicando el rango de categoría que hace incluir estos casos en el marginal condicionado de la variable.

Forma General:

NV NVC1 CI1 CS1 NVC2 CI2 CS2 NVC3 CI3 CS3

Descripción de la tarjeta control de variable:

Variable	Columna	Descripción
NV	1-2	Número de orden que se dará a la variable controlada debe ser el siguiente a la última variable no controlada y así se sigue aumentando
NVC1	4-5	Número de orden de la primera variable de control a usar
CI1	7-8	Límite inferior de categoría de la variable (NVC1) que se incluirá como caso de interés
CS1	10-11	Límite superior de categoría de la variable (NVC1) que se incluirá como caso de interés
NVC2	13-14	Número de orden de la segunda variable de control a usar
CI2	16-17	Límite inferior de categoría de la variable 2 (NVC2) que se incluirá como caso de interés
CS2	19-20	Límite superior de categoría de la variable 2 (NVC2) que se incluirá como caso de interés

Variable	Columna	Descripción
NVC3	22-23	Número de orden de la tercera variable de control a usar
CI3	25-26	Límite inferior de categoría de la variable 3 (NVC3) que se incluirá como caso de interés
CS3	28-29	Límite superior de la categoría de la variable 3 (NVC3) que se incluirá como caso de interés

5. Tarjeta de FORMATO

Es obligatoria. Existe para cada clase de tarjeta. Como el programa es general, se necesitan informaciones sobre la disposición de las variables que serán leidas (columna en que está la variable, número de columnas que ocupa, etc.)

5.1. FORMATO MARGBCD

Forma General:

(NNCCD,NNCCD,NNCCD,NNCCD,NNCCD,.....NNCCD)

Descripción de la Tarjeta de Formato

Columna	Descripción
1	Debe dejarse siempre en blanco
2	Debe perforarse un símbolo "("
3-7	A partir de la columna 3, empiezan los conjuntos NNCCD. Cada conjunto NNCCD se refiere a una o más columnas. NN = indica la repetición de variables de igual número de dígitos colocados en forma continua CC = columna que ocupa la primera variable del conjunto NNCCD. Cuando la variable ocupa más de una columna, se codifica la primera; cuando hay variables seguidas de igual formato y se usa el índice de repetición, se debe colocar la columna inicial de la primera variable. D = número de dígitos de la(s) variable(s) o número de columnas que ocupa(n) la(s) variable(s), es decir, 1 ó 2.
8	Siempre perforar un símbolo ",", continúa igual que 3-7. Se describe el diseño de la tarjeta de izquierda a derecha, según las variables que sean de interés

El conjunto NNCCD se repite cuantas veces sean necesarias para representar todas las variables (columnas 6-7 de la tarjeta DATA). Cuando hay más de una tarjeta de FORMATO, se empieza en la columna 1 de la tarjeta siguiente. Los conjuntos NNCCD son separados por una coma. Se prohíben blancos y se debe terminar el formato con paréntesis derecho ")".

5.2. FORMATO MARBINCO

Forma General:

(CCCCDCCCCDCCCCDCCCCDCCCCDCCCC)

Descripción de la tarjeta de formato

Columna	Descripción
1	Debe perforarse siempre un "("
2-5	Número del byte correspondiente a la primera variable
6	"," ó "-"
7-10	Número del byte de la segunda variable, según el delimitador usado en la sexta columna, etc.
11	"," ó "-"
12-15	Campo de definición
16	Delimitador
17-20	Campo de definición
21	Delimitador
22-25	Campo de definición
26	Delimitador
27-30	Campo de definición
31	Delimitador
⋮	
76	Delimitador
77-80	Campo de definición

Podemos utilizar como delimitador los símbolos "," y "-", al usar "," se trata de variables simples definidas independientemente, al usar el símbolo "-" se consideran las variables contiguas desde el byte antes de guión hasta el byte después de guión. Se termina el formato con "," y no se permiten blancos entre los paréntesis.

NOTA IMPORTANTE SOBRE LA TARJETA FORMATO

Para MARGBCD el formato debe contener especificaciones para un número de variables igual al valor codificado en NV de DATA (columnas 6-7). Para MARBINCO el formato puede contener más variables que el número especificado en NV de DATA (columnas 6-7). En este caso, las NV primeras variables del formato corresponden a las que se calcularán marginales. Las restantes se usan para condicionar los marginales de las primeras.

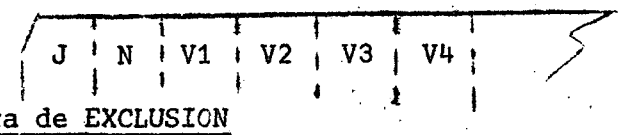
En ambos programas las columnas pueden ser especificadas en cualquier orden, pero se debe observar que este determina automáticamente el número de orden de la variable.

Una misma columna puede ser especificada más de una vez, pero en cada caso cambia el número de orden según su posición en el formato.

6. Tarjetas de EXCLUSION (uso opcional)

Las tarjetas de exclusión contienen informaciones sobre las categorías que no participan de los cálculos estadísticos. Para cada variable se pueden excluir hasta cuatro categorías. El total de tarjetas de exclusión por clase de tarjeta (nótese que para cada variable con exclusión existe una tarjeta de exclusión) debe ser su número total igual al valor codificado en el campo NVE de la tarjeta principal de control.

Forma General:



Descripción de la tarjeta de EXCLUSION

Variable	Columna	Descripción
J	1-2	Número de orden de la variable dentro de la clase de tarjeta en estudio
N	4	Número de categorías que se excluirán de los cálculos estadísticos para la variable en cuestión
V1	6-7	Primera categoría a excluir
V2	9-10	Segunda categoría a excluir (si la hay)
V3	12-13	Tercera categoría a excluir (si la hay)
V4	15-16	Cuarta categoría a excluir (si la hay)

7. Tarjeta de NOMBRE DE VARIABLE

Es obligatoria. Debe ir tantas veces como sea el valor codificado para NV en la tarjeta de parámetro principal DATA. La tarjeta NOMBRE DE VARIABLE contiene las informaciones para los títulos de identificación de la variable (nombre, ubicación, etc.)

Forma General:

NV	NC	NCS	NCAT	NOM

Descripción de la tarjeta NOMBRE DE VARIABLE

Variable	Columna	Descripción
NV	1-2	Esta variable es opcional, ya que no la lee la computadora, pero es recomendable que se codifique aquí el número de orden de la variable dentro de la clase de tarjeta en estudio
NC	4-5	Primera columna en que está perforada la variable (cuando hay más columnas, se codifica la primera)
NCS	7	Número de columnas que ocupa la variable (1 ó 2)
NCAT	9-11	Número de categorías con nombre; cuando no se usan nombres de categorías para la variable en estudio, se deja en blanco.
NOM	13-36	Nombre de la variable; existen 24 espacios reservados para el nombre de la variable.

8. Tarjeta NOMBRE DE CATEGORIA (uso opcional)

Cuando se desea que en la salida aparezcan nombradas las categorías (NTC=1), en la tarjeta se deben codificar las tarjetas nombres de categorías para aquellas variables en las cuales la tarjeta nombre de variable contiene NCAT $\neq$ 0.

Forma General:

L	NOMCA	L	NOMCA	L	NOMCA

Descripción de la tarjeta NOMBRE DE CATEGORIA

Variable	Columna	Descripción
L	1-2	Código de la categoría
NOMCA	4-19	Nombre de la categoría
L	21-22	Código de la categoría
NOMCA	24-39	Nombre de la categoría
L	41-42	Código de la categoría
NOMCA	44-59	Nombre de la categoría

Nota: Cada tarjeta de nombre de categoría acepta hasta tres categorías, luego el número de tarjetas de este tipo dependerá del número total de categorías para los cuales se dará nombre, es decir, depende del valor de NCAT en la tarjeta NOMBRE DE VARIABLE.

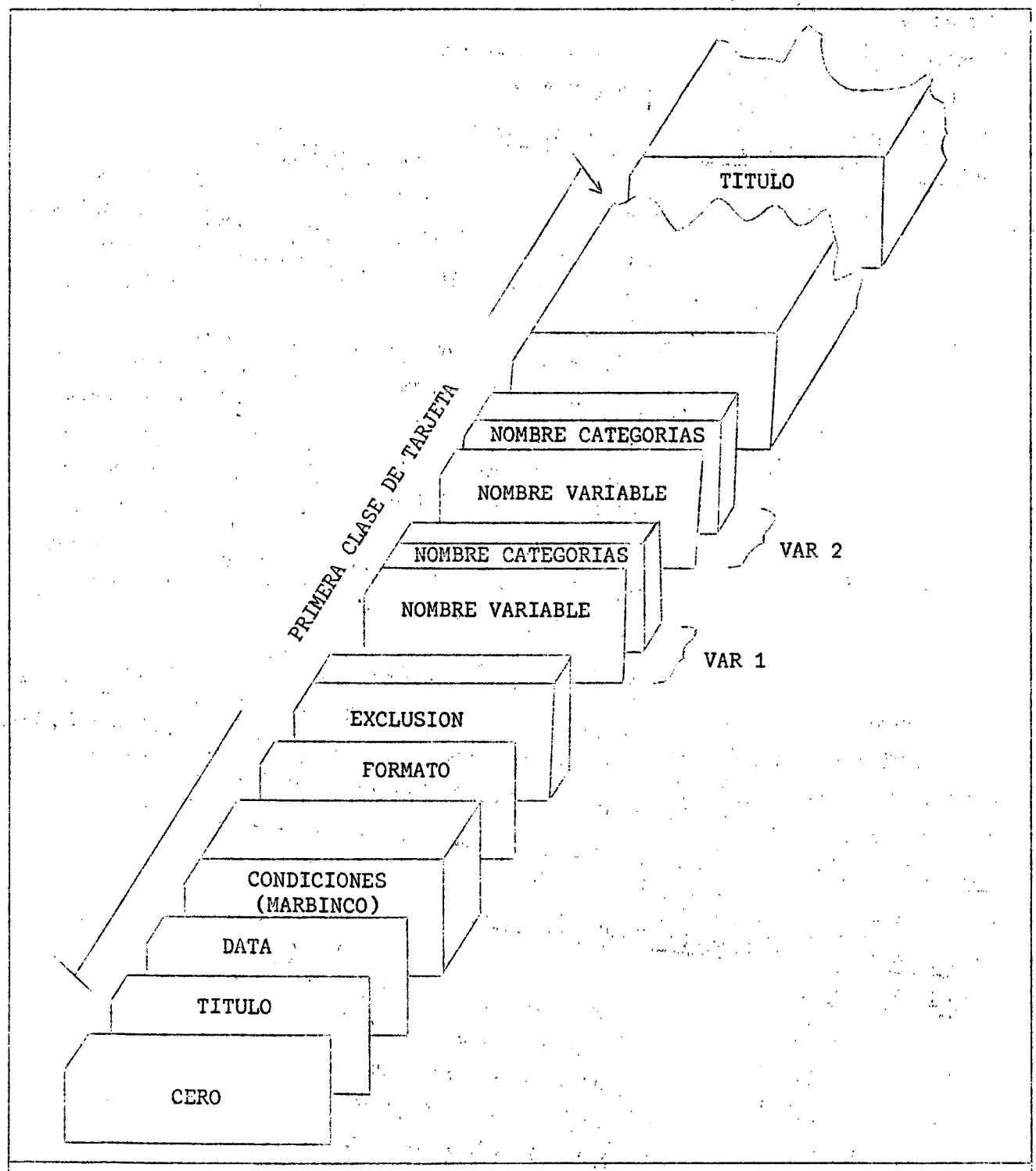


Figura 6: Orden de las tarjetas de control para MARGBCD y MARBINCO.

IV. PROGRAMA MARGINAL

El programa MARGINAL calcula marginales de sólo una variable a la vez. Las principales limitaciones son:

largo máximo del registro: 600 bytes

largo máximo de la variable: caracteres - 99 bytes

binario - 4 bytes

número máximo de variables condicionantes: 4

Las tarjetas de control son: TITULO, PRINCIPAL y CONDICION (opcional). La figura 7 presenta el orden en que deben aparecer.

1. Tarjeta TITULO

Es obligatoria. Para cada página de salida se imprime el título del trabajo.

Columnas	Descripción
1-80	80 caracteres alfanuméricos

2. Tarjeta PRINCIPAL

Es obligatoria. Especifica los parámetros principales del marginal deseado.

Columnas	Descripción
1-4	Columna inicial de la variable en el registro de entrada
5	,
6-7	Largo de la variable
8	,
9-10	Número de líneas de impresión (asume 90 líneas)
11	,
12-18	Tipo de la variable: en blanco - variable en caracteres, caso contrario, en binario
19	,
20-25	Número de registros. En blanco - se procesa hasta fin del archivo
26-80	Comentario

3. Tarjetas de CONDICION

Es opcional. Para cada condición hay una tarjeta. (máximo 4 tarjetas).

Columnas	Descripción
1-4	Columna inicial de la variable condicionante
5	,
6-7	Largo de la variable
8	,
9-13	Límite inferior de la variable condicionante
14	,
15-19	Límite superior de la variable condicionante
20-80	Comentario

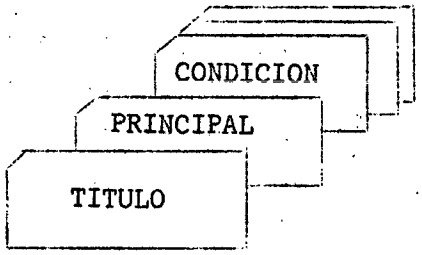


Figura 7: Orden de las Tarjetas.

Ejemplos:

1) col.1

```
** MARGINAL DE LA VARIABLE OCUPACION **
0017,02,55,BINARIO,005000
/*
```

En este ejemplo la variable ocupación (cols. 17-18) está en binario y se calcularán los marginales para los 5000 primeros registros y los resultados serán impresos en 55 líneas por página.

2)

```
** MARGINAL DE OCUPACION PARA PEA 10 AÑOS Y MAS **
0032,04
0011,02,00010,00099      EDAD ≥ 10
0014,01,00001,00002      TIPO ACT = 1,2
/*
```

La variable ocupación (columnas 32-35) está en caracteres y los marginales serán calculados para la población de edad (columnas 11-12) mayor o igual a 10 años y que sea económicamente activa (columna 14, códigos 1 ó 2). Será considerado el archivo completo y los resultados se imprimirán con 90 líneas por página.

Nota: Para procesar en DOS se debe especificar en la tarjeta parámetro DTF, el largo del bloque y del registro que tiene el archivo de entrada.

Formato de la tarjeta DTF.*

Columnas	Descripción
1-3	Debe ir la palabra DTF
5-8	Largo del bloque con un máximo de 4000 bytes. Debe ajustarse a la derecha del campo
9-12	Largo del registro con un número de 600 bytes. Debe ajustarse a la derecha del campo.

*Esta tarjeta debe ser la primera de todas las tarjetas parámetros. Si no se incluye asume los siguientes valores:

1) Largo del bloque = 4000 bytes
2) Largo del registro = 80 bytes

V. MARGBCD32 Y MARBINCO32

Estos programas son muy simplificados y sólo obtienen frecuencias de categorías sin incluir ningún tipo de porcentajes.

Están programados enteramente en lenguaje ASSEMBLER de IBM y ocupan 17327 bytes con la configuración que tiene. Esta configuración presenta las siguientes limitaciones:

1. Soporta un máximo de 30 variables de 1 ó 2 dígitos como máximo.
2. Número de casos o registros, máximos a leer: 999999.
3. Número de categorías por variable: puede aceptar de 00 a 99 (numérico); más, "casos en blanco", "casos con perforación x(-)", "casos con perforación y (&)" y "otros" (signos distintos a los anteriores que caen en un sólo bolsón).
4. Los registros pueden tener hasta un máximo de 80 bytes cada uno.

El programa funciona a base de tarjetas de control, que contienen los parámetros necesarios para identificar cada caso en particular.

1. Tarjeta de Título - (CT-1) (Clase de Tarjeta 1)

Es la primera tarjeta que debe ir siempre, tiene una "clase de tarjeta 1" y sirve para colocar el título o encabezamiento que va a salir en cada página del listado de resultado.

Formato de esta tarjeta

Cols.	Nombre del campo	Descripción
1	CT	Clase de tarjeta (debe ir un 1)
2	-	En blanco
3-80	TIT	Caracteres alfanuméricos que definen el título antes mencionado

2. Tarjeta de parámetros - (CT-2)

Debe ser la segunda tarjeta que siempre debe ir y da los valores que controlan el programa.

Formato de esta tarjeta

Cols.	Nombre del campo	Descripción
1	CT	Clase de tarjeta (debe ir un 2)
2	-	En blanco

Cols.	Nombre del campo	Descripción
3-4	NV	Número de variables (≤30) de las cuales se desea sacar frecuencia marginal
6-11	NCAS	Número de casos. Acepta un máximo de 999999. Si se pone un número menor de dígitos se debe rellenar a la izquierda del campo con ceros (Ej. 004552). Si no se conoce el número exacto de casos, basta con poner un número mayor para que el programa se detenga por fin de archivo.
12	-	En blanco
13	NFM	Número de tarjeta clase 3 (CT3) con formato del registro (NFM ≤ 2), que se incluirán en esa corrida del programa.
14	-	En blanco
15-16	LINES	LINES 90 sale 1 variable por página. LINES 91 salen 2 variables por página.

3. Tarjeta de formato - (CT-3)

Debe ser la tercera tarjeta obligatoriamente y no pueden ir más de 2 de estas tarjetas.

Formato de esta Tarjeta

Cols.	Nombre del campo	Descripción
1	CT	Clase de tarjeta (debe ir un 3)
2	-	En blanco
3	NNCCD	NN: número de veces que se repite el campo D. CC: primera columna que marca el primer campo D. D: número de dígitos de cada campo. Puede ser 1 ó 2.
9	-	Debe ir una coma (,)
10-14	NNCCD	Se repite lo mismo que de 4-9. Para terminar el formato debe ir un paréntesis derecho -)-. No pueden ir blancos en el medio. Si el campo NFM tiene un 2, deben ir 2 tarjetas de formato, en que la segunda tarjeta continúa a partir de la

columna 4 como si fuera la columna 81 de la primera tarjeta. Además, la segunda tarjeta de formato debe llevar también un 3 en la columna 1.

Nota: Debe comprobarse que el número de campos, definido a través del formato, sea igual al número de variables indicado en la CT-2.

Estos tres tipos de tarjetas deben ir siempre. Si el programa encuentra algún error en algún campo de estas tarjetas de control, se cancela la corrida con mensajes de error.

Si no hay errores en las primeras tarjetas de control, el programa lee desde SYS008 los casos hasta encontrar o los NCAS registros o el fin de archivo.

4. Tarjeta nombre de variable

En la etapa de impresión se leen los nombres de las variables "en el mismo orden en que se especificaron en el formato, (CT-3)". Las tarjetas de variables deben ser tantas como el valor NV. El formato de estas tarjetas es:

Cols.	Descripción
1-2	Número de orden de la variable (01-02, etc.) dentro de la definición del formato (CT-3)
3	En blanco
4-27	24 caracteres para el nombre de la variable
28-80	En blanco.

La forma de operar el programa es sumamente sencilla y se describe en la figura siguiente (fig. 8).

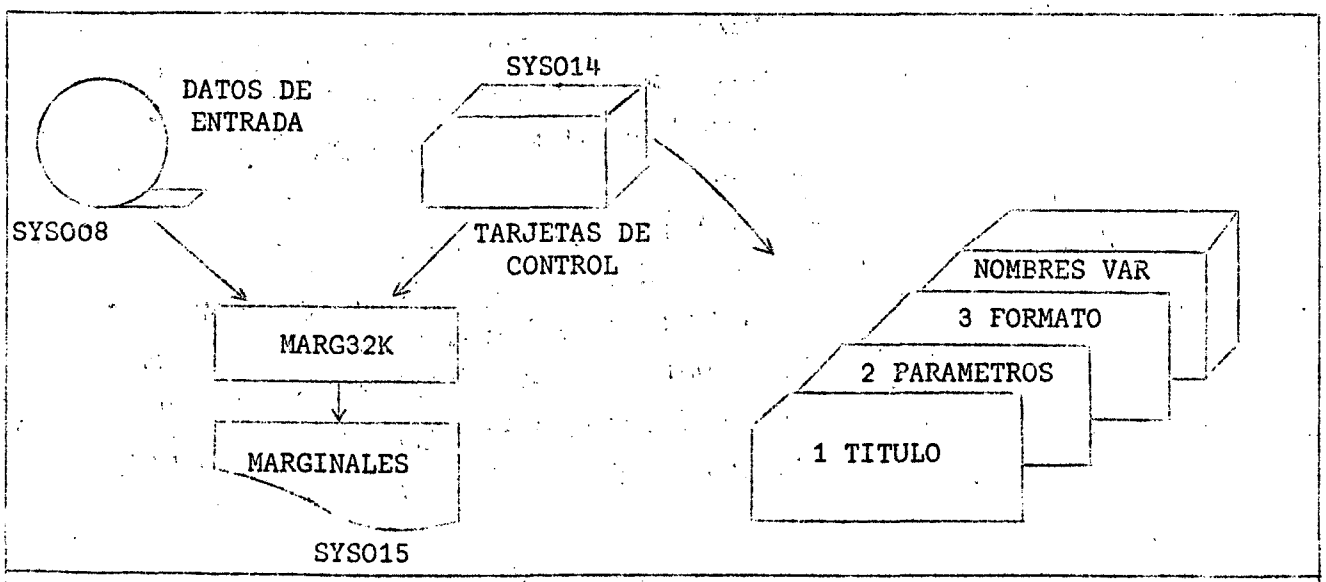


Figura 8: Diagrama de Proceso y Orden de las Tarjetas.

Se utiliza una cinta magnética cuya dirección lógica en el programa es SYS008 (hay que cambiar los parámetros RECSIZE y BLKSIZE de la DTFMT y además la IOAREAL al valor dado).

Se utiliza una lectora de tarjetas cuya dirección lógica en el programa es SYS014 (esta lectura se hace mediante EXP, luego hay que buscar la CCB con nombre LEC).

También se utiliza una impresora de línea de 132 posiciones cuya dirección lógica en el programa es SYS015. Esta impresión se hace mediante EXP, luego hay que buscar la CCB con nombre IMP.

Si se desea ampliar el programa para un número mayor de variables, es necesario modificar el valor de la DC FREC en la instrucción 783. El valor actual es 3120 y se calculan como 104 categorías por 30 variables = 3120 posiciones. Además, se debe cambiar el control de NV en las instrucciones 47 y 53.

CENTRO LATINOAMERICANO DE DEMOGRAFIA
CELADE: J.M. Infante 9. Casilla 91. Teléfono 257806
Santiago (Chile)
CELADE: Ciudad Universitaria Rodrigo Facio
Apartado Postal 5249
San José (Costa Rica)