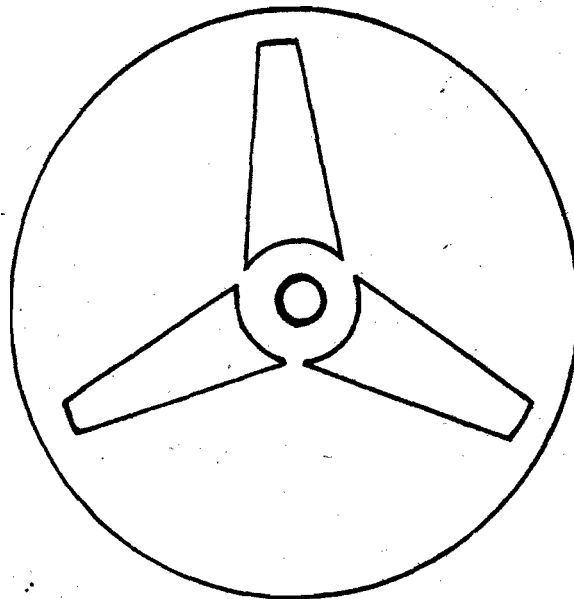


CENTRO LATINOAMERICANO DE DEMOGRAFIA
CELADE - SAN JOSE
NACIONES UNIDAS

LICENCIATURA CENTROAMERICANA
EN SOCIOLOGIA
CSUCA - UNIVERSIDAD DE COSTA RICA



SPSS: UN LENGUAJE ORIENTADO AL ANALISIS DE DATOS

(Statistical Package for the Social Sciences,
Universidad de Chicago)

WALTER VUOLO S.



2761.00
(06834)

CENTRO LATINOAMERICANO DE DEMOGRAFIA
CELADE - SAN JOSE
NACIONES UNIDAS

LICENCIATURA CENTROAMERICANA EN SOCIOLOGIA
CSUCA - UNIVERSIDAD DE COSTA RICA

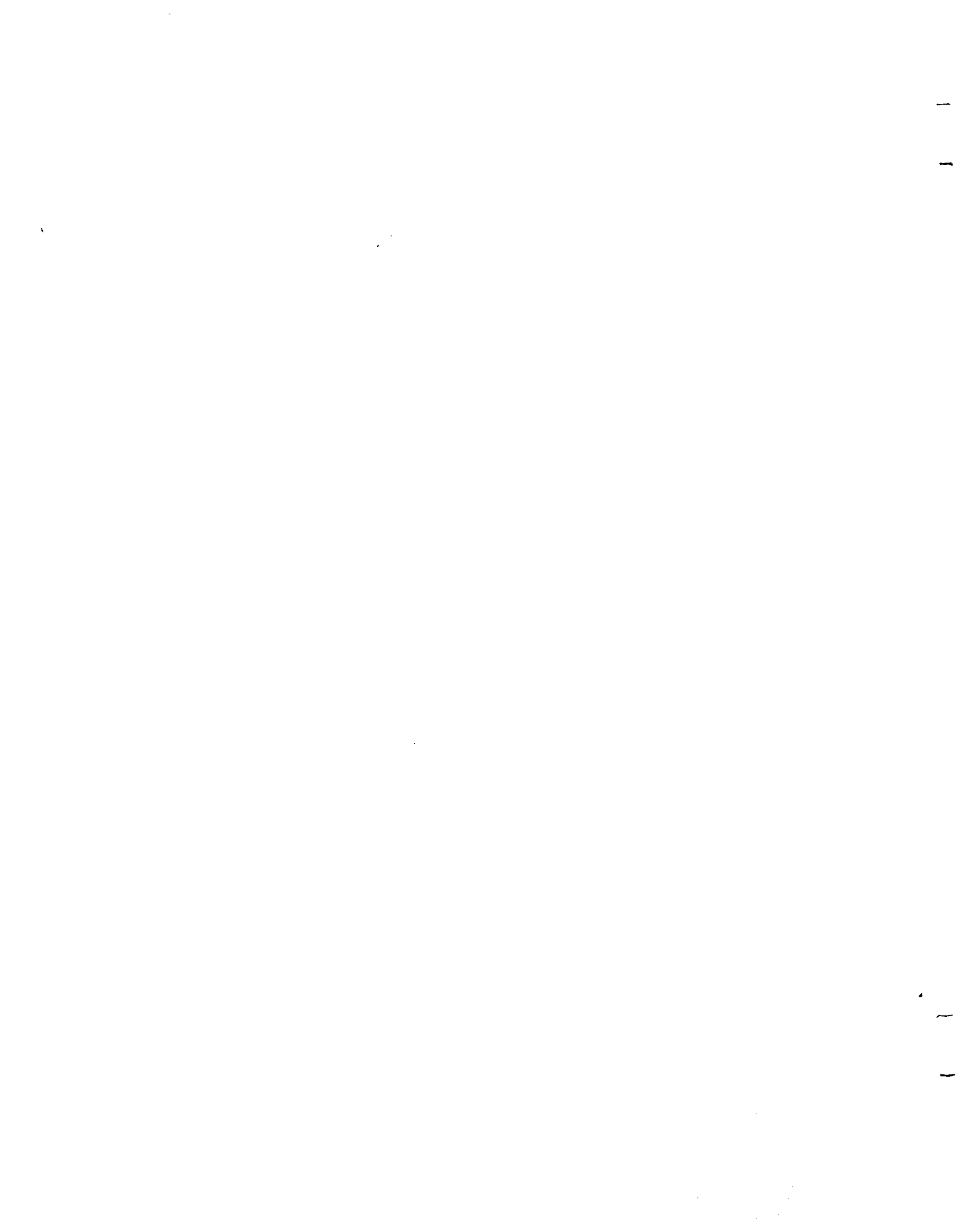
SPSS: UN LANGUAGE ORIENTADO AL ANALISIS DE DATOS

(Statistical Package for the Social Sciences,
Universidad de Chicago)

Cátedra: Metodología IV. - Introducción al
Procesamiento Electrónico
de Datos para Cientistas
Sociales

Prof. Valter Vuolo S.

San José, Costa Rica
Setiembre, 1976



INDICE GENERAL

		Página
I.	UNA VISION GLOBAL DEL SPSS	2
II.	LA ORGANIZACION DE DATOS PARA LA ENTRADA AL SPSS	12
III.	TARJETAS DE CONTROL SPSS	20
IV.	LAS TARJETAS DE DEFINICION DE DATOS	27
V.	LAS TARJETAS DE DEFINICION DE TAREA	48
VI.	LAS TARJETAS DE PROCESO "RUN CARDS"	53
VII.	GENERACION Y PROCESAMIENTO DE ARCHIVOS CON SPSS.	57
*	APENDICE A: El cambio a la versión 6.	68
VIII.	LAS TARJETAS DE MODIFICACION DE DATOS	76
IX.	LAS TARJETAS DE SELECCION DE DATOS	113
X.	LA TARJETA EDIT: chequeo de sintaxis	119
XI.	ESTADISTICAS DESCRIPTIVAS Y DISTRIBUCION DE FRE- CUENCIAS	122
XII.	CROSSTABS: tabulaciones cruzadas y medidas de asociación	140
XIII.	BREAKDOWN: descripción de subpoblaciones	157
XIV.	ANALISIS DE CORRELACION BIVARIADA: PEARSON CORR NONPAR CORR, SCATTERGRAM	170
XV.	PARTIAL CORR: correlación parcial	192
XVI.	REGRESSION: Regresión Múltiple	208
*	APENDICE 1. TARJETAS DE J.C.L.	232

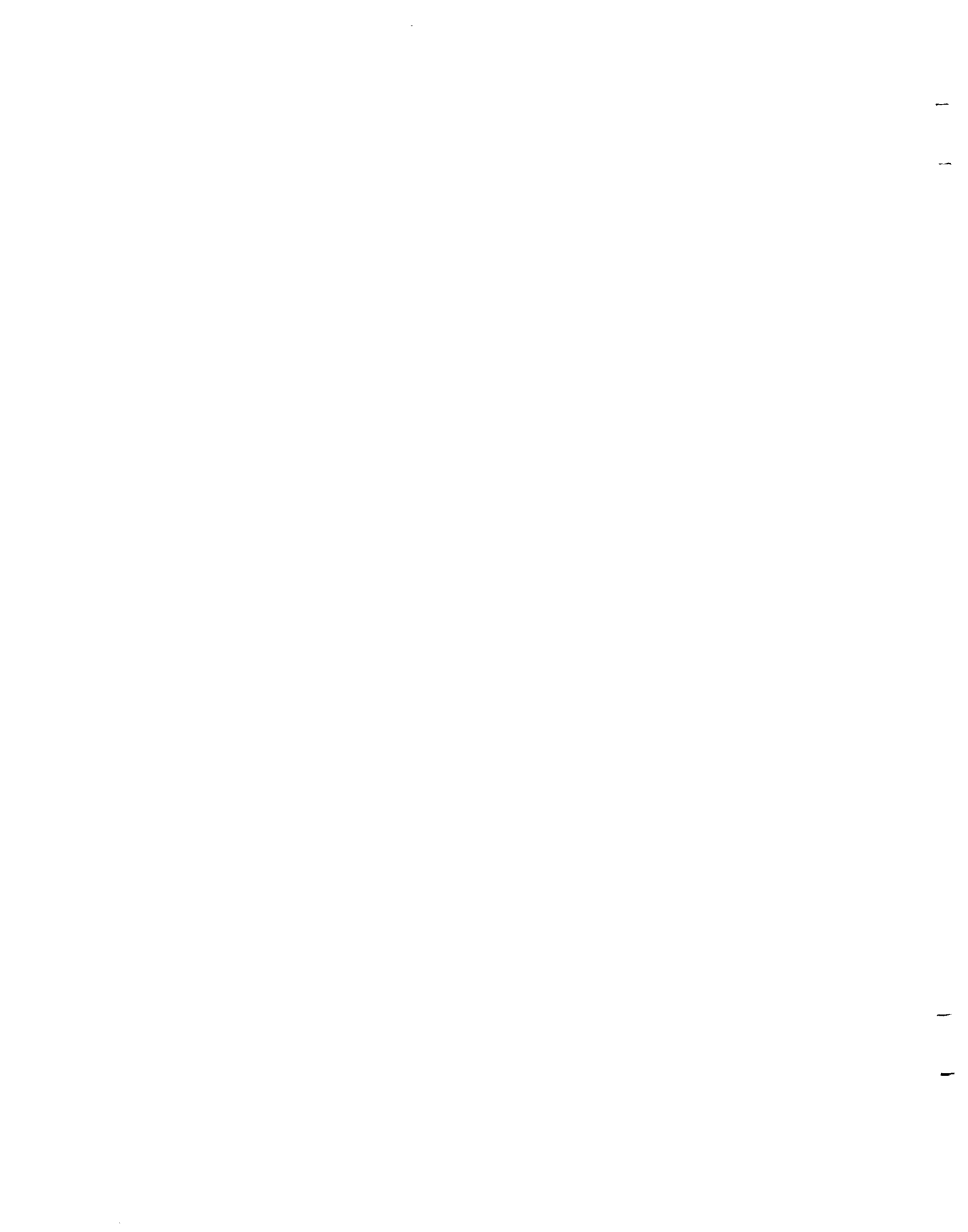
INDICE DEL CONTENIDO

	Página	
I.	UNA VISION GLOBAL DEL SPSS	2
	I.1 Los procedimientos estadísticos	2
	I.2 Características del SPSS	6
II.	LA ORGANIZACION DE DATOS PARA LA ENTRADA AL SPSS .	12
	II.1 El caso como unidad de análisis	12
	II.2 Convenciones de codificación de datos	12
	II.3 Subarchivos: función y estructura	17
	II.4 Ingreso de datos	19
	II.5 Limitaciones sobre la entrada de datos	19
III.	LAS TARJETAS DE CONTROL SPSS	20
	III.1 El formato general	20
	III.2 El campo de control	20
	III.3 El campo de especificación	21
	III.4 Normas generales	25
IV.	LAS TARJETAS DE DEFINICION DE DATOS	27
	IV.1 FILE NAME	28
	IV.2 VARIABLE LIST	28
	IV.3 SUBFILE LIST	32
	IV.4 # OF CASES	33
	IV.5 INPUT MEDIUM	34
	IV.6 INPUT FORMAT	35
	IV.7 MISSING VALUES	43
	IV.8 VAR LABELS	44
	IV.9 VALUE LABELS	45
	IV.10 PRINT FORMATS	46
	IV.11 El orden de las tarjetas de definición de datos	47
V.	LAS TARJETAS DE DEFINICION DE TAREA	48
	V.1 Tarjetas de procedimientos	48
	V.2 La tarjeta OPTIONS	49
	V.3 La tarjeta STATISTICS	49
	V.4 La tarjeta READ INPUT DATA	50
	V.5 La tarjeta PROCESS SBFILES	51
	V.6 Palabras de control de definición de tarea .	52
VI.	LAS TARJETAS DE PROCESO	53
	VI.1 RUN NAME	53
	VI.2 FINISH	54
	VI.3 KEYPUNCH	54
	VI.4 PRINT BACK	54
	VI.5 NUMBERED	54
	VI.6 COMMENT	55
	VI.7 DOCUMENT	55
	VI.8 DUMP	55
	VI.9 El orden de las tarjetas de proceso	56

	Página
VII.	GENERACION Y PROCESAMIENTO DE ARCHIVOS 57
	VII.1 Procesamiento de archivos del usuario 57
	VII.2 SAVE FILE: generación de archivos SPSS .. 61
	VII.3 Proceso de archivos SPSS 63
*	APENDICE A: El cambio a la versión 6 68
	1. Introducción 68
	2. Tarjetas de definición de datos 68
	3. Tarjetas de definición de tarea 71
	4. Tarjetas de proceso 72
	5. Generación y procesamiento de archivos ... 74
	6. WRITE FILEINFO reemplaza a DUMP 75
VIII.	LAS TARJETAS DE MODIFICACION DE DATOS 76
	VIII.1 La tarjeta RECODE 76
	VIII.2 COMPUTE: transformación incondicional ... 82
	VIII.3 IF: transformación condicional 86
	VIII.4 Limitaciones de COMPUTE e IF 93
	VIII.5 COUNT: creación de índices aditivos 93
	VIII.6 ASSIGN MISSING 98
	VIII.7 Inicialización de variables creadas con IF 102
	VIII.8 Declaraciones: DO REPEAT: END REPEAT ... 103
	VIII.9 ALLOCATE: asignación de memoria 106
IX.	TARJETAS DE SELECCION DE DATOS 113
	IX.1 SAMPLE, *SAMPLE: muestra al azar 113
	IX.2 SELECT IF, *SELECT IF: selección de casos 114
	IX.3 WEIGHT, *WEIGHT: ponderación 114
	IX.4 El orden de las tarjetas 117
X.	LA TARJETA EDIT: chequeo de sintaxis 119
	X.1 Funciones de EDIT 119
	X.2 El acceso a archivos 120
XI.	ESTADISTICOS DESCRIPTIVOS Y DISTRIBUCION DE FRE - CUENCIAS 122
	XI.1 Introducción 122
	XI.2 CONDESCRIPTIVE: estadísticos disponibles. 122
	1. Estadísticos disponibles 123
	2. Generación de archivo con variable Z.. 124
	3. Opciones 125
	4. Limitaciones 126
	5. Ejemplos de salida 127
	XI.3 FREQUENCIES: distribución de frecuencias. 131
	1. El modo GENERAL 132
	2. El modo INTEGER..... 132
	3. Estadísticos 133
	4. Opciones 134
	5. Limitaciones 135
	6. Ejemplos de salida 137

XII.	CROSSTABS: TABULACIONES CRUZADAS Y MEDIDAS DE ASOCIACION	140
XII.1	CROSSTABS: modos General, Integer	140
XII.2	CROSSTABS: modo General	141
XII.3	CROSSTABS: modo Integer	144
XII.4	Los VALUE LABELS e inclusión de missing ...	146
XII.5	Estadísticos	146
XII.6	Opciones	147
XII.7	Limitaciones	148
XII.8	Ejemplos de salida	151
XIII.	BREAKDOWN: DESCRIPCION DE SUBPOBLACIONES	157
XIII.1	El subprograma BREAKDOWN	157
XIII.2	BREAKDOWN: modo General	158
XIII.3	BREAKDOWN: modo Integer	159
XIII.4	Estadísticos	160
XIII.5	Opciones	161
XIII.6	Limitaciones	162
XIII.7	CROSSBREAK: salida en forma tabular	164
XIII.8	Ejemplos de salida	166
XIV.	Análisis de correlación:: PEARSON CORR, NONPAR CORR, SCATTERGRAM	170
XIV.1	El subprograma PEARSON CORR	170
	1. La tarjeta de procedimiento	171
	2. Opciones	172
	3. Estadísticos	174
	4. Limitaciones	175
	5. Ejemplos de salida	176
XIV.2	El subprograma NONPAR CORR	178
	1. La tarjeta de procedimiento	179
	2. Opciones	179
	3. Estadísticos	179
	4. Limitaciones	179
	5. Ejemplos de salida	181
XIV.3	El subprograma SCATTERGRAM	183
	1. La tarjeta de procedimiento	183
	2. Opciones	185
	3. Estadísticos	186
	4. Limitaciones	186
	5. Ejemplos de salida	187
XV.	PARTIAL CORR: CORRELACION PARCIAL	192
XV.1	La tarjeta de procedimiento	192
XV.2	Convenciones para ingresar matrices	194
XV.3	Opciones	199
XV.4	Estadísticos	201
XV.5	Limitaciones	201
XV.6	Ejemplos de salida	203

	Página
XVI. REGRESSION: ANALISIS DE REGRESION MULTIPLE ...	208
XVI.1 La tarjeta de procedimiento	208
XVI.2 El nivel de inclusión: modalidad de cálculo	209
XVI.3 Parámetros para el criterio de inclusión	212
XVI.4 RESID: tratamiento de los missing	213
XVI.5 Convenciones para el ingreso de matrices	214
XVI.6 Opciones	215
XVI.7 Estadísticos	217
XVI.8 Limitaciones	219
XVI.9 Descripción de la salida	221
XVI.10 Ejemplos de salida	223
* APENDICE 1: TARJETAS DE J.C.L.	232



INTRODUCCION

En la evolución de los lenguajes de programación se puede citar como un hecho de gran importancia, el desarrollo de los llamados lenguajes orientados o de tercer nivel, que permiten al hombre una mayor facilidad en su comunicación con la máquina. Así, dentro del campo de las ciencias sociales, se dispone de varios lenguajes orientados tales como el SAS, OSIRIS y el SPSS que es el objeto de estos apuntes.

Podemos decir que la característica general de los lenguajes orientados es que permiten una disminución significativa en el tiempo hombre requerido para la solución de un problema, además de que permiten la obtención de resultados en corto plazo.

El grado de conocimientos especializados en computación requerido para capacitarse en el uso de los lenguajes orientados es variado. El SPSS en especial no requiere prácticamente ningún conocimiento especializado, cumpliendo así con uno de los objetivos de su diseño: permitir al investigador, que en general no es un especialista en computación, que pueda por sí mismo, realizar el tratamiento de sus datos; que es a la vez uno de los objetivos básicos de la Cátedra de Metodología IV.

El desarrollo del SPSS se inicia en 1965 en la Universidad de Stamford y tiene su continuidad en la Universidad de Chicago, junto al NORC (National Opinion Research Center) de dicha Universidad. En 1970, el SPSS era utilizado en alrededor de 60 instalaciones. Actualmente es utilizado en cerca de 600 instalaciones, incluyendo las conversiones a más de 20 distintos sistemas operativos.

La actual versión (release 6) trae prácticamente casi el doble de los procedimientos estadísticos disponibles en la versión de 1970, incluye una mayor potencialidad para el manejo de archivos tales como el SORT CASES, SAVE ARCHIVE, GET ARCHIVE, etc.

En este manual, redactado con el propósito de servir de texto a los estudiantes de licenciatura de la Escuela Centroamericana de Sociología, se incluyó solamente los procedimientos estadísticos e instrucciones para el manejo de archivos de uso más frecuente.

De esta forma, este manual no contempla las siguientes instrucciones de manejo de archivos: LIST FILEINFO; LIST CASES; WRITE CASES; DELETE VARS, KEEP VARS; ADD VARIABLES, ADD DATA LIST; ADD CASES, ADD SUBFILES; DELETE SUBFILES; REORDER VARS; SORT CASES; SAVE ARCHIVE, GET ARCHIVE; LIST ARCHINFO. Se excluyó la descripción de los siguientes procedimientos estadísticos: T-TEST, AGGREGATE, ANOVA (análisis de varianza); ONWAY; DISCRIMINANT; FACTOR (análisis factorial); CANCORR (correlación canónica); y GUTTMAN SCALE.

Para el uso de este manual es necesario que para el estudio de los capítulos I al VII, el lector se remita al Apéndice A, al final del Capítulo VII, pues estos primeros capítulos se refieren a la versión 5 del SPSS y a pesar de que sufrieron pocas modificaciones se requiere las especificaciones del apéndice para el uso de la versión 6. Esto se debe al hecho de que dichos capítulos fueron objeto de una edición preliminar para los estudiantes del 2º semestre de 1975, fecha en que todavía no se disponía aún de la versión 6 del SPSS. Por limitaciones de tiempo no se pudo redactar nuevamente dichos capítulos y se optó por especificar los cambios en el apéndice, ya que los cambios que sufrieron los temas ahí tratados fueron mínimos.



I. UNA VISION GLOBAL DEL SPSS

I.1 LOS PROCEDIMIENTOS ESTADISTICOS EN SPSS

I.1.1 Distribuciones de frecuencia en una dirección, medidas de tendencia central y dispersión

CONDESCRIPTIVE

- Para cálculo de estadísticas descriptivas
- Aplicación a variables numéricas continuas, que asumen gran número de categorías; por ejemplo el ingreso familiar.
- Estadísticas disponibles: media, error estándar, varianza, Kurtosis, Asimetría (SKEWNESS), rango, mínimo, máximo.

CODEBOOK

- Para computar distribuciones de frecuencias absolutas y relativas.
Aplicación a variables numéricas y alfanuméricas; recomendable para variables con pocas categorías.
- Estadísticas disponibles: media, error estándar, mediano, moda, desviación estándar, varianza, kurtosis, asimetría, rango, mínimo, máximo.
- Permite si se desea producir histogramas y salidas totalmente documentadas con títulos de categorías.

MARGINALS

- Cálculo de distribuciones de frecuencias absolutas y relativas.

- Aplicación a variables numéricas y alfanuméricas; recomendable para variables con gran número de categorías pues posee una salida impresa más condensada.
- Estadísticas disponibles: igual a CODEBOOK.
- No permite histogramas ni títulos de categorías.

FASTMARG

- Cálculo de frecuencias absolutas y relativas.
- Aplicación a variables numéricas enteras.
- Estadísticas disponibles: igual CODEBOOK.
- La salida impresa es igual a de CODEBOOK, pero no permite histogramas.
- Ventajas sobre el CODEBOOK:
 - Opción para producir salida en 8 1/2" x 11" para fotocopiar.
 - Opción que permite tener salida impresa en cinta o disco para posteriores re-ediciones.
 - La velocidad de procesamiento es varias veces mayor que la del CODEBOOK y permite un mejor aprovechamiento de la memoria disponible.

I.1.2 Cuadros de relaciones entre dos y más variables: Tabulación cruzada y descripción de subpoblaciones.

- Tabulaciones cruzadas

CROSSTABS: para variables numéricas y alfanuméricas.

FASTABS: sólo para variables numéricas y enteras.

Estadísticas:

- Chi cuadrado (para cuadros 2 x 2; texto exacto de FISCHER si hay menos de 21 casos; el chi cuadrado corregido de YATES se aplica a los demás cuadros de 2 x 2).
- PI para 2 x 2; Indicador de CRAMER V para cuadros más grandes.
- Coeficiente de Contingencia.
- Lambda simétrico y asimétrico.
- Coeficiente de incertidumbre simétrico y asimétrico.
- Indicador TAU B de Kendall.
- Indicador TAU C de Kendall.
- Gamma.
- Indicador D de Somer (simétrico y asimétrico).

- Descripción de subpoblaciones

BREAKDOWN: provee una técnica simple para examinar la media, desviación estándar y varianza de una variable dependiente, entre varios subgrupos de una población total.

La salida es análoga a una tabulación cruzada, sólo que para cada combinación de categorías de las variables involucradas computa la media, varianza, desviación estándar y frecuencia.

I.1.3 Análisis de correlación bivariada

Hay disponibles dos subprogramas:

- PEARSON CORR: produce coeficiente de correlación de orden cero del producto momento.
- NONPAR CORR: coeficientes de correlación no paramétrica. Permite al usuario computar correlaciones de orden de rango Spearman o Kendall o ambos.

Estos dos procedimientos permiten proveer las matrices de correlación para la entrada a cálculos posteriores.

I.1.4 Correlación parcial y regresión múltiple

- PARTIAL CORR: este programa puede operar con datos crudos o con matrices de coeficientes de correlación obtenidos por PEARSON CORR o NONPAR CORR. Permite computar hasta 5 órdenes de parciales -el usuario tiene el control total sobre las órdenes y parciales a obtener.
- REGRESSION: igual a PARTIAL CORR este subprograma puede operar con datos crudos o matrices de coeficientes de correlación. Este procedimiento permite al usuario controlar la forma en que desea la introducción de variables en el cálculo; puede forzar la inclusión de ciertas variables y después dejar que el procedimiento automáticamente seleccione aquellas que tienen alguna contribución significativa a la ecuación de regresión.

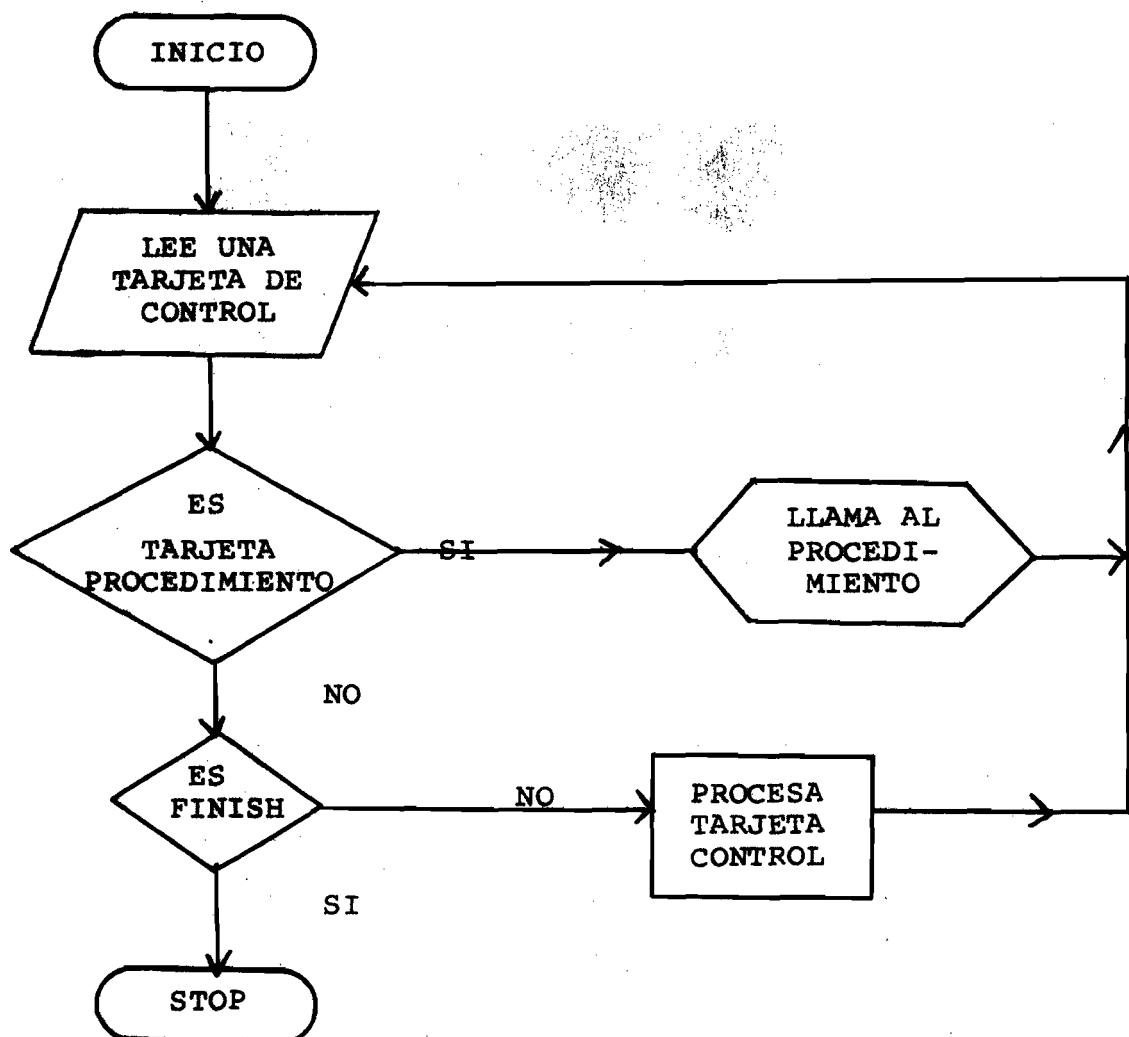
I.1.5 Escalas de GUTTMAN y Análisis Factorial

Existen además los procedimientos GUTMAN SCALE y FACTOR que no serán detallados en estos apuntes.

I.2 CARACTERISTICAS DEL SPSS

I.2.1 La secuencia de control

La secuencia de funciones del SPSS es determinada por las "tarjetas de control" y se muestra en el diagrama.



I.2.2 Entrada y Procesamiento de Datos

Los datos se organizan dentro del sistema SPSS en unidades llamadas archivos.

Hay dos formas de entrar datos al sistema:

- ARCHIVO DEL USUARIO

Son los datos del usuario en tarjetas o cintas en carácter (BCD). Para su proceso con el SPSS se debe dar por tarjetas de control SPSS la información asociada que define y describe esos datos. El ejemplo N° 1 a continuación muestra el proceso con un archivo del usuario y se guarda un archivo del sistema con SAVE FILE.

- ARCHIVO DEL SISTEMA SPSS

Está constituido por:

- Datos del usuario.
- Información asociada que definen y describen los datos.

El archivo del sistema SPSS es almacenado en binario y la rutina para la lectura es más eficiente que la rutina para la lectura del archivo del usuario. Esto implica una velocidad de proceso con archivo SPSS varias veces superior a la velocidad de proceso con un archivo del usuario (claro está que en el ejemplo que se muestra la diferencia en el tiempo de proceso sería insignificante debido al número de casos. No en tanto con archivos reales en los cuales el número de casos asciende de 1.000 a 30.000 o más casos la diferencia es significativa).

Además del archivo SPSS ser una entidad permanente de auto-documentación; hay que considerar el efecto potencial sobre la manera que el investigador actúa con sus datos en sus procesos diarios y también minimiza los posibles errores en las tarjetas de control SPSS.

EJEMPLO N° 1: Procesamiento con Archivo del Usuario

	1		16	
	RUN NAME		EJEMPLO DE PROCESO CON ARCHIVO DEL USUARIO	
	FILE NAME		ARCH1	
Tarjetas de definición de datos	VARIABLE LIST		NOMBRE, EDAD, SEXO, INGRESO, ORFAN	
	INPUT FORMAT		FIXED (A8, 2x, F2.0, 2x, F1.0, 2x, F5.0, 1x, F1.0)	
	PRINT FORMAT		NOMBRE (A)/EDAD TO ORFAN (0)	
	# OF CASES		10	
	INPUT MEDIUM		CARD	
	VAR LABELS		NOMBRE, NOMBRE DEL ENCUESTADO/ EDAD, EDAD SIMPLES/ SEXO, SEXO/ INGRESO, INGRESO FAMILIAR DEL ENCUESTADO/ ORFAN, CONDICION DE ORFANDAD DEL ENCUESTADO/ SEXO (1) HOMBRES (2) MUJERES/ ORFAN (1) HUERFANO (2) NO HUERFANO/ (9) IGNORADO	
	VALUE LABELS		ORFAN (9)/SEXO (3,8,9)	
	MISSING VALUES			
		CROSSTABS		EDAD BY SEXO BY ORFAN
	Tarjetas de definición de tarea	OPTIONS		3,5
STATISTICS			ALL	
	READ INPUT DATA			
Tarjetas de datos del usuario	JUAN	25	1 02300 1	
	JOSE	20	1 04570 2	
	MATILDE	27	2 04600 1	
	CLEOTILDE	01	2 00000 1	
	Enrique	30	1 01980 1	
	RICARDO	40	1 10600 9	
	NAPOLEON	89	1 05700 2	
	LUIS XV	50	1 08900 9	
	ROBESPIE	40	1 00030 9	
	JUDAS	60	1 89980 9	
	SAVE FILE			
	FINISH			

EJEMPLO N° 2: Procesamiento con Archivo del Sistema
(Archivo SPSS)

1	16
RUN NAME	PROCESO CON ARCHIVO SPSS
GET FILE	ARCH1
CROSSTABS	EDAD BY SEXO BY ORFAN
OPTIONS	EDAD
STATISTICS	ALL
FINISH	

I.2.3 Estructura de Sub-archivos

Los datos entrados al SPSS pueden ser estructu-
rados en sub-archivos. Una vez creada la estructura
de sub-archivos el manejo de estos puede ser: (1) se-
leccionando sub-archivos individuales, (2) combinando
sub-archivos o (3) ignorando la estructura. Por ejem-
plo para una encuesta a nivel nacional podemos gene-
rar una estructura de sub-archivos para cada provincia.
Al realizar digamos una tabulación cruzada obtendre-
mos los cuadros para cada provincia, o si deseamos pa-
ra un conjunto de provincias o también ignorando la
estructura tendremos cruces para el país.

I.2.4 Valores Faltantes

Por medio de la tarjeta MISSING VALUES se pue-
de definir hasta 3 valores faltantes para cada varia-
ble. Valores faltantes son aquellas categorías de
una cierta variable en que se registró "NO SABEN";
"NO RESPONDEM" o "IGNORADO". Cada procedimiento esta-
dístico de SPSS posee opción para incluir o excluir
los casos declarados como MISSING en el cálculo espe-
cífico que realiza.

I.2.5 Recodificación de datos

Es muy usual que el sistema de codificación utilizado en la recolección de los datos no sea lo más adecuado para la etapa de análisis. Por ejemplo para la variable sueldo en general para el análisis se desea ciertos rangos y no todas las posibilidades registradas. Por medio de la tarjeta RECODE podemos recodificar dicha variable a las categorías que interesan al análisis. Por ejemplo para el caso de desear análisis con grupos decenales de edades:

1	16					
RECODE	EDAD	(00 THRU 9 = 1)	(10 THRU 19 = 2)	(20 THRU 29 = 3)		
		(30 THRU 39 = 4)	(40 THRU 49 = 5)	(50 THRU 99 = 6)		

I.2.6 Transformación de variables

El SPSS permite lograr una gran variedad de transformaciones de variables por medio de declaraciones de variables por medio de declaraciones tipo FORTRAN.

Por ejemplo:

```
COMPUTE          Z = Y + X
```

es una transformación incondicional en que la variable creada Z tendrá el valor de Y + X.

Transformación condicional se obtiene por:

```
IF              (H EQ 1)  Z = Y + X
```

I.2.7 Muestreo, selección y ponderación de datos

La tarjeta de control SAMPLE permite al usuario seleccionar una muestra al azar de sus datos (el 10% por ejemplo) antes de realizar un cruce, una regresión o cualquier procedimiento.

SELECT IF: permite al usuario seleccionar los casos que desea (por ejemplo solo hombres).

WEIGHT: permite la ponderación de los datos de acuerdo a algún factor existente o generado por medio de COMPUTE e IF.

I.2.8 Recuperación de datos desde el sistema

El SPSS permite que los datos entrados al sistema sean perforados en tarjetas, grabados en cinta o disco para que así se genere un archivo que será el INPUT a otros programas o paquete estadístico. De esta forma se puede utilizar todo el potencial de transformaciones, y recodificación y selección de casos que permite el SPSS para generar un archivo de entrada a programas posteriores. El usuario tiene el control total de las variables de salida, formato y selección de casos.

II. LA ORGANIZACION Y CODIFICACION DE DATOS PARA LA ENTRADA AL SPSS

II.1 EL CASO COMO UNIDAD DE ANALISIS

Un caso es la unidad básica de estudio para la cual una serie de medidas han sido observadas.

Cada unidad o caso se compone de los valores para una o más medidas que se han tomado. Estas medidas se denominan variables y cada caso tendrá un valor para cada variable.

Un caso o unidad de estudio puede ser "un individuo", "una vivienda", "una industria", "una ciudad", etc.

II.2 CONVENCIONES DE CODIFICACION DE DATOS

Normalmente las medidas recogidas en un estudio son posteriormente codificadas y perforadas en tarjeta o traspasada a cualquier otro medio legible por la máquina.

Una unidad de estudio (caso o cuestionario) puede dar origen a una o más tarjetas perforadas. En el caso de que una unidad de estudio ocupe dos o más tarjetas, todas las tarjetas llevarán la misma identificación y una variable denominada clase de tarjeta que indica si la tarjeta es la 1ra, 2da, 3ra, ... de una unidad de estudio.

El SPSS sólo trabaja con unidades de estudio que tienen un número fijo de tarjetas por unidad de estudio. Por ejemplo: encuestas del tipo de fecundidad que traen una tarjeta con las características de la madre y una tarjeta adicional por cada hijo tenido requerirá algún manipuleo previo antes de poder hacerse uso del SPSS.

II.2.1 Variables y tipos de variables

La codificación de las variables para la entrada al SPSS puede originar dos tipos:

- Variables numéricas (valores con 1 o más dígitos).
- Variables alfanuméricas (dígitos, letras u otros caracteres).

Las variables numéricas permiten el manipuleo matemático y ciertos cálculos estadísticos como la media y desviación estándar, sólo pueden ser computados para las numéricas.

Distribuciones de frecuencias y cruces pueden ser obtenidas tanto para variables numéricas y alfanuméricas. De todas formas el SPSS provee la instrucción RECODE mediante la cual podemos transformar una variable alfanumérica a numérica.

II.2.2 La codificación de valores faltantes

El SPSS permite que declaremos (opcional) desde 1 hasta 3 valores faltantes para cada variable. Los valores faltantes de las variables a ser investigados pueden originarse en distintas formas: el entrevistado contesta que "NO SABE" o se rehusa a contestar "NO RESPONDE"; por falla del entrevistado que no registró la información "IGNORADO". Hay también los casos que la variable puede asumir un valor que "NO SE APLICA" como sería el caso de número de hijos para personas de sexo masculino.

II.2.3 La organización de los datos

El SPSS permite al usuario dos alternativas para organizar las variables en la tarjeta o registro:

- A. Formato de campo fijo.
- B. Formato de campo libre.

A. Organización de datos en formato de campo fijo

Esta organización de datos es aquella en que los valores de cada variable siempre aparecerán localizados en las mismas columnas (o campo) para todos los casos (o unidades de estudio).

Ejemplo I: una tarjeta por caso

	número cuestion.	edad	sexo	es casado?
1er caso	001	10	1	no
2° caso	002	25	1	SI
3er caso	003	39	2	SI
4to caso	004	65	1	NO
5to caso	005	07	2	NO
...				

Ejemplo II: dos tarjetas por caso

	tarij.1	numero	ct.	edad	sexo	es casado	
	tarj.2	cuestion		nivel			ingreso
1er caso		001	1	10	1	NO	
		001	2		0		00010.0
2do caso		002	1	35	1	SI	
		002	2		1		01000.0
3er caso		003	1	40	2	NO	
		003	2		2		02550.0
4to caso		004	1	15	1	NO	
		004	2		1		0.0
...							
117° caso		117	1	30	2	SI	
		117	2		3		00500.0

En el ejemplo II un caso trae dos tarjetas: o sea un caso continúa físicamente de la primera tarjeta en la segunda tarjeta. No hay un límite del número de tarjetas que pueda tener un caso pero el SPSS sólo permite trabajar con casos que siempre tengan el mismo número de tarjetas. El orden lógico de las variables está condicionado por el orden físico en que los valores están perforados y en el ejemplo II es: EDAD, SEXO, CONDICION DE ESTADO CIVIL, NIVEL DE INSTRUCCION e INGRESO.

B. Organización de datos en formato de campo libre

En esta organización no hay una correspondencia entre las variables y las columnas como ocurre en el caso de campo fijo. Al contrario del formato en campo fijo en la organización de formato de campo libre uno o más casos pueden ocupar una misma tarjeta.

El valor de cada variable en un caso debe estar separado por uno o más blancos y/o comas. Esta separación es la que determina donde terminó el valor de una variable y donde comienza el valor de la próxima.

Las variables con codificación alfanumérica deben traer sus valores encerrados entre comillas simples como por ejemplo si una variable asume el valor A debe ser perforado 'A', si asume el valor alfanumérico 1 debe ser perforado '1'.

001	10	1	'NO'	0	10.0	002	35	1
'SI'	1	1000.0	003				

ejemplo de formato de campo libre

C. Asignación de identificación de caso

Independiente de la identificación que el usuario utilice para los casos el SPSS crea automáticamente una variable de identificación de secuencia con el nombre SEQNUM que asume los valores desde 1 hasta n, donde n sería el número total de casos leídos. Esta variable puede ser utilizada por el usuario y la referencia a ella se hace por su nombre: SEQNUM.

II.3 SUBARCHIVOS: SU FUNCION Y ESTRUCTURA

Frecuentemente en ciencias sociales el investigador desea analizar sus datos agrupándolos en dos o más juegos pues esto le permite una mejor manera de efectuar su estudio. Para una encuesta a nivel nacional por ejemplo cada grupo podría ser el subconjunto de datos de cada provincia, o también la agrupación por estratos sociales, etc.; o sea hay una infinidad de situaciones en las cuales se desea hacer agrupaciones en base a una o más variables críticas.

En SPSS estos grupos son denominados subarchivos.

II.3.1 El uso de SUBARCHIVOS

La estructura de subarchivos permite varios tipos de análisis: puede ser utilizada para examinar o probar tanto la distribución o tendencia central como en comparaciones de relaciones de variables en distintos subarchivos.

En cualquiera de los tipos de análisis, el SPSS permite procesar cada subarchivo en forma separada o procesar todos los subarchivos como si fuera sólo uno.

Permite además, obtener los resultados para ciertos conjuntos de subarchivos: en nuestro ejemplo de subarchivos por provincias podríamos obtener resultados para ciertos conjuntos de provincias que conformarían una o más regiones. Las distintas maneras de procesar los subarchivos están especificadas por la tarjeta de control ~~PR~~CESS SUBFILES.

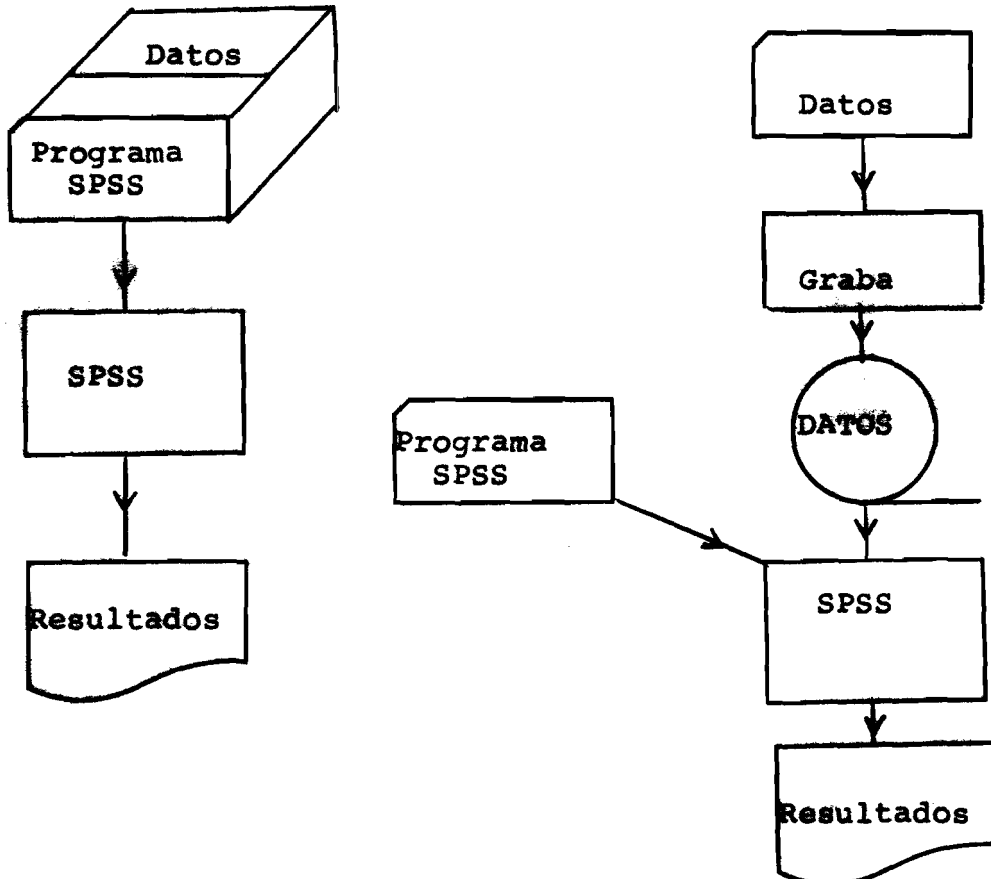
II.3.2 La estructura de subarchivo

Todos los subarchivos de un archivo deben contener las mismas variables y las variables deben estar en el mismo formato. Los casos que contiene cada subarchivo pueden variar. Los casos de cada subarchivo deben estar agrupados todos juntos de tal forma que el primer juego de datos conforma el primer subarchivo, a continuación vendría el segundo juego de datos que conformaría el segundo subarchivo y así sucesivamente como se ilustra en la figura.

	Número cuestionario	Provincia	Sexo	Edad	
Caso 1	0001	1	1	10	Primer subarchivo (203 casos)
Caso 2	0005	1	2	11	
Caso 3	0007	1	1	13	
⋮	⋮	⋮			
Caso 201	0978	1	1	20	
Caso 202	0997	1	1	40	
Caso 203	1003	1	2	50	
Caso 1	0002	2	1	60	Segundo subarchivo (511 casos)
Caso 2	0003	2	2	71	
Caso 3	0004	2	1	20	
Caso 4	0011	2	2	07	
⋮	⋮	⋮			
Caso 507	0101	2	1	06	
Caso 510	0950	2	1	50	
Caso 511	2521	2	2	73	
Caso 1	0013	3	1	19	Tercero subarchivo (1004 casos)
Caso 2	0015	3	1	20	
Caso 3	0971	3	2	31	
		3	1	28	
		⋮			
Caso 1001	2010				
Caso 1002	2071	3	2	20	
Caso 1003	3074	3	2	35	
Caso 1004	4011	3	1	45	

II.4 INGRESO DE DATOS

Una vez que los datos hayan sido codificados y perforados pueden entrar al SPSS desde tarjetas perforadas o pueden previamente ser grabados en disco o cinta o cualquier otro medio legible por máquina.



II.5 LIMITACIONES SOBRE LA ENTRADA DE DATOS

No existe un límite del número de casos que puedan conformar un archivo de datos para la entrada al SPSS. Más bien tal limitación está dada por los medios de almacenamiento que se utilice, que varía de equipo a equipo. Si se utiliza cintas magnéticas prácticamente el número de casos es ilimitado.

Independientemente del número de casos en un archivo, el SPSS no permite manipular más de 500 variables por caso ni más que 100 subarchivos.

III. TARJETAS DE CONTROL SPSS

El SPSS es llevado a través de sus variadas funciones por medio de tarjetas de control, que permiten al usuario controlar el procesamiento de sus datos.

A pesar de que cada tarjeta de control desarrolla una función única, todas comparten un formato común y la información en ellas es codificada en forma similar.

El lenguaje de las tarjetas de control SPSS es bastante sencillo, constituyéndose de: 1) nombres 2) valores 3) palabras claves 4) reglas de puntuación y espaciamento.

III.1 EL FORMATO GENERAL

Campo de control y campo de especificación.



CODEBOOK
REGRESSION
INPUT FORMAT
CROSSTABS

III.2 EL CAMPO DE CONTROL

La palabra de control debe ser perforada en columnas 1-15 siempre comenzando en columna 1. La palabra o juego de palabras de control identifican la tarjeta de control e informan al sistema qué tipo de especificaciones seguirán en el campo de especificación.

III.3 EL CAMPO DE ESPECIFICACION

El campo de especificación está constituido por la combinación de seis elementos:

1. Nombres
2. Valores
3. Palabras claves
4. Títulos
5. Delimitadores
6. Operadores

los cuales pueden ser perforados libremente en columnas 16-80 y continuar en otra tarjeta dejando el campo de control (col. 1-15) en blanco -se puede usar tantas tarjetas de continuación como se desee.

III.3.1 Nombres

Los nombres son utilizados por el usuario para definir y/o referirse a: (1) sus variables, (2) sus subarchivos (si los hay) y (3) el archivo que será creado.

Reglas:

- 1) Sólo se puede usar caracteres alfabéticos o numéricos (no se permiten los caracteres especiales).
- 2) Primer carácter de los ocho, máximo permitidos, debe ser alfabético.
- 3) Entre los caracteres de un mismo nombre no debe haber blancos.

Ejemplo: SEXO, EDAD, N, VAR1, VAR10, ØC111

III.3.2 Valores

Los valores que pueden ser numéricos y/o alfanuméricos, entran como elementos en el campo de especificación.

Valores numéricos: pueden ser enteros o decimales y su uso en general es en transformación de variables, por ejemplo:

```
COMPUTE      VARA = VARB ** 2 (el 2 = valor numérico)
IF           (VARX EQ 0.01) VARZ = VARY * 10.1
            (el 0.01 y 10.1 son valores numéricos)
```

Valores alfanuméricos: deben ir entre comillas simples y su uso en general es para recodificación o comparaciones de una variable leída en alfanuméricos.

```
IF           (ALFA EQ 'AB') BETA = ALFA
RECODE      ALFA ('AA' = 1) ('AB' = 2) ...
            ('AZ' = 30) ('--' = 40)
            ('- ' = 50)
```

En el manejo de variables y valores alfa - numéricos hay que tomar la siguiente precaución:

el valor alfanumérico '1' ≠ '01' ≠ 'Ø1'

o sea la representación interna de una variable leída en alfanuméricos y que tiene perforado los códigos válidos 1, 2, 3, 4, 5, 6, 7, 8, 9 -tendrá la representación interna: '1', '2', '3', '4', '5' ... '9', '-' que es distinta de la representación interna numérica de 1, 2, 3, 4, ..., 9

III.3.3 Palabras claves

Las palabras claves son configuraciones pre-establecidas de carácter que poseen un significado especial en el SPSS.

Entre ellas tenemos

TØ; WITH; BY; AND; ØR; EQ; NE; etc.

Como las palabras claves poseen un significado especial dentro de un cierto contexto, son descritas en más detalle en el uso de las tarjetas de control a las cuales se asocian.

III.3.4 Títulos (LABELS)

El sistema SPSS posee una gran capacidad y flexibilidad para titular variables o categorías de variables y archivos. Los títulos extendidos permitidos pueden ser cualquier combinación de caracteres (alfabéticos, numéricos y especiales) excluyendo algunos caracteres especiales que son utilizados como de limitadores especiales.

III.3.5 Delimitadores

Los nombres de variables, valores y palabras claves deben estar separados por delimitador:

Hay dos tipos de delimitadores.

- Delimitadores comunes: el blanco y la coma, son para separar los elementos en el campo de especificación y se pueden usar indistintamente y cuantos se deseen para mejorar la lectura.

- Delimitadores especiales: el slash, paréntesis izquierdo y paréntesis derecho "/"; "(";")". Su uso depende de la tarjeta de control pero en general el "/" es utilizado para separar listas de nombres de variables y los paréntesis para encerrar listas de valores.

III.3.6 Operadores

Los operadores en general son usados en las tarjetas de transformación de variables (COMPUTE e IF) y son los operadores matemáticos conocidos.

exponenciación	**
multiplicación	*
división	/
suma	+
resta	-
asignación	=

Observe que la división "/" es el mismo símbolo del delimitador especial ya descrito; pero no causará confusión pues son utilizados en contextos distintos. O sea en las tarjetas de control COMPUTE e IF donde se utilizarán estos operadores, nunca se utilizará el delimitador especial "/".

III.4 NORMAS GENERALES PARA LA PREPARACION DE TARJETAS DE CONTROL

1. Las palabras de control deben ser escritas y espaciadas correctamente.
2. Los nombres deben tener un máximo de ocho caracteres; el primero debe ser alfabético, las demás letras o números, no se permiten caracteres especiales.
3. Los valores pueden ser enteros, decimales o alfanuméricos (letras, números, caracteres especiales). Los valores alfanuméricos deben aparecer entre comillas simples.
4. Las palabras claves deben ser escritas y espaciadas exactamente como se presentan en el texto.
5. Los "labels" pueden constituirse de cualquier caracter válido, excluyéndose los delimitadores especiales.
6. Los nombres, palabras claves y valores no deben llevar delimitadores entremedio; ni pueden comenzar en una tarjeta y seguir en la siguiente. Los labels sí pueden iniciarse en una tarjeta y tener la continuación en la siguiente.
7. Los nombres, palabras claves y valores deben estar separados unos de otros por delimitadores.

IV. LAS TARJETAS DE DEFINICION DE DATOS

111

111

8. Si el campo de especificación requiere más de una tarjeta de control puede ser continuado en tarjetas sucesivas de columnas 15-80, dejando siempre el campo de control en blanco.

IV. LAS TARJETAS DE DEFINICION DE DATOS

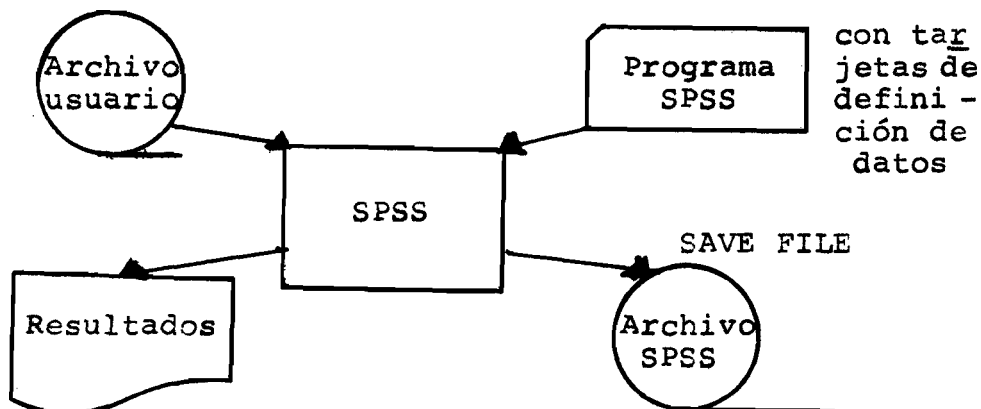
Los datos son manipulados por el SPSS en grupos denominados "archivos". Un "archivo" queda definido con la siguiente información que definen y describen los datos:

1. Nombre del archivo.
2. Una lista de variables (serán leídas).
3. Estructura de subarchivos (si hay).
4. Número de casos.
5. Medio de entrada (donde residen los datos).
6. El formato de entrada.

En forma opcional:

7. Valores faltantes (Missing values).
 8. Títulos extendidos para variables (Var Labels).
 9. Títulos extendidos de categorías (Value Labels).
 10. Formato de impresión (Print Formats).
- (Este último es obligatorio para variables alfanuméricas).

Esta información es entrada al sistema SPSS por medio de las "tarjetas de definición de datos" que conjuntamente con los casos leídos conforman un archivo completo que si deseamos podemos almacenarlo como un archivo del sistema por medio de la tarjeta "SAVE FILE". Las tarjetas de definición de datos sólo son requeridas si estamos procesando el archivo del usuario.



IV.1 La tarjeta "FILE NAME"

Un "nombre de archivo" identifica cada archivo SPSS. O sea el nombre con lo cual bautizamos un archivo SPSS nos permite posteriormente utilizar tal archivo llamándolo por su nombre. La tarjeta "FILE NAME" es obligatoria si vamos a guardar un archivo del sistema (archivo SPSS) mediante el uso de la tarjeta SAVE FILE. Caso contrario esta tarjeta es opcional.

La estructura general de la tarjeta es:

1	16
FILE NAME	nombre del archivo, [Título extendido]

O sea en columnas 1-15 (campo de control) debe ir la palabra de control FILE NAME comenzando en la columna 1.

Ejemplos:

1	16
FILE NAME	ARCH1, esto es un título extendido
FILE NAME	NAPØLEØN, archivo bautizado con el nombre de Napoleón
FILE NAME	JØSEFINA

Los nombres de archivos deben seguir las reglas para nombres (máximo 8 caracteres, el primero debe ser alfabético).

En el primer ejemplo el nombre del archivo es "ARCH1", todo lo demás son comentarios que ayudan a describir el archivo. El segundo ejemplo bautizamos el archivo con el nombre de "NAPØLEØN". Y el tercero, con el nombre de "JØSEFINA". El largo máximo del título extendido es de 64 caracteres.

IV.2 La tarjeta "VARIABLE LIST"

Esta tarjeta desarrolla la función de definir los nombres de las variables que serán leídas desde el archivo del usuario. El

nombre que asignamos a cada variable a ser leída es arbitrario, siguiendo las reglas para la construcción de nombres ya establecidas. Existen tres nombres de variables pre-definidas "SEQNUM, CASWGT y SUBFIL que no pueden ser utilizados como nombres de variables. Esto también es válido para las siguientes 20 palabras claves: "TO, WITH, BY, SQRT, EXP, LN, SIN, COS, ATAN, AND, LG10, GE, LE, GT, LT, EQ, NE, ØR y NØT.

La estructura general de la tarjeta es:

1	16
VARIABLE LIST	lista de nombres de variables

La lista de nombres puede estar conformada por una serie de nombres mnemotécnicos que recuerdan la naturaleza de cada variable y el orden debe ser el mismo en que cada variable se encuentra perforada en la tarjeta(s) de cada caso. Algunos ejemplos de esta tarjeta:

- | | |
|---|----|
| 1 | 16 |
|---|----|
- (a) VARIABLE LIST EDAD, SEXØ, NIVEL, ESTCIV, ØCUP
- (b) VARIABLE LIST RELJEF, ØCUPJEF, ØCUPENT, EDJES, EDENT

La lista de nombres de variables puede definir un máximo de 500 variables para un archivo dado.

El hecho de tener que definir 100, 200 a 500 nombres de variables para el proceso de nuestros datos utilizando nombres mnemotécnicos tiene tres desventajas:

- es una tarea tediosa,
- requiere confección de muchas tarjetas; y
- puede acarrear errores ya que podemos perfectamente repetir algún nombre ya definido.

Para estos casos existe una convención especial para la lista de nombres, más segura y más fácil de usar:

VARxxx TØ VARyyy

Donde xxx, yyy son números de tres dígitos debiendo xxx ser menor que yyy. El prefijo de los 3 dígitos debe ser "VAR" y la palabra clave "TO" debe ser entre los dos números de variables separados por uno o más delimitadores comunes.

Ejemplo:

```
1          16
VARIABLE LIST VAR001 TØ VAR081
```

En este ejemplo estamos definiendo 81 nombres de variables a ser leídas, que son: VAR001, VAR002, VAR003, ... VAR080, VAR081.

El uso de esta convención requiere siempre:

- a) uso del prefijo VAR seguido de 3 dígitos.
- b) la palabra clave TO
- c) el dígito a la izquierda del TO debe ser siempre menor que el dígito de la derecha.

Las dos convenciones pueden ser mezcladas en una lista de variables como se muestra:

```
1          16
VARIABLE LIST NOMBRE, EDAD, VAR001 TØ VAR003, SEXØ, ØCUP, VAR004
              TØ VAR008
```

Para este ejemplo tendríamos la siguiente tabla considerando una tarjeta por caso:

Posición de la variable en el caso	Columna en que la variable está perforada	Nombre de la variable
001 (primera)	1 - 8	NOMBRE
002 (segunda)	9 - 10	EDAD
003 :	11 - 12	VAR001
004 :	21 - 24	VAR002
005	33 - 35	VAR003
006	41	SEXØ
007	42 - 44	ØCUP
008	51	VAR004
009	52	VAR005
010	53	VAR006
011	61 - 62	VAR007
012	60	VAR008

También podemos mezclar la convención VARxxx de la siguiente forma:

```
1                   16
VARIABLE LIST  VAR001 TØ VAR007, VAR011 TØ VAR020
```

SUMARIO DE REGLAS Y CONVENCIONES PARA LA TARJETA "VARIABLE LIST"

1. La palabra de control es "VARIABLE LIST"
2. La lista de nombres puede estar compuesta de nombre mnemotécnicos y/o nombres definidos por la convención VARxxx TØ VARyyy.
3. Todas las normas que rigen la confección de nombres deben ser usadas.
4. No se puede declarar dos veces ningún nombre de variable o sea el nombre para cada variable debe ser único.

5. El orden de la lista de nombres debe corresponder al orden en que las variables aparecen en los casos.
6. No se pueden definir más de 500 variables.

INTERFASE OSIRIS - SPSS

El SPSS permite la lectura de archivos generados por el paquete OSIRIS (Organized Set of Integrated Routines for Investigation with Statistics) por medio de la tarjeta OSIRIS VARS en vez de VARIABLE LIST. El OSIRIS es otro paquete de programas desarrollado por el "Institute for Social Research" de la Universidad de Michigan y el "Inter-University Consortium for Political Research" que posee las siguientes capacidades en el manejo de datos:

- generación de datos
- corrección de archivos
- transformaciones lógicas y aritméticas de datos tanto entre registros como a través de registros.
- el análisis estadístico incluye una variedad de programas de análisis multivariado y no-paramétrico.

En estos apuntes no entraremos en los detalles de la interfase ya que eso requiere previo conocimiento del paquete OSIRIS.

IV.3 La tarjeta SUBFILE LIST

El uso de esta tarjeta es opcional y sólo se requiere si estamos creando una estructura de subarchivos. Esta tarjeta informa al sistema los nombres de cada subarchivo que entran a componer nuestro archivo global. La estructura es similar a la VARIABLE LIST y los nombres de los subarchivos deben ser especificados en el mismo orden en que cada subarchivo será leído.

1	16
SUBFILE LIST	lista de nombres mnemotécnicos

Ejemplo:

SUBFILE LIST SANJOSE, PUNTAREN, GUANAC, LIMON

Los nombres de los subarchivos se constituyen en los valores de una variable especial llamada SUBFIL que puede ser utilizada para el análisis como cualquier otra.

El número máximo de subarchivos a ser creados es de 100.

IV.4 La tarjeta # OF CASES

Esta tarjeta informa al sistema el número de casos en el archivo del usuario. Si tenemos una estructura de subarchivos esta tarjeta informa el número de casos de cada subarchivo.

IV.4.1 Si no hay estructura de subarchivo el formato de esta tarjeta es:

1	16
# OF CASES	{ número exacto de casos a leerse ESTIMATED número casos }

Ejemplo:

1	16
# OF CASES	2351

En este ejemplo serán leídos 2351 casos.

Si nuestro archivo posee más casos solamente serán leídos los 2351 primeros y sólo éstos serán procesados.

Para el caso más general que inicialmente desconocemos el número exacto de casos que contiene nuestro archivo existe la convención ESTIMATED cuyo uso es el siguiente:

1	16
# OF CASES	ESTIMATED 5000

El número colocado después de la palabra clave **ESTIMATED** debe ser siempre mayor que el número real de casos. En el ejemplo anterior si nuestro archivo tiene un número de casos menor o igual a 5000 serán leídos todos los casos, pero si tiene más de 5000 sólo serán leídos 5000. O sea el uso de la convención **ESTIMATED** implica que sepamos en forma aproximada el número real de casos existentes. Es recomendable siempre que se use la convención **ESTIMATED**, para especificar un número bien superior al que creemos sea el número real de casos en vez de especificar un número cerca de los límites del número real. Si estimamos por ejemplo que nuestro archivo tiene aproximadamente 10000 casos la tarjeta recomendable sería:

OF CASES ESTIMATED 50000

IV.4.2 Si estamos creando una estructura de subarchivos se debe especificar claramente el número exacto de casos para cada subarchivo. En este caso, esta tarjeta está directamente asociada a la tarjeta **SUBFILE LIST**.

Ejemplo:

1	16				
SUBFILE LIST	SANJOSE,	PUNTAREN,	GUANAC,	LIMÓN	
# OF CASES	1501,	958,	2030,	857	

O sea el primer subarchivo cuyo nombre es **SANJOSE** posee 1501 casos, el 2° **PUNTAREN** posee 958 casos y así sucesivamente.

IV.5 La tarjeta INPUT MEDIUM

Esta tarjeta informa al sistema el medio donde reside el archivo del usuario.

La estructura general es:

1	16				
INPUT MEDIUM	[TAPE]		
		CARD			
		DISK			
		OTHER			

O sea lleva la palabra de control INPUT MEDIUM y en el campo de especificación una y sólo una de las palabras claves que se muestra.

Ejemplo: si nuestros datos están almacenados en cinta esta tarjeta sería:

1	16
INPUT MEDIUM	TAPE

Si el archivo del usuario está en cinta, disco u otro medio se debe especificar en las tarjetas de control del sistema operativo la unidad lógica FT08. Si el archivo será leído desde tarjetas no requiere la especificación de FT08 pero tampoco permite el uso de la convención ESTIMATED en la tarjeta de # OF CASES.

IV.6 La tarjeta INPUT FORMAT

Esta tarjeta especifica la siguiente información que describen los datos de un archivo del usuario:

- a) Organización de los casos:
 - Formato de campo fijo
 - Formato de campo libre
- b) El tipo de cada variable:
 - numérica
 - alfanumérica
- c) La ubicación de cada variable: tarjeta en que se encuentra y columna(s).

IV.6.1 Formato de campo fijo o libre

Como ya se describió en capítulos anteriores el SPSS permite leer datos en dos tipos de formatos:

El primer elemento de formato especifica la información al respecto de la primera variable de la lista de VARIABLE LIST y así sucesivamente de izquierda a derecha.

IV.6.3 Los elementos de formato

IV.6.3.1 VARIABLES TIPO A (alfanuméricas)

VARIABLES alfanuméricas son aquellas que puede asumir valores codificados con caracteres alfabéticos, numéricos y caracteres especiales.

Un elemento de formato en la forma Aw indica que la variable es alfanumérica y que ocupa w columnas.

La forma general es:

<u>nAw</u>	donde	n = factor de repetición
		A = elemento alfanumérico
		w = largo de la variable(s) (w máximo = 8)

El factor de repetición se utiliza para el caso de que hay n variables adyacentes todas con el mismo largo de w

Ejemplos:

- a) 20A3 20 variables alfanuméricas adyacentes con tres columnas cada una.
- b) 30A1 30 variables alfanuméricas adyacentes con una columna por variable.
- c) 1A2 \emptyset A2 una variable alfanumérica que ocupa dos columnas.

NOTA: no debe dejarse blancos entre n, A, w.

IV.6.3.2 VARIABLES TIPO F (numéricas)

VARIABLES numéricas son aquellas cuyos valores son siempre caracteres numéricos (0, 1, 2, 3, 4, 5, 6, 7, 8, 9,)

La forma general es

$nFw.d$

donde

n = factor de repetición

F = elemento de formato numérico

w = largo de la(s) variable(s) (número de columnas) incluyendo el signo y el punto decimal si son perforados

d = número de decimales ($d < w$)

Para el caso más general de datos de encuestas en que la mayor parte de los valores registrados no existe casas decimales tendríamos $d = 0$ o sea por ejemplo para leer edad y sexo que ocupan 2 y 1 columna respectivamente.

edad: F2.0

sexo: F1.0

Si se encuentra perforado el punto decimal este tiene preferencia sobre el número de decimales especificados en el elemento de formato.

Ejemplos:

Elemento del formato	Valor perforado	Valor leído y almacenado en memoria
F4.0	1000	1000.
F4.1	1000	100.0
F4.2	1000	10.00
F4.0	1.00	1.0
F4.0	1.58	1.58
F4.3	1.58	1.58
F4.2	1.58	1.58
F4.2	1058	10.58
F4.3	0100	0.100

IV.6.3.3 Salto de columnas con X

Forma general: nX

El elemento X permite saltar el número de columnas especificado por n.

1X - salta una columna

12X - salta doce columnas

IV.6.3.4 Transferencia a una columna específica con Tr

El elemento Tr indica la columna a partir de la cual (inclusive) debe leerse la(s) variable(s) cuyo elemento de formato se especifica inmediatamente a continuación:

Supongamos que tenemos

VARIABLE LIST EDAD, SEXO, OCUP

y que

EDAD - perforada en columna 11-12

SEXO - perforada en columna 20

OCUP - perforada en columna 71-73

tendríamos

1 16
INPUT FORMAT FIXED (T11, F2.0, T20, F1.0, T71, F3.0)

o su equivalente, utilizando X

INPUT FORMAT FIXED (10X, F2.0, 7X, F1.0, 50X, F3.0)

Note que para la primera variable (EDAD) utilizando T tenemos T11 al paso que utilizando X tenemos 10X o sea salta las 10 primeras columnas y lee a partir de la columna 11.

IV.6.3.5 Salto de registro (tarjetas) con el "/".

El elemento "/" da la instrucción para que el sistema se posicione inmediatamente al próximo registro para continuar la lectura o saltar tarjetas cuyas variables no interesan al estudio para el cual se realiza el proceso.

El "/" puede ser utilizado de tres formas distintas como se ilustra: supongamos que cada caso de nuestro archivo posee 4 tarjetas.

- a) Saltar tarjetas iniciales para posicionarnos en aquella que nos interesa leer datos:

```
INPUT FORMAT  FIXED (///, 10X, F2.0 ...)
```

Salta tres tarjetas e inicia la lectura de la primera variable en las columnas 11-12 de la 4° tarjeta.

- b) Saltar tarjetas finales para que la lectura del próximo caso sea correcto

```
INPUT FORMAT  FIXED (20X, F3.0, ///)
```

Lee una variable en la primera tarjeta y salta las 3 que restan.

- c) Saltar tarjetas intermedias

```
INPUT FORMAT  FIXED (F2.0, F1.0, ///, F5.2)
```

Lee dos variables en la primera tarjeta, salta la segunda y la tercera y lee una variable en la cuarta tarjeta.

Hay una sencilla regla para el uso del "/" que es:

1. Para saltar tarjetas iniciales o finales
 n "/" salta n tarjetas
2. Para saltar tarjetas intermedias
 n "/" salta $(n - 1)$ tarjetas

Ejemplos de uso de la lista de variables y lista de formatos

Ejemplo 1: 1 tarjeta por caso

```
1          16
VARIABLE LIST  VAR001 TØ VAR006
INPUT FORMAT  FIXED(T11, F2.0, 3A3, 2X, 2F1.0)
```

<u>Nombre de la variable</u>	<u>Tipo de variable</u>	<u>Tarjeta</u>	<u>Columna</u>
VAR001	F	1	11-12
VAR002	A	1	13-15
VAR003	A	1	16-18
VAR004	A	1	19-21
VAR005	F	1	24
VAR006	F	1	25

El formato informa al sistema:

1. Posiciones en la columna 11
2. Lea variable numérica en columnas 11-12
3. Lea tres variables alfanuméricas adyacentes de tres columnas cada una.
4. Salté dos columnas (22-23)
5. Lea dos variables numéricas adyacentes de una columna cada una.
6. Repita de 1 a 5 para cada caso.

Ejemplo 2: 5 tarjetas por caso

```

1          16
VARIABLE LIST  EDAD, SEXO, NIVEL, VAR001 TØ VAR007
INPUT FORMAT  FIXED(11X, F2.0, T51, F1.0, /, 40X, F2.0, T60, 2A3,
              //, 21X, 3F1.0, T47, F2.0, F3.0, /)
    
```

<u>Nombre variable</u>	<u>Tipo de variable</u>	<u>Tarjeta</u>	<u>Columna</u>
Edad	F	1	12-13
SEXO	F	---1---	51
NIVEL	F	2	41-42
VAR001	A	2	60-62
VAR002	A	---2---	63-65
VAR003	F	4	22
VAR004	F	4	23
VAR005	F	4	24
VAR006	F	4	47-48
VAR007	F	4	49-51

La lista de formato informa:

1. Salte 11 columnas
2. Lea variable numérica de dos columnas
3. Posiciónese en columna 51
4. Lea variable numérica (en columna 51) de largo 1
5. Salte a la 2da. tarjeta
6. Salte 40 columnas
7. Lea variable numérica de dos columnas
8. Posiciónese en la columna 60
9. Lea (partir de columna 60 inclusive) dos variables alfa-numéricas de tres columnas
10. Salte a la 4ta. tarjeta
11. Salte 21 columnas
12. Lea tres variables numéricas de una columna cada una
13. Posiciónese en columna 47.
14. Lea variable numérica de dos columnas
15. Lea variable numérica de tres columnas
16. Salte a la tarjeta 5
17. REPITA pasar de 1 a 16 para cada caso

IV.7 La tarjeta MISSING VALUES

En muchas investigaciones el usuario desea tratar los casos que se registraron "NO SABE", "NO SE APLICA", "IGNORADO" en forma distinta a aquellos casos donde el registro de la información es una categoría válida. El SPSS permite al usuario declarar para cada variable hasta un máximo de tres valores faltantes. Al momento de realizar algún procedimiento en específico aquellos casos que traen los valores faltantes declarados son manipulados en forma específica dependiendo de la opción que se utiliza. Cada procedimiento posee una manera específica de cómo tratar los valores faltantes declarados.

El formato general es:

```

1           16
MISSING VALUES { nombre de variable } (valores faltantes)/
                 { o
                 { lista de nombres }
... / { nombre de var } (valores faltantes)
      { o
      { lista de vars }

```

Ejemplo

```

MISSING VALUES EDAD(97, 98, 99)/VAR001, VAR003 TØ
VAR007 (9)/NIVEL, ESTCIV (0,9)/
ALFA, BETA ('A', '0')/INGRESO(9800.90)

```

Limitación: el campo de especificación para una instrucción de MISSING VALUES no puede contener más de 250 elementos (nombres, valores y delimitadores especiales). La siguiente parte del ejemplo anterior:

```
VAR001, VAR003 TØ VAR007 (9)
```

posee ocho elementos:

```
3 nombres TØ, VAR001, VAR003, VAR007
```

```
1 valor: 9
```

```
3 delimitadores especiales: "(", ")", "/"
```


Esta limitación es para una instrucción MISSING que puede tener varias tarjetas de continuación con el campo de control en blanco. Si nuestros datos requieren definir valores faltantes y para tal objeto necesitamos 480 elementos ponemos dos instrucciones de MISSING, por ejemplo:

MISSING VALUES lista con 240 elementos (varias tarjetas de continuación)

MISSING VALUES lista con 240 elementos (varias tarjetas de continuación)

Para la declaración de valores faltantes de variables creadas por medio de tarjetas IF, COMPUTE y COUNT hay la tarjeta de control ASSIGN MISSING que se explicita junto con las tarjetas de transformación de datos.

IV.8 La tarjeta VAR LABELS

Esta tarjeta permite al usuario especificar un título extendido para cualquiera o todas las variables de su archivo. El título saldrá impreso en todas las salidas que utilicen la o las variables.

Formato general:

1	16
<u>VAR LABELS</u>	nombre de var ₁ , título / nombre var ₂ , título / .../ nombre var _n , título /

Normas para la confección de la tarjeta VAR LABELS:

- Solo se puede especificar un nombre de variable y un título a la vez o sea "nombre var₁, nombre var₂, título" no es válido.
- La variable a la cual se le asigna un título debe estar previamente definida.

- c) Se permite un título extendido de un largo máximo de 40 caracteres válidos del juego IBM 360 con excepción de los caracteres "(", ")", "/", que son los delimitadores especiales.
- d) para la continuación se requiere el delimitador "/"
- e) La cuenta de caracteres empieza con el primer caracter

```
1          16
VAR LABELS  EDAD, EDAD SIMPLES /
            VAR001, SEXO / VAR002, NIVEL DE INSTRUCCION /
            VAR003, ESTADO CIVIL /
            VAR010, INGRESO FAMILIAR
```

IV.9 La tarjeta VALUE LABELS

Permite al usuario especificar un título extendido para algunos o todos los valores (categorías) y para algunas o todas las variables del archivo.

Formato general:

```
1          16
VALUE LABELS { nombre de var } (valor1) título1, (valor2) título2...
              { o }
              { lista de vars } (valorn) títulon /

              { nombre var } (valor1) título1, (valor2) título2
              { o }
              { lista de vars } .... (valorn) títulon /
```

Normas de uso

- a) No es necesario especificar títulos para todas las categorías de una variable, si deseamos, sólo especificamos para algunos códigos.

- b) El máximo de caracteres para un título es de 20 caracteres. El "(", ")", "/" no se pueden utilizar por el hecho de que desarrollan la función de delimitadores especiales. La salida en cruces sólo utilizará 16 caracteres y en 2 dos filas de 8 caracteres cada una.
- c) La continuación de una tarjeta a otra se hace dejando el campo de control en blanco y pasar de una lista de variables a otra requiere el "/".
- d) La cuenta de caracteres después de (valor) empieza al detectar el primer caracter distinto de blanco o coma.

Ejemplo:

```
1          16
VALUE LABELS  VAR001, VAR007 TØ VAR015 (1) SI (2) NØ
              (9) IGNØRADØ / VAR002 TØ VAR006 (1)
              PRIMARIA (2) SECUNDARIA (3) UNIVERSITARIA
              (9) IGNØRADØ
```

IV.10 La tarjeta PRINT FØRMATS

Esta tarjeta informa al sistema con cuantos decimales se desea la salida impresa en el caso de variables numéricas y si la impresión es de una variable alfanumérica. Para el caso de variables alfanuméricas esta tarjeta es obligatoria; para las numéricas la impresión estándar es con dos decimales: o sea si una variable leída con formato F1.0 (puede asumir valores 0, 1, 2 ... 9) y no se especifica con cuantos decimales se debe imprimir, saldrá: 0.00, 1.00, 2.00 ... 9.00

Formato general:

```
1          16
PRINT FØRMATS { nombre de var } (valor) / { nombre var } (valor)
               { o }
               { lista de vars } { lista vars }
```

El valor especificado entre paréntesis puede ser un número entero 0, 1, 2, 3, 4, 5 para variables numéricas o A para variables alfanuméricas.

<u>Tipo de variable</u>	<u>Leída en:</u>	<u>Valor en PRINT FØRMATS</u>	<u>Ejemplo SALIDA (impresa)</u>
numérica	F1.0	(0)	1.
"	F1.0	(1)	1.0
"	F1.0	(2)	1.00
"	F1.0	(3)	1.000
"	F1.0	(4)	1.0000
"	F1.0	(5)	1.00000
alfanumérica	A1 (A2, ...A8)	(A)	1

IV.11 El orden de las tarjetas de definición de datos

<u>Status</u>	<u>Palabra de control</u>	<u>Observaciones</u>
opcional	FILE NAME	Requerida si se va guardar Archivo SPSS
obligatoria	VARIABLE LIST	
opcional	SUBFILE LIST	Requerida si hay estructura de subarchivo

obligatoria	INPUT FORMAT	} El orden de estas tarjetas es arbitrario
obligatoria	INPUT MEDIUM	
obligatoria	# ØF CASES	
opcional	MISSING VALUES	
opcional	VAR LABELS	
opcional	VALUE LABELS	
opcional	PRINT FØRMATS	Requerida para variables alfanuméricas

1
1
1
1

1
1
1

1
1
1
1

1
1
1

V. LAS TARJETAS DE DEFINICION DE TAREA

Las tarjetas de control SPSS, llamadas tarjetas de definición de tarea, desarrollan la función de activar la lectura de un archivo, definir y controlar los cálculos a ser realizados sobre los datos.

V.1 TARJETAS DE PROCEDIMIENTOS

Los procedimientos disponibles en el SPSS están constituidos por una o varias subrutinas dentro del sistema. A un procedimiento se denomina el juego de subrutinas que realizan una tarea específica sobre nuestros datos como puede ser: una distribución de frecuencia, un cruce, una regresión, etc.

El juego de subrutinas que conforman un procedimiento es activado por medio de una tarjeta de procedimiento que lleve en el campo de control (col. 1-15) la palabra de control correspondiente.

Ejemplos: Para solicitar una distribución de frecuencia

<u>1</u>	<u>16</u>
CODEBØØK	EDAD, SEXØ, ØCUP, NIVEL

Para solicitar un cruce:

<u>1</u>	<u>16</u>
CROSSTABS	EDAD BY SEXO / EDAD BY NIVEL

Cada procedimiento es activado por medio de una única palabra de control; cada uno de los procedimientos disponibles posee un formato propio que se detalla en la descripción de cada uno, pero todos comparten reglas de sintaxis comunes.

V.2 LA TARJETA OPTIØNS

Esta tarjeta está directamente asociada a la tarjeta de procedimiento y permite al usuario informar al sistema sobre cómo realizar los cálculos. Por ejemplo es por medio de la tarjeta OPTIØNS que el usuario solicita la producción de un histograma en el caso de estar utilizando el procedimiento CODEBØØK. Cada procedimiento posee su lista de opciones asociada.

El formato general es:

1	16
OPTIØNS	lista de números

Ejemplo:

1	16
OPTIØNS	1, 2, 5

El uso de esta tarjeta es opcional, pero si se usa debe aparecer inmediatamente después de la tarjeta de procedimiento. Si no se usa cada procedimiento posee opciones estándares que son asumidas.

V.3 LA TARJETA STATISTICS

Esta tarjeta es similar a la de OPTIØNS y su función es solicitar al sistema el cálculo de las estadísticas asociados al procedimiento en cuestión. Igual que la tarjeta de OPTIØNS sólo puede ser utilizada junto a una tarjeta de procedimiento.

El formato general es:

<u>1</u>	<u>16</u>
STATISTICS	{ lista de números o ALL }

Podemos entonces especificar una lista de números solicitando sólo algunos estadísticos o la palabra clave ALL que informará al sistema para que calcule todos los estadísticos asociados al procedimiento.

Esta tarjeta es opcional y si no se usa no se calcula ningún estadístico. Si se está utilizando la tarjeta OPTIØNS la tarjeta STATISTICS debe aparecer a continuación.

Ejemplos:

1	16
CODEBOOK	EDAD, NIVEL, SEXØ
ØPTIØNS	1, 4
STATISTICS	1, 2, 3

1	16
CØDEBØØK	EDAD, NIVEL, SEXO
STATISTICS	ALL

V.4 LA TARJETA READ INPUT DATA

Esta tarjeta sólo se utiliza para procesar archivos del usuario y su función es informar al sistema para que inicie la lectura de datos. Esta tarjeta debe aparecer inmediatamente después de las tarjetas OPTIONS y STATISTICS del primer procedimiento. Se especifica una sola vez y el campo de especificación (col. 16-80) debe ir en blanco. Es obligatoria su especificación.

Ejemplo:

<u>1</u>	<u>16</u>
⋮	
tarjetas de definición de datos	
⋮	
CØDEBØØK	EDAD, SEXO, NIVEL
ØPTIØNS	1, 4
STATISTICS	ALL
<u>READ INPUT DATA</u>	
CROSSTABS	EDAD BY SEXØ
STATISTICS	ALL

V.5 LA TARJETA PROCESS SBFILES

Esta tarjeta se utiliza sólo cuando nuestros datos se encuentran con una estructura de subarchivo y en este caso su uso es obligatorio.

La tarjeta PROCESS SBFILES debe ir inmediatamente antes de la primera tarjeta de procedimiento y su acción se hace efectiva para todos los procedimientos que le siguen hasta que detecte otra tarjeta PROCESS SBFILES.

Los nombres en el campo de especificación deben estar anteriormente definidos en una tarjeta SUBFILE LIST.

Volvamos a nuestro ejemplo hipotético en que estamos creando una estructura de subarchivo para cada provincia y los nombres de los subarchivos son: PRØV1, PRØV2, PRØV3, PRØV4, PRØV5 y que deseamos obtener distribuciones de frecuencias, cruces u otros cálculos para cada provincia en particular, haciendo ciertas agrupaciones y obviamente para el total del país.

Tenemos:

1. Para cada Provincia

1	16
PROCESS SBFILES EACH	

2. Sólo para provincia 2 y 5

1	16
PROCESS SBFILES (PROV2) (PROV5)	

3. Agrupando provincia 1 con la 3; y agrupando provincias 2, 4, 5; o sea estamos creando 2 regiones.

1	16
PROCESS SBFILES (PRØV1, PRØV3) (PRØV2, PRØV4, PRØV5)	

4. Provincias 1 y 2 individualmente y agrupando 3, 4, 5

1	16
PROCESS SBFILES (PRØV1) (PRØV2) (PRØV3, PRØV4, PRØV5)	

5. Agrupando todas las provincias para obtener los cálculos a nivel del país

1	16
PROCESS SBFILES ALL	

Como se puede observar hay dos palabras claves que simplifican la elaboración de nuestro programa SPSS:

EACH .. el sistema tratará cada subarchivo en forma individual

ALL ... el sistema ignora la estructura, tratando todos los subarchivos como uno solo.

V.6 PALABRAS DE CONTROL DE LAS TARJETAS DE DEFINICION DE TAREA

1

16

CØNDESCRIPTIVE
CØDEBØØK
MARGINALS
FASTMARG
CRØSSTABS
FASTABS
SCATTERGRAM
BREAKDØWN
PEARSØN CØRR
NONPAR CØRR
PARTIAL CØRR
REGRESSIØN
FACTØR
GUTTMAN SCALE

} palabras de control para las
tarjetas de procedimiento

ØPTIONS
STATISTICS
PRØCESS SBFILES
READ INPUT DATA

} palabras de control de las tarjetas
de definición de tarea asociadas a
la tarjeta de procedimiento

VI. LAS TARJETAS DE PROCESO "RUN CARDS"

La palabra RUN significa una corrida o un proceso. Un proceso en SPSS implicará uno o más cálculos: podemos en una corrida elaborar distribuciones de frecuencia, cruces, regresiones, etc. sobre nuestro juego de datos.

VI.1 LA TARJETA RUN NAME

Esta tarjeta permite al usuario definir un título para el proceso que realiza. El título especificado por medio de esta tarjeta saldrá impreso como encabezado de todas las páginas utilizadas en una corrida. El largo máximo del título es de 64 caracteres del conjunto estándar IBM/360. Conviene observar que este título en general describirá los datos (encuesta) sobre la cual realizamos el proceso.

El formato general es:

1	16
RUN NAME	título

Ejemplo:

1	16
RUN NAME	ESTO ES UN EJEMPLO DE TITULO

El uso de esta tarjeta es opcional.

VI.2 LA TARJETA FINISH

Esta tarjeta informa al sistema el final del proceso. Su especificación es obligatoria y debe ser la última tarjeta del programa SPSS.

1	16
FINISH	

VI.3 LA TARJETA KEYPUNCH

Esta tarjeta sólo se usa cuando las tarjetas del programa SPSS están perforadas en máquina IBM-026. Por omisión el SPSS asume 029.

El formato es:

1	16
KEYPUNCH	026

VI.4 LA TARJETA PRINT BACK

El sistema SPSS normalmente realiza la impresión de las tarjetas de control (del programa). Si por alguna razón el usuario desea que no se impriman, deberá especificar:

1	16
PRINT BACK	NØ

VI.5 LA TARJETA NUMBERED

Normalmente el campo de especificación del sistema se extiende de columna 16-80. La tarjeta NUMBERED informa al sistema que el campo de especificación se reduce a columnas 16-72; reservando así las columnas 73-80 para un número de secuencia del programa especialmente útil para programas muy largos.

La declaración es:

1	16
NUMBERED	YES

VI.6 LA TARJETA COMMENT

Permite insertar comentarios en cualquier parte de las tarjetas de nuestro programa. Permite cuantas tarjetas de continuación como se desee, dejando el campo de control (col. 1-15) en blanco.

1	16
COMMENT	comentarios

VI.7 LA TARJETA DOCUMENT

Esta tarjeta permite al usuario retener en el archivo SPSS en forma permanente toda la información relevante relacionada a transformaciones, recodificaciones y creación de nuevas variables que realiza sobre su juego de datos. A pesar de que su uso puede tornarse tedioso en la medida que requiere duplicar en el campo de especificación lo que declaramos como una instrucción, hay que resaltar que permite al usuario tener un archivo SPSS auto-documentado. La información entrada por medio de DOCUMENT se recupera posteriormente por medio de la tarjeta DUMP.

Permite tantas tarjetas de continuación que se requiera. El formato es:

1	16
DOCUMENT	mensajes relevantes

Ejemplo:

DOCUMENT	LA VARIABLE ING ... INGRESO DE LA FAMILIA HA SIDO RECODIFICADA DE LA SIGUIENTE MANERA: 1 ... 0 - 1000 colones; 2 ... 1001 - 3000, 3 ... 3001 - 5000; 4 ... 5001 y +
----------	---

VI.8 LA TARJETA DUMP

Esta tarjeta permite al usuario recuperar la información que describe un archivo SPSS (del sistema). El formato es:

1	16
DUMP	lista de palabras claves

La lista puede contener una o más de las siguientes palabras claves:

- VARLIST imprimirá una lista de todas las variables del del archivo en el orden que fueron grabadas. Incluye las variables declaradas en VARIABLE LIST y aquellas creadas por transformaciones
- VARINFØ imprimirá una lista de los formatos de impresión (PRINT FØRMATS) y valores faltantes declarados para cada variable del archivo.
- SUBDIRECTORY imprimirá una lista de todos los subarchivos y el número de casos en cada uno.
- LABELS imprimirá una lista de todos los títulos asociados al archivo y sus variables.
- DØCUMENT..... imprimirá toda la información ingresada al archivo por medio de la instrucción DOCUMENT
- SØRTVARS imprimirá una lista de todas las variables en orden alfabético.
- LABCARDS imprimirá los títulos de variables y títulos de categorías de variables tal como se especificaron originalmente.
- COMPLETE impresión para cada variable del nombre, títulos asociados, valores faltantes y formatos de impresión.

VI.9 EL ORDEN DE LAS TARJETAS DE PROCESO

<u>1</u>	<u>16</u>	
NUMBERED	YES	opcional
KEYPUNCH	026	obligatoria si se usa IBM-026
PRINT BACK	NØ	opcional
RUN NAME	título	opcional

VARIABLE LIST	lista vars	} tarjetas de definición de datos entre las cuales se puede insertar COMMENT y DOCUMENT
:		
COMMENT	comentarios	
:		
DOCUMENT	mensajes	
:		

CODEBOOK		
READ INPUT DATA		
:		
FINISH		obligatoria

VII. GENERACION Y PROCESAMIENTO DE ARCHIVOS CON SPSS

VII.1. PROCESAMIENTO DE ARCHIVOS DEL USUARIO (BCD)

Como ya se describió anteriormente, un archivo del usuario es aquel que contiene los datos originales, provenientes de una etapa de digitación mediante la cual se almacenó la información bajo una forma legible por máquina. Los datos en esta forma BCD (Binary Coded Decimal) puede residir en tarjetas perforadas, cintas magnéticas o discos. Para procesar los datos en BCD, utilizando el SPSS se requiere la especificación de las tarjetas de definición de datos ya descritos. Un proceso inicial del archivo del usuario puede permitir la obtención de un primer juego de cálculos y además la generación de un archivo del sistema SPSS (archivo SPSS) mediante el uso de la tarjeta SAVE FILE. En el archivo SPSS, toda la información de definición y descripción de los datos tales como nombres de variables, sus formatos, nombres de categorías conjuntamente con los mismos datos quedan almacenados en forma permanente en un medio legible a máquina.

El orden de las tarjetas de control SPSS requeridas para procesar un archivo del usuario (BCD) se muestra en la tabla. Los 3 status básicos son: (a) obligatoria- o sea siempre se requiere la especificación; (b) condicional -son requeridas dependiendo de alguna acción especial y (c) opcional -son aquellas que solo se especifica si se desea; en general desarrollan una función de descripción.

TABLA: Orden de las tarjetas de control para procesar archivos BCD

STATUS	CAMPO DE CONTROL	OBSERVACIONES	
Opcional	RUN NAME		
Condiciona	FILE NAME	Obligatoria si se va a crear un archivo SPSS	↑ TARJETAS DE DEFINICION DE DATOS ↓ ↑ TARJETAS DE DEFINICION DE LA 1a. TAREA ↓ ↑ TARJETAS DE DEFINICION DE TAREA PARA LAS SUBSIGUIENTES ↓
Obligatoria	VARIABLE LIST	Obligatoria si hay estructura de sub-archivos	
Condiciona	SUBFILE LIST		
Obligatoria	INPUT MEDIUM	El orden de estas tarjetas es arbitrario Obligatoria para variables alfa numéricas	
Obligatoria	# OF CASES		
Obligatoria	INPUT FØRMAT		
Opcional	MISSING VALUES		
Opcional	VAR LABELS		
Opcional	VALUE LABELS		
Condiciona	PRINT FØRMATs		
Condiciona	PRØCESS SBFILES	Obligatoria solo para archivos con sub-archivos	
Obligatoria	*Tarjeta de Procedimiento*		
Opcional	OPTIONS		
Opcional	STATISTICS		
Obligatoria	READ INPUT DATA		
(Las tarjetas conteniendo los casos -datos aparecieron aquí si se definió INPUT MEDIUM como CARD)			
	PRØCESS SBFILES		
	Tarjeta de Procedimiento		
	ØPTIONS		
	STATISTICS		
Obligatoria	FINISH		

OBSERVACION: En el caso que se utilice la convención ESTIMATED en la tarjeta # OF CASES la tarjeta INPUT MEDIUM deberá aparecer antes.

En el siguiente ejemplo N^o1 se muestra el mínimo de tarjetas requeridas para un sencillo proceso de obtención de distribuciones de frecuencia con el procedimiento CØDEBØØK:

Ejemplo N^o1

1 16

VARIABLE LIST EDAD, SEXO, EDUC, INGRESØ
INPUT MEDIUM CARD
ØF CASES 10
INPUT FORMAT FIXED (5X, F2.0, 2F1.0, 2X, F7.0)
CØDEBØØK ALL
STATISTICS ALL

READ INPUT DATA

1	3013	0006000
1	1021	0000000
1	1513	0006000
2	2024	0008000
2	4013	0005000
2	5021	0000600
3	6021	0000050
3	2514	0006000
3	3615	0009000
3	4125	0008000

FINISH

En el ejemplo N^o2 se muestra el programa SPSS para el proceso de un archivo del usuario (BCD) con estructura de subarchivos y utilizando todas las tarjetas opcionales y condicionales del sig tema. Estas últimas tarjetas están señaladas con un asterisco ya que su presencia no es obligatoria sino que ayudan a describir nuestros datos.

Ejemplo N°2

```

1                16
RUN NAME          EJEMPLO DE PROCESO SPSS CON ARCHIVO BCD
(*) FILE NAME     NAPOLEON, ESTE ES EL TITULO EXTENDIDO
      .TABLE LIST  PRØVIN, EDAD, SEXO, EDUC, INGRESØ
SUBFILE LIST      PRØV1, PRØV2, PRØV3
INPUT MEDIUM     CARD
# ØF CASES       3, 3, 4
INPUT FØRMAT     FIXED(JX, F1.0, 3X, F2.0, 2F1.0, 2X, F7.0)
(*) MISSING VALUES EDAD(97,98,99)/SEXO, EDUC(9)
(*) VAR LABELS   PROVIN, PROVINCIA /
                  EDAD, EDAD DEL ENTREVISTADØ /
                  EDUC, NIVEL DE EDUCACION /
                  INGRESO, INGRESØ FAMILIAR /
VALVE LABELS     PRØVIN, (1) PROVINCIA CENTRAL (2) PROVINCIA NØRTE
                  (3) PROVINCIA SUR /
                  SEXO, (1) HOMBRES (2) MUJERES (9) IGNØRADO /
                  EDUC, (1) NINGUNO (2) PRIMARIA (3) SECUNDARIA
                  (4) UNIVERSITARIA (9) IGNØRADO
(*) PRINT FØRMATS PRØVIN TØ INGRESØ (0)
PROCESS SBFILES EACH
CØDEBOOK        EDAD TØ INGRESØ
(*) STATISTICS   ALL
READ INPUT DATA
  1   3013   0006000 }
  1   1021   0000000 } subarchivo PRØV1
  1   1513   0006000 }
  2   2024   0008000 }
  2   4013   0005000 } subarchivo PRØV2
  2   5021   0000600 }
  3   6021   0000050 }
  3   2514   0006000 } subarchivo PRØV3
  3   3615   0009000 }
  3   4125   0008000 }

```

1 16

NØNPAR CØRR EDAD TØ INGRESØ WITH EDAD TØ INGRESO
ØPTIØNS 1, 3
STATISTICS 1, 2
PROCESS SBFILES (PRØV1, PRØV2) (PRØV3)
CROSSTABS EDAD BY SEXO, EDUC
STATISTICS ALL
SAVE FILE
FINISH

VII.2. LA TARJETA SAVE FILE: GENERACION DE ARCHIVOS DEL SISTEMA SPSS

El archivo del sistema SPSS (archivo SPSS) es una entidad permanente de autodocumentación además de garantizar al usuario una máxima eficiencia de proceso ya que es almacenado con bloqueos especiales en forma binaria y la rutina de lectura de los archivos SPSS es más eficiente que la de lectura de archivos BCD.

Un archivo SPSS puede ser generado por medio de la tarjeta "SAVE FILE" como finalización de un proceso preliminar de los datos en BCD. El archivo SPSS contiene:

- Los casos (datos) del usuario.
- Toda la información que definen y describen los datos.

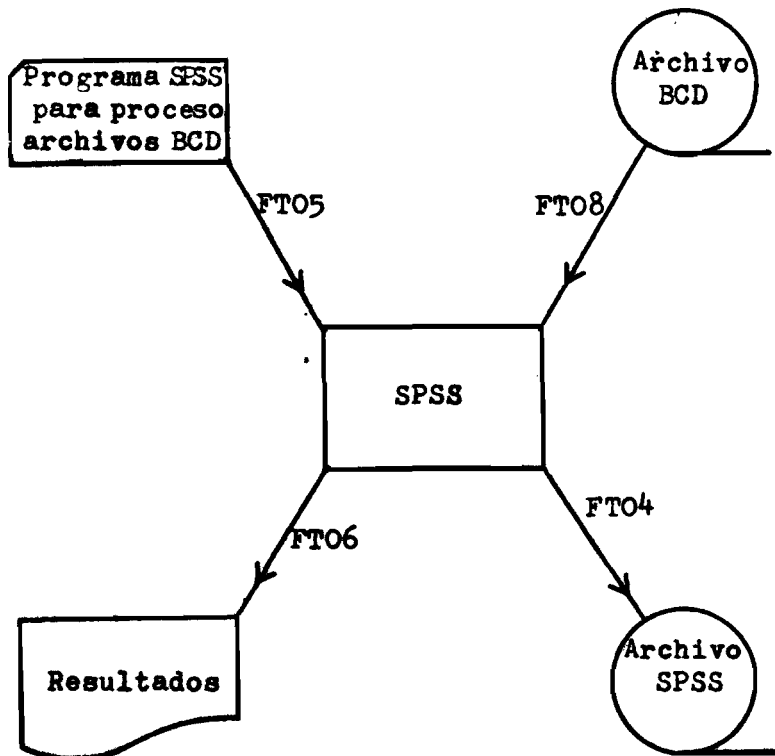
El formato general de la tarjeta es:

1	16
SAVE FILE	

Las siguientes unidades lógicas deberán ser especificadas en JCL (JOB control Language) para leer archivo del usuario y grabar un archivo SPSS, si la lectura del archivo del usuario es desde cinta o disco:

- FT08 Lectura de archivo BCD
- FT04 Grabación de archivo SPSS.

Esquemáticamente tendríamos:



La estructura en que es almacenado el archivo SPSS consta de 2 partes: la primera donde se encuentra toda la información de definición y descripción de los datos y una segunda parte que son los valores de los casos mismos, como se muestra en la figura siguiente.

La tarjeta "SAVE FILE" DEBE IR INMEDIATAMENTE ANTES DE LA TARJETA "FINISH". En el ejemplo N°2 se puede ver la posición de la tarjeta SAVE FILE.

Estructura del archivo del Sistema SPSS

INFORMACION DE DEFINICION DEL ARCHIVO:

- nombre del archivo y título extendido
- nombres de variables, títulos y formatos
- valores de categorías y títulos de las categorías
- indicadores de valores faltantes

VARIABLE →	Primera	Segunda	Ultima
Primer Subarchivo	Primer caso		...	
	Segundo caso			
	Tercer caso			
	...			
	Ultimo caso			
Segundo Subarchivo	Primer caso			
	Segundo caso			
	Ultimo caso			
...				
	Primer caso			
	Segundo caso			
	Ultimo caso			

VII.3. PROCESO DE ARCHIVOS DEL SISTEMA SPSS

El hecho de haber generado un archivo SPSS no implica que los datos no puedan ser nuevamente alterados mediante nuevas recodificaciones, transformaciones, re-definición de valores faltantes, títulos, etc.

Todo eso es perfectamente factible teniendo como entrada un archivo SPSS y se puede además generar otro archivo SPSS para disponer así de otro archivo con las transformaciones realizadas.

Las unidades lógicas y un diagrama para procesar un archivo SPSS, generando además un nuevo archivo SPSS se muestra a continuación:

Programa para proceso

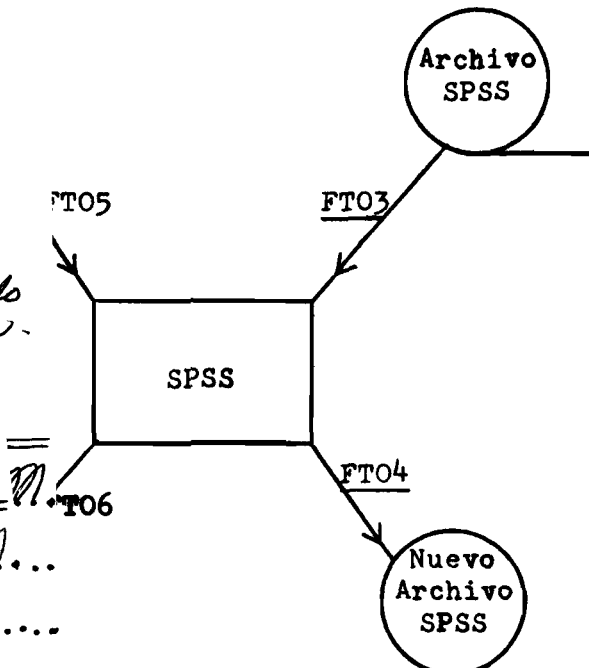
Nº: 276100

Nº: 276100

Autor:

Título:

*Unos datos de SPSS
de análisis de datos*



Fecha	Nombre	Firma
29/10/80	<i>[Handwritten Name]</i>	<i>[Handwritten Signature]</i>
.....
.....
.....
.....
.....
.....
.....
.....
.....

...inir la unidad lógica FT03 en el JCL pa-
SPSS y la unidad FT04 si va a generar un

[LE

l tiene la función de activar la lectu-
ado en un proceso anterior. El formato

1	16
GET FILE	nombre del archivo

VII.3.2. El orden de las tarjetas de Control para procesar archivos SPSS

Para procesar el archivo SPSS generado en el ejemplo N°2, que posee una estructura de subarchivos y generar un nuevo archivo, las tarjetas de control SPSS serían:

```
1           16
RUN NAME      EJEMPLO DE PROCESO DE ARCHIVOS  SPSS
GET FILE      NAPØLEØN
FILE NAME     NAPØLEØ2, SEGUNDO ARCHIVO NAPOLEØN
PRØCESS SBFILES EACH
CØDEBØØK      ALL
STATISTICS    ALL
PROCESS SBFILES ALL
CRØSSTABS     EDAD  BY  SEXO  BY  EDUC
SAVE FILE
FINISH
```

El orden general sería:

```

RUN NAME      opcional
GET FILE      obligatoria
-----
:
tarjetas de
definición de datos
:
-----
:
tarjetas de
definición de
tarefas
-----
SAVE FILE     opcional
FINISH        obligatoria
```

A título de ilustración y en forma aproximada, un proceso para 70 000 casos leyéndose unas 12 variables numéricas, el proceso con el archivo BCD tardó 2 horas al paso que la lectura desde un archivo SPSS solamente tardó unos 20 minutos.

VII.3.3. Información de Definición de datos para procesos de archivos SPSS

El SPSS permite al usuario cambiar o agregar la siguiente información de definición de datos durante el proceso de un archivo SPSS :

- (a) Nombre del archivo y título extendido (generando un nuevo archivo con SAVE FILE).
- (b) Valores faltantes - MISSING VALUES
- (c) Títulos de categorías - VALUE LABELS
- (d) Títulos de variables - VAR LABELS
- (e) Formatos de impresión - PRINT FORMATS

Si se desea retener en forma permanente las decisiones o cambios entrados en el proceso de un archivo SPSS, sencillamente se genera otro archivo SPSS por medio de la tarjeta SAVE FILE. Si no se genera un nuevo archivo SPSS las adiciones y/o cambios introducidos solamente son efectivos para el proceso que se realiza, no afectando el archivo de entrada.

En general las modificaciones a la información ya existente se hace por medio de reemplazo siguiéndose las siguientes normas:

1. Un nuevo FILE NAME asociado a un SAVE FILE causa la generación de un nuevo archivo SPSS con el nuevo nombre definido.
2. Cada VALUE LABEL se constituye en una entidad independiente. Esto significa que para una cierta variable podemos solo realizar un reemplazo en el título de una categoría específica sin que los títulos de las demás categorías de esa misma variable sean alteradas.
3. EL PRINT FORMATS Y VAR LABELS asociados a cada variable también pueden ser considerados como entidades independientes y pueden ser reemplazados sin alterar las demás variables. Si tenemos una variable sin título extendido definido anteriormente, podemos igualmente definir un título en el proceso de archivo SPSS mediante la tarjeta VAR LABELS.
4. Los valores faltantes asociados a una variable o un grupo de variables no son entidades independientes pero pueden ser tratadas en grupo.

La inserción de una nueva declaración MISSING VALUES para una variable o un grupo de variables, ocasiona que todos los valores faltantes anteriormente declarados sean borrados y sustituidos por los nuevos declarados.

Por ejemplo, si tenemos una variable con declaración de los valores 98,99 como faltantes y deseamos ahora tener como valores faltantes 0,98,99 debemos declarar MISSING VALUES nombre (0,98,99).

El usuario puede además eliminar (borrar) los valores faltantes declarados para una o varias variables utilizando los paréntesis sin nada "()". El formato general para esto sería:

```
1                16

MISSING VALUES { nombre de var } ()
                  o
                  { lista de vars }
```

Una ilustración, utilizando el ejemplo N°2 de este capítulo, sería:

```
1                16

RUN NAME          PROCESO ARCHIVO NAPOLEON
GET FILE          NAPØLEØN
FILE NAME        NAPØLEØ2
VAR LABELS       INGRESØ, INGRESØ DEL ENTREVISTADØ
VAR LABELS       EDUC, EDUCACION DEL JEFE
VALUE LABELS     SEXO, (8) IGNØRADO (9) INDEFINIDØ
PROCESS SBFILES  ALL
CØDEBØØK        ALL
SAVE FILE
FINISH
```

APENDICE A - Capítulos I al VII

EL CAMBIO A LA VERSION 6

1. Introducción

Considerando que a la fecha (julio/76) de escribir estos apuntes del Capítulo VIII en adelante ya se dispone de la nueva versión del SPSS (Versión H, release 6) el propósito de este apéndice es mostrar los cambios en relación a lo anteriormente escrito y publicado para permitir seguir la redacción sobre la versión H, release 6 que trae una serie de nuevos procedimientos.

A pesar de que el release 6 sigue la misma filosofía de las versiones anteriores, los autores del SPSS lo sometieron a una verdadera re-estructuración y uniformización del lenguaje de tal forma que las diferencias con las versiones anteriores es bastante grande.

Como el tiempo no permite una nueva redacción de los capítulos ya publicados, optamos a editar este apéndice con los cambios y seguir los próximos capítulos ya dentro de las reglas de la nueva versión. Cabe aclarar además, que felizmente la parte descrita en los capítulos I al VII se constituye en aquella que menos modificaciones ha sufrido.

2. Tarjetas de Definición de datos

2.A. Variable List

Antes el uso de la convención TO exigía el prefijo VAR; ahora permite el uso de cualquier prefijo.

Ejemplos:

```
-VARIABLE LIST  V1 TO V50
  genera       V1, V2, ... V10, V11, ... V50

-VARIABLE LIST  V001 TO V050
  genera       V001, V002, ... V050 que es distinto del ejemplo anterior
  V1 ≠ V001

-VARIABLE LIST  EDAD,SEXO,INGR1 TO INGR50
  genera       INGR1,INGR2, ... INGR10, ... INGR49,INGR50
```

2.B Subfile list

Antes esta tarjeta estaba asociada a la # OF CASES. Ahora el número de casos de cada subarchivo se especifica en la misma tarjeta SUBFILE LIST y no se debe especificar N OF CASES.

Formato General

1	16
SUBFILE LIST	nombre subarchivo ₁ (n° casos subarchivo ₁), ... nombre subarchi _i (n° casos i), nombre subarchn _n (n° casos n)

Ejemplos:

```
SUBFILE LIST  SJOSE (1055), PUNTAR (1976),
               LIMON (2567)
```

O sea:

```
subarchivo    SJOSE trae 1055 casos
"             PUNTAR " 1976  "
"             LIMON  " 2567  "
```

2.C N of cases

1	16
N OF CASES	{ n° de casos en el archivo o } { UNKNOWN }

Ejemplos

```
N OF CASES    12567
N OF CASES    UNKNOWN
```

Al especificarse la palabra clave UNKNOWN el sistema leerá el archivo hasta detectar fin de archivo. La tarjeta N OF CASES no debe ser especificada si se está declarando SUBFILE LIST. En el caso que se especifique el sistema ignora la tarjeta N OF CASES, prevaleciendo lo declarado en SUBFILE LIST

2.D Input Format

Variables tipo A - máximo es A4

Esta tarjeta permite ahora la lectura de variables en binario.

Formato General

1	16
INPUT FORMAT	{ FREEFIELD FIXED BINARY }

BINARY

Las variables deben estar almacenadas en "palabras". El tipo de variables no necesita ser especificada pues ya están en la representación binaria interna de la máquina - para su manipuleo posterior si necesitamos saber si una variable es tipo A o F. Para la lectura sólo se requiere saber la posición de las variables.

Hay 4 elementos de formato

V - debe haber uno para cada var declarado en la VARIABLE LIST

S - saltar variables

P_n - posicionese en la variable n

/ - para saltar registros

Ejemplos

(a)

VARIABLE LIST VAR1,VAR10,VAR20,VAR22, TO VAR26
INPUT FORMAT BINARY (V,P10,V,P20,V,S,5V,10S)

En este caso un total de 8 variables serán leídas de un caso con un total de 36 variables. Las variables leídas son la 1, 10, 20, 22, 23, 24, 25, 26.

(b)

VARIABLE LIST X3, X6, X9, X12, X15, X18
INPUT FORMAT BINARY (6(2S.V))

De un total de 18 variables serán leídas:
3a, 6a, 9a, 12a, 15a, 18a

2.E Data list

Esta tarjeta es una fusión de VARIABLE LIST e INPUT FORMAT en una sola instrucción. A pesar de que es interesante su uso para ciertos casos nos limitamos a hacer mención a su existencia pues su descripción de uso es relativamente compleja.

3. Tarjetas de Definición de Tareas

3.A Run Subfiles

Esta tarjeta reemplaza PROCESS SUBFILES y el cambio es sólo de la palabra de control. El campo de especificación es el mismo descrito para PROCESS SUBFILES

3.B RAW OUPUT UNIT

El SPSS posee una serie de procedimientos que permiten la salida de datos tanto para usarlos como entrada a otro programa como para volver al uso del SPSS tales como WRITE CASES, PEARSON CORR (perforación o grabación de las matrices de correlación), etc.

La unidad lógica estándar que asume el SPSS es la FT09 que sería equivalente a especificarse "RAW OUPUT UNIT 09". Por una serie de razones en un mismo proceso puede tornarse incómodo y hasta inconveniente tener todas las salidas por la misma unidad lógica, lo cual se soluciona mediante esta instrucción RAW.

Formato General

1	16
RAW OUPUT UNIT	n

Valores permitidos para n: 9, 15, 16, 17, 18, 19, 20

La instrucción RAW OUTPUT UNIT debe preceder al procedimiento(s) al cual(es) afectará y queda en efecto hasta que detecte otra RAW OUTPUT UNIT.

Ejemplo

1	16
RUN NAME	EJEMPLO DE RAW OUTPUT UNIT
GET FILE	NAPOLEON
WRITE CASES	(F6.0,F2.0,F1.0) SEQNUM, EDAD, EDUC
RAW OUTPUT	16
PEARSON CORR	EDUC,EDAD,OCUP,INGRES
OPTIONS	4
NONPAR CORR	X1 TO X10
OPTIONS	4
FINISH	

Los casos grabados (perforados) por la tarjeta WRITE CASES se harán hacia la unidad lógica FT09 que es la que asume el sistema si no se especifica RAW OUTPUT UNIT.

Opción 4 para PEARSON y NONPAR CORR ocasiona la grabación (perforación) de las matrices de correlación. En este caso la salida se realiza hacia la unidad lógica FT16 ya que se especificó antes de estos procedimientos "RAW OUTPUT UNIT 16"

3.C. Resumen de las tarjetas de Definición de Tarea

1	16	
RAW OUTPUT UNIT RUN SUBFILES	}	asociadas a tarjetas de procedimientos
CONDESCRIPTIVE FREQUENCIES AGGREGATE CROSSTABS BREAKDOWN T-TEST PEARSON CORR NONPAR CORR SCATTERGRAM PARTIAL CORR REGRESSION ANOVA ONEWAY DISCRIMINANT FACTOR CANCORR GUTTMAN SCALE		}
OPTIONS STATISTICS	}	

4. Tarjetas de proceso "RUN CARDS"

RUN NAME
<u>TASK NAME</u>
FINISH
<u>PAGESIZE</u>
PRINT BACK
COMMENT
DOCUMENT
NUMBERED

Las dos subrayadas son nuevas y de gran utilidad como veremos. Las demás permanecen como se describió en Capítulo VI.

4.A Tarjeta TASK NAME

Esta tarjeta permite la especificación de título adicional asociado a una tarjeta de procedimiento o sea asociado a una tarea.

El título especificado en una TASK NAME queda en efecto hasta que detecte obra TASK NAME. Se debe insertar esta tarjeta inmediatamente antes de la tarjeta de procedimiento. Así el encabezado de las tareas tendrán el título especificado en RUN NAME y adicionalmente el título especificado en TASK NAME.

Ejemplos

1	16
*WEIGHT	HIJOS
TASK NAME	PONDERADO POR NUMERO DE HIJOS
CROSSTABS
:	
*WEIGHT	EMBAR
TASK NAME	PONDERACION POR EMBARAZOS
CROSSTABS
OPTIONS	
STATISTICS	

4.B Tarjeta PAGESIZE

Formato General

1	16
PAGESIZE	{ n } { 0 } { NOEJECT }

n: especifica el número de líneas por página. Si no se especifica el SPSS asume n = 55 líneas por página - esto es para papel de 11 pulgadas y cinta impresora de 6 líneas/pulgada. Para instalaciones con cinta de control de impresora a 8 líneas por pulgada el valor de n puede variar de 75 a 80 según se desee.

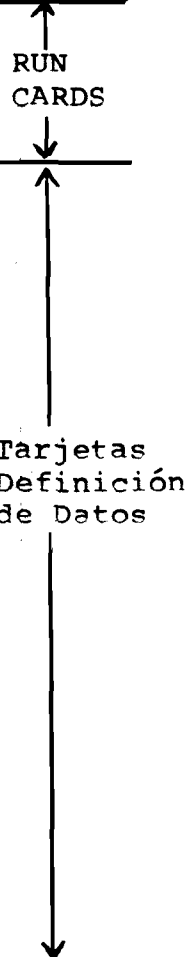
NOEJECT: especifica que no se llevará control de salto de página -esto puede ocasionar un poco de dificultad en la lectura de las tablas pero puede ahorrar mucho papel como es el caso de muchos cruces de 2 x 2.

El efecto de una instrucción PAGESIZE es válido hasta que detecte otra instrucción PAGESIZE.

5. Generación y Procesamiento de Archivo

El Capítulo VII prácticamente no ha sufrido cambios por lo cual nos limitamos a reproducir la tabla con el orden para el proceso de archivos BCD

status	Campo de control	Observaciones	Precedencia
opcional	NUMBERED		1
opcional	PRINT BACK		2
opcional	PAGESIZE		2
opcional	RUN NAME		3
condicional	FILE NAME	obligatoria si se va salvar archivo	4
requerida	{ VARIABLE LIST o DATA LIST }		
requerida	INPUT MEDIUM		6
condicional	SUBFILE LIST	requerida si hay estructura de subarchivos	7
condicional	N OF CASES	requerida si no hay estructura de subarchivos	7
condicional	INPUT FORMAT	requerida si se usa VARIABLE LIST	8
opcional	MISSING VALUES		8
opcional	VAR LABELS		8
opcional	VALUE LABELS		8
condicional	PRINT FORMATS	requerida para variable tipo A (alfanumérica)	8



status	Campo de control	Observaciones	Precedencia
opcional	TASK NAME		"mejor posición"
opcional	RAW OUTPUT UNIT		9
condicional	RUN SUBFILES	usado sólo para archivos con subarchivos	Tarjetas de Definición de la tarea
requerida	"tarjeta procedimiento"		10
opcional	OPTIONS		11
opcional	STATISTICS		11
requerida	READ INPUT DATA		12
tarjetas de datos aparecerían aquí si los datos están almacenados en tarjetas			
	TASK NAME		
	RAW OUTPUT UNIT		
	RUN SUBFILES		tarjetas de las tareas subsiguientes
	"tarjeta de procedimiento"		
	OPTIONS		
	STATISTICS		
opcional	DOCUMENT		mejor posición
opcional	SAVE FILE	para salvar archivo del sistema	
requerida	FINISH		última tarjeta

6. WRITE FILEINFO reemplaza a DUMP

formato general

```

1          16
WRITE FILEINFO  VARIABLES=ALL / palabras claves e ALL
    
```

Palabras claves:

FILENAME: edita el nombre y título extendido definidos en FILE NAME

VARLIST : listado de todas las variables definidas en VARIABLES=

SUBFLIST: nombres de subarchivos y respectivos números de casos.

NCASES : edita la tarjeta N OF CASES y el número de casos.

MISVALS : variables y respectivos missing values declarados.

PRNFMTS : edita los print formatos

VARLABS : edita los VAR LABELS

VALLABS : edita los VALUE LABELS

DOCUMENTS: información ingresada por DOCUMENT

ALL : es equivalente a la especificación de todas las palabras claves - edita toda la información asociada al archivo.

Suponiendo que la VAR002 es continua, de la cual desconocemos sus límites inferior y superior, la función que desarrolla las dos palabras claves LOWEST e HIGHEST es impedir posibles errores originados por tal motivo.

Así la VAR002 recodificada sería:

<u>valores originales</u>	<u>nuevos valores</u>
mínimo hasta 10	1
10.1 hasta 100	2
100.1 hasta máximo	3

El THRU indica los valores entre los cuales, incluyendo los extremos, que deberán ser recodificados hacia el nuevo valor.

Ejemplos de listas de valores:

(3,4,5 THRU 10 = 1) o sea los valores 3,4,5,6,7,8,9,10 serán recodificados al valor 1.

(LOWEST THRU 1,2,3 = 1) desde el mínimo hasta 1, el 2 y 3 serán recodificados al nuevo valor 1.

La palabra clave BLANK: diferenciación entre blanco y cero para variables numéricas

En muchas encuestas se utiliza el cero como un código válido y el blanco para el caso en que la pregunta no se aplica. Normalmente, el SPSS trata los blancos, leídos como variables numérica (formato Fw.d.) como cero; a no ser que utilizemos la palabra clave BLANK para recodificar lo que es blanco a otro valor.

Ejemplos

(a) Uso correcto

1	16
RECODE	OCUP (BLANK = 9) (0 THRU 200 = 1) (201 THRU 500 = 2) (501 THRU 999 = 3)
VALUE LABELS	OCUP (1) PROFESIONALES (2) EMPLEADOS PUBLICOS (3) OBREROS Y OTROS (9) NO SE APLICA

(b) Uso incorrecto

```
1          16
RECODE     OCUP (0 THRU 200  1)(201 THRU 500 = 2)
           (501 THRU 999 = 3)(BLANK = 9)
```

El error del ejemplo (b) está en que el (BLANK = valor) aparece posterior a la recodificación del cero y por lo tanto los blancos, tratados como cero serán recodificados al valor 1 y no 9 como se desea.

Entonces la regla para el uso correcto de la palabra clave BLANK es que la recodificación (BLANK = valor) debe venir antes de la recodificación del valor cero. Si esto no se cumple se comete un error lógico y los blancos serán recodificados al mismo valor del cero, sin que el sistema detecte error de sintaxis pues en realidad la sintaxis es correcta.

La palabra clave ELSE

```
RECODE     VAR001,VAR050 (0,1,2 = 1)(3 = 2)
           (4,5,9 = 3)(ELSE = 4)
```

O sea cualquier valor de las variables que no estén especificadas en las listas de recodificación anteriores serán recodificados al valor 4.

El tratamiento de variables alfanuméricas: CONVERT

En muchos casos el usuario es obligado a ingresar variables bajo la lectura alfanumérica debido a la codificación adoptada. El hecho de no poder realizar operaciones aritméticas con variables alfanuméricas y además tomando en consideración que los procedimientos que permiten manipular variables numéricas son más eficientes y eficaces que aquellos que permiten tratar variables alfanuméricas, fundamenta la necesidad y lo recomendable de recodificar tales variables a valores numéricos.

Algunos ejemplos ilustran el uso del RECODE y la palabra clave CONVERT:

```
1          16
RECODE     X1 TO X10 ('0' = 0)('1' = 1)('2' = 2)
           ('3' = 3)('4' = 4)('5' = 5)('6' = 6)('7' = 7)
           ('8' = 8)('9' = 9)('-' = 11)('&' = 12)
           (ELSE = 13)
```

La instrucción equivalente a la anterior sería:

```
RECODE          X1 TO X10 (CONVERT)
```

O sea, la palabra CONVERT para el caso de variables alfanuméricas de 1 columna realiza la siguiente equivalencia numérica

```
'0', '1', ... '9'  —————> numérico 0, 1, ... 9
                    ' '  —————>          "    11
                    '&'  —————>          "    12
```

blanco, o cualquier otro carácter —————> numérico 13

El uso del CONVERT para variables alfanuméricas de 2 o más columnas debe ser cauteloso, pues como se muestra a continuación, no garantiza la equivalencia única.

El algoritmo que utiliza el SPSS para estos casos es el siguiente:

- (a) transforma el carácter de la extrema izquierda hacia su equivalente numérico utilizando la regla anteriormente mostrada cuando se trata de variables de 1 columna. Llamemos a este valor de X
- (b) Multiplica X por 10^{w-1} donde w = largo del campo.

Este algoritmo es utilizado siempre que detecta un carácter no numérico (inclusive el blanco [' ']) en la composición de la variable; no importando que dicho carácter sea el 1º, 2º, ... etc.

Ejemplos: [donde \emptyset = blanco]

<u>Valor</u> <u>Alfanumérico</u>	<u>Equivalente</u> <u>numérico</u>	<u>Carácter</u> <u>de la</u> <u>izquierda</u>	<u>w</u>	<u>Cálculo</u>
' \emptyset 3' 130	' \emptyset '	2	$13 \cdot 10^{2-1} = 130$
' \emptyset 8' 110	'-'	2	$11 \cdot 10^{2-1} = 110$
'8 \emptyset ' 80	'8'	2	$8 \cdot 10^{2-1} = 80$
'&88' 1200	'&'	3	$12 \cdot 10^{3-1} = 1200$
'70&' 700	'7'	3	$7 \cdot 10^{3-1} = 700$
'010' 10			No se aplica pues todos los caracteres son numéricos

El ejemplo siguiente ilustra la mejor forma y la facilidad que representa el uso del CONVERT para variables alfa - numéricas de 2 o más columnas (caracteres).

Llamemos de OCUPAC el nombre de la variable ocupación que requirió la lectura en formato A3 por la existencia de caracteres no numéricos en la composición del código. Sobre esta variable ya tenemos previo conocimiento que posee códigos desde '001', '002', ... '999' y además '---', '-&&', '&&&', '&--'. El conocimiento previo sobre los códigos se puede obtener tanto de las normas de perforación o en forma más segura de un proceso previo en lo cual obtuvimos la distribución de frecuencia de dicha variable en forma alfanumérica.

```
RECODE          OCUPAC ('---' = 1001) ('-&&' = 1002)
                ('&--' = 1003) (CONVERT)
```

El CONVERT indica al sistema que todos los demás valores deberán ser convertidos a su equivalente numérico o sea

'001'	...	1
'002'	...	2
'.....'	...	
'999'	...	999

Observaciones

1. El usuario deberá especificar una nueva tarjeta PRINT FORMATS para las variables que son transformadas a numéricas en el caso que ya hubiera especificado anteriormente (o si está recuperando los datos desde un archivo SPSS) para dichas vars un PRINT FORMATS como alfanuméricas.
2. Las palabras claves LOWEST, HIGHEST pueden ser utilizadas en forma abreviada como LO Y HI y sólo se aplican a variables numéricas. Por otro lado el ELSE se aplica tanto a variables numéricas como alfanuméricas.
3. La instrucción RECODE ocasiona la destrucción de la variable original. En el caso de que el usuario desee realizar una recodificación permanente pero desea además guardar los valores originales podrá usar el COMPUTE antes de recodificar como se muestra:

COMPUTE	EDADEC=EDAD
RECODE	EDADEC(0 THRU 10 = 1)(11 THRU 20 = 2) (21 THRU 30 = 3)(30 THRU HI = 4)

O sea en EDADEC tendrá la edad en grupos decenales hasta 30 y + y la variable original EDAD permanecerá inalterada.

VIII.1.1 Recodificación temporal: *RECODE

Todas las recodificaciones anteriormente ejemplificadas se trataban de RECODE permanentes y en el caso de utilizarse el SAVE FILE para la creación de archivo SPSS, los valores que se almacenan en el archivo serán los valores recodificados (recuérdese que la recodificación permanente destruye los valores originales).

En muchos casos, el usuario desea realizar una recodificación temporal. Esto se obtiene mediante *RECODE e implica que los valores originales no se destruyen y la recodificación temporal sólo tiene efecto para la tarea que sigue inmediatamente después.

Ejemplos:

1	16
RUN NAME	EJEMPLO DE RECODE
FILE NAME	NAPOL
VARIABLE LIST	EDAD, NIVEL, SUELDO
INPUT FORMAT	FIXED(10X, F2.0, F1, 0, T40, F6.1)
INPUT MEDIUM	DISK
N OF CASES	UNKNOWN
COMPUTE	EDADEC=EDAD
RECODE	EDADEC (0 THRU 9 = 1)(10 THRU 19 = 2) (20 THRU 29 = 3)(30 THRU 39 = 4) (40 THRU 49 = 5)(50 THRU 98 = 6)(99 = 7)
CROSSTABS	TABLES = EDADEC BY NIVEL
STATISTICS	ALL
READ INPUT DATA	
*RECODE	SUELDO (0 THRU 100 = 1)(101.1 THRU 500 = 2)(500.1 THRU 1000 = 3) (1000.1 THRU HIGHEST = 4)
FREQUENCIES	GENERAL = ALL
STATISTICS	ALL
SAVE FILE	
FINISH	

Este ejemplo ilustra el uso del RECODE permanente y *RECODE temporal: el primer RECODE recodifica la variable EDADEC en forma permanente y los valores de esta var en el archivo SPSS serán: 1,2,3, ... 7; el segundo *RECODE sólo tiene efecto para la tarea que le sigue: FRECUENCIAS y la variable SUELDO en el archivo SPSS mantendrá sus valores originales de 6 dígitos y 1 decimal.

VIII.1.3 Limitaciones

- (a) Las tres variables automáticas SEQNUM, SUBFILE y CASWGT no deben ser recodificadas.
- (b) No se permite en una instrucción RECODE o *RECODE más de 250 elementos. Los elementos del campo de especificación de RECODE son: nombres de variables, palabras claves, valores individuales, paréntesis izquierdo, paréntesis derecho, signo =

RECODE	X1 TO X10 (CONVERT)	posee 6 elementos que son "X1", "TO", "X10", "(", "CONVERT", ")".
--------	---------------------	---

RECODE	VARI (1, 2, 3 = 1) (ELSE = 2)	posee 13 elementos que son: "VARI"; "(", "1", "2", "3", "=", "1", ")", "(", "ELSE", "=", "2", ")".
--------	-------------------------------	--

- (c) Máximo de 400 variables pueden estar involucradas en el conjunto de todas las instrucciones RECODE; *RECODE; COUNT, *COUNT

VIII.2 La tarjeta COMPUTE

Formato General

1	16
COMPUTE	variable computada = expresión aritmética

Esta instrucción es también llamada transformación incondicional.

Veamos algunos ejemplos que ilustran el uso de esta instrucción.

- (a) Supongamos que en una encuesta se registró para cada persona 3 fuentes de ingreso: sueldo, ingreso por comisiones y otros ingresos; y se desea obtener el ingreso total que sería la suma de estas 3 vars.

COMPUTE INGTOT = SUELDO + COMIS + OTROS

(b) Otra ilustración del uso de la instrucción COMPUTE es: para cada persona disponemos de dos variables: estado civil y sexo; deseamos crear una 3ra. variable que sería la combinación de estas dos.

<u>Nombre variable</u>	<u>categorías</u>
ESTACIV	1 = soltero 2 = casado 3 = unido 4 = separado, viudo, divorciado 5 = otro
SEXO	1 = hombres 2 = mujeres

La instrucción:

COMPUTE ESTSEX= (SEXO-1)*5 + ESTACIV

produce la creación de una nueva variable ESTSEX con las siguientes categorías posibles.

1 = hombre, soltero	6 = mujer, soltera
2 = " , casado	7 = " , casada
3 = " , unido	8 = " , unida
4 = " , separado, etc.	9 = " , separada, etc.
5 = " , otros	10 = " , otros

Este tipo de combinación de variables permite la obtención de relaciones entre variables en una forma que facilite el análisis. Obviamente la realización de este tipo de transformación exige que conozcamos previamente las categorías existentes: si hubiera algún caso con sexo = 1 y estado civil = 6, esta persona aparecería como mujer soltera lo que obviamente es incorrecto.

La construcción de una expresión aritmética puede estar constituida por funciones y operadores aritméticos. Toda función posee una jerarquía superior a cualquier operador aritmético y por lo tanto se evalúa antes.

*** Funciones ***

<u>Símbolo</u>	<u>Función</u>	<u>Ejemplo</u>
SQRT	Raíz cuadrada	VX = SQRT (A)
LN	Logaritmo natural	VX = LN (VAR1)
LG10	Logaritmo base 10	VX = LG10 (VAR1)
EXP	Exponencial (e^{argum})	VX = EXP (VAR1)
SIN	Seno (*)	VX = SIN (VAR1)
COS	Coseno (*)	VX = COS (VAR1)
ATAN	Arco tangente (*)	VX = ATAN (VAR1)
RND	Redondeo al entero más próximo.	VX = RND (VAR1 + VZ)
ABS	Valor absoluto	VX = ABS (VAR1)
TRUNC	Trunca la parte decimal	VX = TRUNC (VZ)
MOD10	Da como resultado el resto de la división por 10.	VX = MOD10 (VAR1)

* Argumento en radianos

NOTA: los argumentos de cualquier función pueden ser una expresión aritmética.

*** Operadores aritméticos ***

<u>Símbolo</u>	<u>Operador</u>	<u>Ejemplo</u>	<u>Jerarquía</u>
**	Exponenciación	VX = VA**2	3 (más alta)
*	Multiplicación	VX = VA*4	2
/	División	VX = VA/VB	2
+	Suma	VX = VA + VB	1
-	Resta	VX = VA - VB	1 (más baja)

Se puede observar que el símbolo "/" en el contexto de la instrucción COMPUTE es el símbolo del operador división, al paso que en otros contextos es utilizado como delimitador especial. Nunca habrá confusión pues el campo de especificación del COMPUTE jamás utilizará el "/" como delimitador.

Las reglas para la evaluación de la expresión aritmética al lado derecho del operador de asignación "=" son las mismas del lenguaje FORTRAN:

- a) Calcula antes las operaciones de mayor jerarquía.
- b) Entre operaciones de misma jerarquía evalúa de izquierda \longrightarrow derecha

Ejemplos

(a) COMPUTE $M = A/B/C$
 Equivale a $M = \frac{\frac{A}{B}}{C}$ en forma matemática

pero $\frac{\frac{A}{B}}{C} = \frac{A}{B*C}$

(b) Si deseamos computar $M = \frac{A}{\frac{B}{C}} = \frac{A*C}{B}$

se escribiría el "compute" necesariamente con paréntesis pues éstos sirven para modificar la jerarquía o sea: se evalúa antes lo que está encerrado entre paréntesis:

COMPUTE $M = A/(B/C)$

(c) COMPUTE $M = A + B/C + C/D**2**A$

Podemos ilustrar las reglas de la evaluación mostrando como evoluciona el cálculo.

1. $M = A + B/C + C/ X**A$ $X = D**2$
2. $M = A + B/C + C/ Y$ $Y = X**A$
3. $M = A + Z + W$ $Z = B/C; W = C/Y$
4. $M = Z1 + W$ $Z1 = A + Z$

(d) El siguiente ejemplo ilustra cómo separar dígitos de una variable: una variable de 6 dígitos leída en formato F6.0 es la identificación geográfica de la unidad de estudio y sus dígitos según la posición nos dan Provincia, Cantón, Distrito, Zona.

Así el valor

1 6 5 2 3 2 ... 1 ... provincia San José
 65 ... cantón 65
 23 ... Distrito 23
 2 ... zona rural

Si la variable la llamamos de IDENT podemos obtener las demás mediante el siguiente juego de instrucciones

COMPUTE	PROVIN = TRUNC (IDENT/100000)
COMPUTE	CANTON = TRUNC (IDENT/1000) - (PROVIN*100)
COMPUTE	DISTRIT = TRUNC (IDENT/10) - (PROVIN*10000 + CANTON*100)
COMPUTE	ZONA = MOD10 (IDENT)

Primer COMPUTE produce $1.65232 - \text{TRUNC} \rightarrow 1 = \text{PROVIN}$

Segundo COMPUTE : $\text{TRUNC (IDENT/1000)} = 165$
 $\text{PROVIN} * 100 = 1 * 100 = 100$
luego $165 - 100 = 65 \rightarrow 65 = \text{CANTON}$

Tercer COMPUTE: $\text{TRUNC (IDENT/10)} = 16523$
 $\text{PROVIN} * 10000 + \text{CANTON} * 100 = 1 * 10000 + 65 * 100$
 $16523 - 16500 = 23 = \text{CANTON}$

Cuarto COMPUTE: MOD10 (IDENT) resto es 2 = ZONA

VIII.2.1 *COMPUTE: Cálculo temporal

En forma similar al *RECODE la instrucción *COMPUTE es temporal y su efecto sólo se extiende a la tarjeta de procesamiento (CROSSTABS, REGRESSION) que le sigue.

VIII.3 La tarjeta "IF": transformación condicional

Formato general

1	16
IF	(expresión lógica) var computada = expresión aritmética

Una "expresión lógica" consiste en una o más expresiones de relación

En forma general

expresión lógica = (expresión aritmética 1) operador de relación (expresión aritmética 2)

$$\text{expresión lógica} = (\text{expresión lógica 1}) \text{ operador lógico } (\text{expresión lógica 2})$$

Dos (2) expresiones aritméticas deben estar ligadas por uno y sólo un operador de relación.

Expresión lógica puede ser: Verdadera o Falsa

Llamemos EXP1 = expresión aritmética 1
EXP2 = expresión aritmética 2

TABLA DE OPERADORES DE RELACION

Relación	Símbolo del operador	Definición
\geq	GE	Si EXP1 \geq EXP2 la relación es verdadera, si no es falsa
\leq	LE	Si EXP1 \leq EXP2 - verdadera
$>$	GT	Si EXP1 $>$ EXP2 - verdadera
$<$	LT	Si EXP1 $<$ EXP2 - verdadera
$=$	EQ	Si EXP1 = EXP2 - verdadera
\neq	NE	Si EXP \neq EXP2 - verdadera

Ejemplo

$$\text{IF } (\text{VAR1 GE VAR2}) \text{ X} = \text{A*B}$$

Veamos para algunos casos el valor que se obtiene para X; convencionando V = verdadera F = Falsa

N° caso	Valor VAR1	Valor VAR2	Resultado expresión VAR1 ≥ VAR2	Valor A	Valor B	X
1°	10	9	V	2	1	2
2°	10	10	V	3	2	6
3°	10	11	F	3	4	0 (*)
4°	8	7	V	5	2	10
5°	8	8	V	6	3	18
6°	8	9	F	7	7	0 (*)

(*) Los casos que la expresión lógica es falsa, el sistema inicializa la variable a ser calculada en cero. Si deseamos asignar otro valor a X debemos usar la instrucción ASSIGN MISSING que se explica más adelante.

Operadores lógicos

AND: este operador enlaza dos expresiones lógicas dando resultado a una expresión lógica que será verdadera sólo si la expresión lógica 1 "y" la expresión lógica 2 son verdaderas. Basta que una de las expresiones sea falsa para que toda la expresión sea falsa.

OR: este operador enlaza dos expresiones lógica y la expresión resultante será verdadera si la expresión 1 es verdadera "o" la expresión 2 es verdadera. O sea, la expresión resultante sólo será falsa si ambas expresiones (1 y 2) son falsas.

NOT: al contrario de los 2 anteriores que se aplican entre 2 expresiones lógicas este operador se aplica a una expresión lógica y su efecto es invertir el resultado. O sea:

NOT (expresión lógica 1) = Verdadera
 si expresión lógica 1 es falsa,
 NOT (expresión lógica 1) = Falsa
 si expresión lógica 1 es verdadera.

Ejemplos:

(a)

```

1          16
IF          (SEXO EQ 1 AND EDAD GE 10) VX=1
IF          (SEXO EQ 2 AND EDAD GE 10) VX=2
    
```

En este ejemplo estamos creando una nueva variable "VX" que tendrá valores 0, 1, 2, según las características de cada caso (persona); estamos por lo tanto creando una variable que subdivide nuestro conjunto de datos según ciertas características y que son:

<u>Valor de VX</u>	<u>Características</u>
1	Hombres de 10 años y más (SEXO = 1 "y" EDAD \geq 10)
2	Mujeres de 10 años y más
0	Hombres y mujeres menores de 10 años

(b)

1	16
IF	(A GT B OR A GT C) D=1

En este caso la nueva variable "D" tendrá el valor 1 si $A > B$ o $A > C$; o sea cualquier de las dos relaciones que se cumpla nos produce una expresión verdadera y por lo tanto computará el valor de $D = 1$

(c)

1	16
IF	(A GT B AND A GT C) D=1

En este ejemplo sólo computará "D=1" si ambas relaciones son verdaderas o sea, si el valor de A es mayor que B pero además que dicho valor de A sea también mayor que C.

Veamos algunas situaciones:

<u>Valor A</u>	<u>Valor B</u>	<u>Valor C</u>	<u>Resultado expresión lógica</u>	<u>Valor D</u>	<u>Observaciones</u>
1	0	0	Verdadera	1	
1	1	0	Falsa	0	A = B
2	1	3	Falsa	0	A < C
3	2	0	Verdadera	1	
3	2	3	Falsa	0	A = C

(d)

```
1          16
IF          (SEXO EQ 2 AND (EDAD LT 10 OR
            EDAD GE 50)) VX=1
IF          (SEXO EQ 1) VX=2
```

La nueva var VX sólo asume el valor 1 para aquellos casos que son mujeres y tengan edad menor que 10 años o mayor que 49 años; asumirá el valor 2 para hombres. Por complemento VX=0 para el universo de mujeres con edad entre 10 y 49 años.

Este ejemplo ilustra la forma de una expresión lógica más compleja en que hay 3 relaciones y un cambio en la jerarquía de la evaluación por el uso de los paréntesis para encerrar "(EDAD LT 10 OR EDAD GE 50)"; la cual pasa como un todo a ser la expresión lógica 2 que se une mediante AND a la expresión lógica 1 que es "SEXO EQ 2".

(e) El siguiente ejemplo ilustra el uso de los paréntesis:

```
1          16
IF          (A EQ 9 OR (B GT 0 AND B LT 10))D=1
IF          (A EQ 9 OR B GT 0 AND B LT 10)D=1
```

La primera instrucción IF dejará D=1 cuando "A=9" o cuando "0<B<10".

El segundo IF como se puede observar es idéntico al primero a excepción de los paréntesis pero su evaluación es totalmente distinta. En este caso D=1 cuando "A=9 y B < 10" o entonces cuando "B > 0 y B < 10". Así el 2º IF sería equivalente a:

```
IF          ((A EQ 9 OR B GT 0) AND B LT 10)D=1
```

(f) El uso del operador NOT:

```
IF          (EDAD GE 10) X=1
```

La variable X asumirá el valor 1 siempre que el valor de la variable EDAD sea mayor o igual (\geq) a 10. Si a esta expresión lógica aplicamos el operador NOT tendríamos

```
IF          (NOT (EDAD GE 10)) X=1
```

cuyo efecto es invertir el resultado de la expresión lógica a la cual se aplica. Así "X" asumirá el valor 1 siempre que la EDAD sea menor que ($<$) 10 que es la negación de $EDAD \geq 10$

Este ejemplo resulta sencillo visualizar la negación, pero cuando el NOT se aplica sobre una expresión lógica compleja, el usuario debe preocuparse de estar completamente seguro sobre el resultado que va a obtener, motivo por el cual es recomendable bastante cuidado en el uso de este operador.

VIII.3.1 Operandos y operadores de relación implícitos del IF.

En muchos casos las expresiones lógicas resultan ser bastante extensas en su redacción, como sería el siguiente ejemplo:

```
1          16
IF          (A EQ 1 OR A EQ 3 OR A EQ 5 OR A EQ 7) X=1
```

El SPSS permite una notación implícita que torna menos aburrida la redacción de las expresiones lógicas. El ejemplo anterior lo podríamos escribir:

```
IF          (A EQ 1 OR 3 OR 5 OR 7) X=1
```

La siguiente serie de ejemplos ilustran el uso de la notación implícita; la primera instrucción muestra la notación normal y la 2° la notación equivalente implícita.

(a)

```
IF          (A GT 1 OR A EQ 0) X=1
IF          (A GT 1 OR EQ 0) X=1
```

(b)

```
IF          (X EQ 1 OR X EQ 3 OR X EQ 5) Y=1
IF          (X EQ 1 OR 3 OR 5) Y=1
```

(c)

```
IF          (B EQ 1 OR B EQ 3 OR B GT 10) Y=1
IF          (B EQ 1 OR 3 OR GT 10) Y=1
```

(d)

```
IF          (A EQ B AND A EQ C OR A EQ D) X=1
IF          (A EQ B AND C OR D) X=1
```

Esta expresión será verdadera si "A=B y A=C" o cuando "A=D".

(e)

```
IF      (A EQ B OR (A GE C AND A LE D)) X=1
IF      (A EQ B OR (GE C AND LE D)) X=1
```

La expresión será verdadera si "A=B" o " $C \leq A \leq D$ ".

(f)

```
IF      (A EQ Z AND B EQ Z) Y=9
IF      (Z EQ A AND B) Y=9
```

Como la notación implícita no permite la omisión del 2º operando de una relación, este caso requirió que se cambiara el orden dejando Z como el primer operando.

Reglas de la Notación implícita del IF

- 1) Los operadores lógicos AND, OR no pueden nunca ser omitidos.
- 2) El segundo operando de una relación nunca puede ser omitido (ejemplo f).
- 3) La omisión del primer operando de las relaciones tiene efecto a través de los paréntesis (ejemplo e).

VIII.3.2 El IF temporal: *IF

Igual que *RECODE, *COMPUTE la instrucción con la palabra de control "*IF" realiza creación condicional de variables en forma temporal que se aplican sólo al procedimiento que les sigue en forma inmediata.

Ejemplo

```
1          16
*COMPUTE   A = X + Y*2
*COMPUTE   B = X1
*RECODE    A ...
*IF        (B EQ A) C = 1
PARTIAL CORR .....
REGRESSION
```

Las variables A, B, C son creadas en carácter temporal y sólo las podemos utilizar en el procedimiento PARTIAL CORR que les sigue en forma inmediata.

Si intentamos utilizar las variables A, B, C en el procedimiento REGRESSION, se produciría un error pues a ese momento dichas variables ya no existen.

VIII.4 Limitaciones de COMPUTE e IF

- 1) Las tres variables automáticas SEQNUM, SUBFILE, CASWGT no deben ser nunca transformadas, o sea, no deben aparecer a la izquierda del signo [=] de COMPUTE, *COMPUTE, IF, *IF.
- 2) Se permite un máximo de 250 elementos. Los elementos son: nombre de variable, palabra clave, paréntesis izquierdo, paréntesis derecho, operador lógico, operador de relación, constante numérica y signo igual.

COMPUTE VARX = VAR1 + VARB*2
contiene 7 elementos

- 3) Se permite un máximo de 4 tarjetas de continuación.

VIII.5 La tarjeta "COUNT": creación de índices aditivos

Esta instrucción permite al investigador crear índices aditivos que le permita en forma más cómoda, analizar el comportamiento del universo en estudio en base a ciertas características conjuntamente.

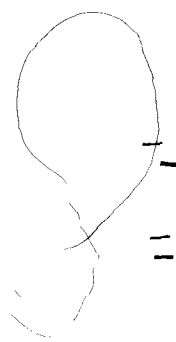
Por ejemplo, en una encuesta de opinión se preguntó:

(i)	Escucha la radio A	<input type="checkbox"/>	SI	<input type="checkbox"/>	NO
(ii)	Escucha la radio B	<input type="checkbox"/>	SI	<input type="checkbox"/>	NO
(iii)	Escucha la radio C	<input type="checkbox"/>	SI	<input type="checkbox"/>	NO

Suponiendo que la muestra utilizada contempló 100 personas; el investigador desea saber:

- (a) Cuántas no escuchan ninguna radio.
- (b) Cuántas escuchan por lo menos 1 radio.
- (c) Cuántas escuchan por lo menos 2 radios.
- (d) Cuántas escuchan las tres radios.

La creación de una nueva variable que indique estas posibilidades se puede perfectamente realizar por medio de una serie de instrucciones RECODE, COMPUTE, IF, lo que constituye en ciertos casos una tarea bastante tediosa de redactar.



Por medio de la instrucción COUNT la creación de esta variable, llamando la RADIO sería

```
1          16
COUNT    RADIO = RADA, RADB, RADC (1)
```

que es equivalente a la secuencia:

```
1          16
COMPUTE  RADIO = 0
IF       (RADA EQ 1) RADIO = RADIO + 1
IF       (RADB EQ 1) RADIO = RADIO + 1
IF       (RADC EQ 1) RADIO = RADIO + 1
```

La obtención de la distribución de frecuencias de la variable RADIO sería en forma hipotética:

<u>Nombre de categoría</u>	<u>código</u>	<u>Frecuencia</u>
No escucha ninguna.....	0	20
Escucha 1.....	1	30
Escucha 2.....	2	30
Escucha las 3.....	3	20

Formato General

```
1          16
COUNT    variable1 = lista11 de variables de decisión
           (lista11 valores) (lista12 de vars de decisión)
           (lista12 valores)...] / ...
           variablen = listan1 vars decisión (listan1 valores)
           [listan2 de vars decisión (listan2 valores)] .
```

Los siguientes ejemplos ilustran las diversas formas de especificación de la lista de valores.

(a) Variables numéricas

```
1          16
COUNT    Z = A,B,C (1)
COUNT    Z = A,B,C (1,2,7)
COUNT    Z = A,B,C (BLANK,1,9) (*)
COUNT    Z = A,B,C (0,1,9) (*)
COUNT    Z = A,B,C (LOWEST THRU 3,4,7)
COUNT    Z = A,B,C (LO THRU 3,4,7) (**)
COUNT    Z = A,B,C (1,2,5 THRU 10.99)
COUNT    Z = A,B,C (1,9,10 THRU HIGHEST)
COUNT    Z = A,B,C (1,9,10 THRU HI) (**)
```

(*) En general el SPSS trata el blanco para variables numéricas como si fuera cero.
En forma similar a RECODE, la instrucción COUNT acepta la palabra clave BLANK para diferenciar el blanco del cero en variables numéricas.
Así la lista (BLANK,1,9) incrementa la variable Z en una unidad para cada variable de decisión, si detecta valores: blanco, 1 o 9. Por otro lado la lista de valores (0,1,9) realiza el incremento si detecta los valores: blanco y cero, 1,9 -en este caso el cero incluye el blanco.

(**) LO = abreviación de LOWEST
HI = abreviación de HIGHEST

b) Variables alfanuméricas

```
COUNT          Z = A,B,C ('1','2','9')
```

O sea los valores deben ser especificados entre apóstrofes (comillas simples).

Las palabras claves BLANK, HIGHEST, LOWEST no pueden ser utilizadas para variables alfanuméricas.

El siguiente ejemplo ilustra el uso de COUNT con más de un valor en la lista de valores.

```
COUNT          Z = A (1,2), B,C (3,7)
```

Sería equivalente a la siguiente serie de instrucciones

```
COMPUTE        Z = 0  
IF              (A EQ 1 OR A EQ 2) Z = Z+1  
IF              (B EQ 3 OR 7) Z=Z+1  
IF              (C EQ 3 OR 7) Z=Z+1
```

Si por equivocación se hubiera especificado dos veces el mismo valor en la lista de valores, esto no implica que se incremente dos veces la variable que se crea. Así, la instrucción:

```
COUNT          Z = A(1,2,1),B,C(3,7,3)
```

es exactamente equivalente al ejemplo antecedente.

Por otro lado, si se nombra una variable más de una vez en la lista de variables de decisión, ocasiona que se incremente la variable que se está creando tantas veces cuantas se nombró la misma variable. Esto permite al investigador crear su índice aditivo ponderando las variables de decisión. El siguiente ejemplo ilustra este aspecto de la instrucción COUNT.

```
COUNT      Z = A1, A2, A3, A1, A2 (1)
```

La secuencia de instrucciones equivalente es:

```
COMPUTE    Z = 0
IF          (A1 EQ 1) Z = Z + 1
IF          (A2 EQ 1) Z = Z + 1
IF          (A3 EQ 1) Z = Z + 1
IF          (A1 EQ 1) Z = Z + 1
IF          (A2 EQ 1) Z = Z + 1
```

Para ilustrar el uso del COUNT ponderado supongamos que en un examen de selección se utilizó el sistema de test: de un total de 10 preguntas, 5 son de conocimientos generales con sólo una alternativa correcta y 5 son de físico-matemática con 1 alternativa correcta y otra semi-correcta. Para la evaluación se adoptó el siguiente criterio.

- (a) conocimientos generales, correcta, vale 1 punto.
- (b) Físico-matemática, correcta, vale 2 puntos.
- (c) Físico-matemática, semi-correcta, vale 1 punto.

Por lo tanto, la NOTA de evaluación final podrá variar desde 0 hasta un máximo de 15 puntos.

La tabla siguiente provee información sobre las variables:

<u>Nombre</u>	<u>Códigos válidos</u>	<u>Código de respuesta correcta</u>	<u>Código de respuesta semi-correcta</u>
CONG1	1, 2, 3, 4, 5	3	-
CONG2	1, 2, 3, 4, 5, 6	3	-
CONG3	1, 2, 3	3	-
CONG4	1, 2, 3, 4	2	-
CONG5	1, 2, 3, 4	2	-
FMAT1	1, 2, 3, 4, 5	1	3
FMAT2	1, 2, 3, 4, 5	1	3
FMAT3	1, 2, 3	1	3
FMAT4	1, 2, 3, 4, 5	2	4
FMAT5	1, 2, 3, 4, 5	2	4

La instrucción:

1	16
COUNT	NOTA=CONG1 TO CONG3(3), CONG4, CONG5(2), FMAT1 TO FMAT3, FMAT1 TO FMAT3(1), FMAT4, FMAT5, FMAT4, FMAT5(2), FMAT1 TO FMAT3(3), FMAT4, FMAT5(4)

genera la variable NOTA con un rango de 0-15, ponderando las respuestas correctas de físico-matemáticas con 2 puntos; las semi-correctas con un punto y las de conocimientos generales con un punto.

Suponiendo además que a dicho examen se sometieron 1000 personas; la distribución de frecuencias de la variable NOTA indicaría al investigador, cuántas personas tuvieron nota cero, cuántas nota 1, etc.

*COUNT: Instrucción temporal

En forma análoga a *RECODE, *IF, *COMPUTE, tenemos la instrucción *COUNT que sólo afecta al procedimiento inmediatamente posterior.

Ejemplo:

1	16
*COUNT	Z = A,B,C(1,2)/X = V1,V2,V5(3)

VIII.5.1 Limitación de COUNT

1. Las variables automáticas SEQNUM, SUBFILE, CASWGT, no pueden aparecer a la izquierda del signo "=" de una instrucción COUNT o *COUNT
2. Un máximo de 250 elementos puede aparecer en una instrucción COUNT o *COUNT

La siguiente instrucción posee 10 elementos.

1	16
COUNT	Z = A1 TO A10 (1,2,3)

Nótese que la lista de variables de decisión A1 TO A10, cuenta como tres elementos.

3. Un máximo de 400 variables pueden ser nombradas (implícita o explícitamente) en el conjunto de todas las declaraciones RECODE, *RECODE, COUNT, *COUNT.
4. Las limitaciones de memoria para el conjunto de las instrucciones de transformación de datos se encuentran en los párrafos siguientes.

VIII.6 La tarjeta "ASSIGN MISSING"

En general las variables creadas por medio de COMPUTE, IF, COUNT, utilizan variables originales o anteriormente creadas. Muchas veces las variables involucradas poseen un valor declarado como MISSING y a menos que el usuario realice su propio control sobre el uso o no de los casos con valores faltantes, sencillamente las instrucciones COMPUTE, IF, COUNT utilizan dichos valores en el cómputo de la variable que se crea; lo cual puede motivar la obtención de resultados impredecibles.

Para evitar la tarea tediosa del usuario de tener que realizar los controles sobre estos casos el SPSS dispone de la instrucción ASSIGN MISSING cuyo formato general es:

```
1          16
ASSIGN MISSING { nombre de var } (valor faltante) / ....
                o
                { lista de var }

                { nombre de var } (valor faltante)
                o
                { lista de vars }
```

La instrucción ASSIGN MISSING, colocada posteriormente en las instrucciones de transformación de datos, asignará a la variable computada el valor faltante especificado entre paréntesis siempre y cuando algunas de las variables involucradas en el cálculo posea un valor declarado anteriormente en la instrucción MISSING VALUES o que traiga un valor faltante que fue asignado por el ASSIGN MISSING.

Consideremos un juego de datos donde disponemos registrado el sueldo del jefe, otro ingresos del jefe, sueldo del cónyuge y otros ingresos del cónyuge y el nombre de las variables son respectivamente SUELDO, OTROING, SCONY, OTCONY y deseamos obtener el ingreso total de la pareja y el promedio de ingreso por persona.

```
1          16
MISSING VALUES SUELDO,OTROING,SCONY,OTCONY(0,9999)
...
COMPUTE      INGTOT=SUELDO+OTROING+SCONY+OTCONY
COMPUTE      PROMED=INGTOT/2
ASSIGN MISSING INGTOT,PROMED (9999)
```

Para el primer COMPUTE a la variable INGTOT se asignará el valor 9999 si cualquiera de las cuatro variables trae un valor declarado como missing.

Para el 2° COMPUTE la variable PROMED asume el valor 9999 siempre que la variable INGTOT, computada anteriormente posea el valor 9999 producido por el ASSIGN MISSING o sea si INGTOT trae el valor faltante asignado. Esta es una importante característica de acción acumulativa del ASSIGN MISSING.

Para declaraciones IF, el valor faltante será asignado cuando:

- a) cualquier variable involucrada en la expresión lógica trae un valor declarado como missing, o
- b) a pesar de ser verdadera la expresión lógica, alguna de las variables involucradas en el cómputo de la variable que se crea (derecha del =) trae un valor missing.

Ejemplos

(a)

```
MISSING VALUES ESTCIV(9),EDAD(99)
...
IF          (EDAD GE 50) NUEV = (NIVEL-1)*5+ESTCIV
ASSIGN MISSING NUEV (99)
```

La variable NUEV tendrá asignado el valor missing 99 cuando:

- a) La variable EDAD es missing.
- b) A pesar de EDAD ser ≥ 50 y con un valor no missing (EDAD=60 por ejemplo) alguna de las variables NIVEL o ESTCIV es missing.
- c) Suponiendo la variable NUEV, la estamos creando en esta declaración IF, asumirá el valor 99 para todos los casos en que EDAD es < 50.

En el caso de que la variable NUEV estuviese creada en alguna declaración anterior al IF o ya existiera en el archivo su valor permanecería inalterado para todos los casos en que EDAD < 50.

- (b) En el caso que estamos creando una nueva variable por medio de varias declaraciones IF, el ASSIGN MISSING actúa sobre cada declaración en forma independiente.

Veamos un ejemplo:

```
MISSING VALUES VAR050,VAR071,VAR063(0,99),TIPO(9)
IF          (TIPO EQ 1) EDAD = VAR050
IF          (TIPO EQ 2) EDAD = VAR071
IF          (TIPO EQ 3) EDAD = VAR063
ASSIGN MISSING EDAD(99)
```

Así para el primer IF, en el caso de que TIPO=1, la asignación del valor missing 99 se dará sólo en el caso que VAR050 sea missing. Igual raciocinio se aplica a las otras dos declaraciones IF, siempre que sea verdadera la expresión lógica. En los casos que la variable TIPO sea 0, 4, 5, 6, 7, 8, 9 (o sea ninguna de las expresiones lógicas se cumplen), también el ASSIGN MISSING actuará asignando a la variable edad el valor 99 pues esta variable está siendo creada en este conjunto de declaraciones -o sea el ASSIGN MISSING es también utilizado para la inicialización de variables (sustituye el antiguo TMISS).

- (c) Otro caso a analizar es cuando la variable que se está transformando en forma condicional ya existe en el archivo o fue creada en instrucciones anteriores.

supongamos una variable "ZX" ya existente en el archivo:

```
1          16
MISSING VALUES ZX(9,10,11),X1,X2(8,9)
...
...
IF          (TIPO EQ 1) ZX = (X1 -1)*5+X2
ASSIGN MISSING ZX(99)
```

La secuencia de instrucciones arriba causan una modificación de la variable ZX sólo cuando se cumple la expresión lógica TIPO = 1. En los casos en que TIPO \neq 1 el valor de ZX permanece inalterado.

El ASSIGN MISSING ocasiona que la declaración anterior de MISSING para ZX se anule y por lo tanto sólo se considera el valor "99" como el valor faltante de ZX; o sea los casos en que ZX asume los valores 9,10,11 y que TIPO \neq 1 serán tratados como valores no missing. El ASSIGN MISSING en este caso tendrá efecto para los casos en que X1 y/o X2 sean missing.

Si deseamos que los valores 9, 10, 11 sean tratados como missing debemos recodificarlos al valor 99 como se muestra:

```
1          16
MISSING VALUES ZX(9,10,11),X1,X2(8,9)
...
...
RECODE      ZX(9,10,11 = 99)
IF          (TIPO EQ 1) ZX = (X1 -1)*5 + X2
ASSIGN MISSING ZX(99)
```

Observación: Si posteriormente a la secuencia de instrucciones que se muestra arriba hubieramos especificado:

```
MISSING VALUES ZX(99)
```

el efecto sería:

1. Estamos declarando para ZX un único valor faltante: el "99"
2. El ASSIGN MISSING pierde el efecto y no se realizará ninguna asignación de valores faltantes.

NOTAS SOBRE EL USO DEL ASSIGN MISSING

1. El ASSIGN MISSING sólo tiene efecto si la relación lógica no se cumple y para variables que están siendo creadas o cuando las variables involucradas en el cálculo asumen un valor declarado como missing.

2. El ASSIGN MISSING debe ser puesto a continuación del juego de instrucciones IF, COMPUTE, COUNT sobre las cuales deseamos su efecto. Para el caso de *IF, *COMPUTE, *COUNT el ASSIGN MISSING debe ser colocado a continuación del juego de instrucciones temporales y antes de la tarjeta de procedimiento.
3. No se debe nombrar una variable con asignación de valores faltantes mediante el ASSIGN MISSING en una instrucción posterior de MISSING VALUES pues esto ocasiona la anulación del efecto del ASSIGN MISSING.
4. La declaración ASSIGN MISSING sólo se aplica a variables computadas mediante IF, COMPUTE, COUNT. Para variables transformadas mediante la declaración RECODE debe usarse la declaración MISSING VALUES.

VIII.7 Inicialización de variables creadas con IF

Quando creamos una variable nueva en una declaración IF, podemos indagar: qué valor asume la variable si la condición lógica no se cumple? o entonces al obtener los resultados utilizando dicha variable observamos que hay una alta frecuencia del valor "cero" y por qué aparece dicho valor si la expresión aritmética después del IF nunca puede dar como resultado el valor cero?

Veamos el siguiente ejemplo:

```
IF          (EDAD GE 30) NUEVA = VA*VB+10
```

En el cual obviamente la variable NUEVA nunca podrá asumir el valor cero si se cumple $EDAD \geq 30$. El sistema no en tanto inicializa la var NUEVA con el valor cero y éste será el valor de la variable para los casos en que EDAD es menor que 30.

En otras palabras el valor cero es el valor de inicialización estándar del sistema. Por diversas razones podemos desear que dicho valor sea distinto de cero, principalmente en los casos en que el resultado de la expresión aritmética puede ser cero. La inicialización se realiza mediante la instrucción ASSIGN MISSING:

```
IF          (EDAD GE 30) NUEVA = VA*VB + 10  
ASSIGN MISSING NUEVA (999)
```

Nótese que la inicialización sólo se realiza si la variable está siendo creada en la declaración IF y no tendrá efecto si la variable a ser computada ya es una variable existente y le estamos cambiando el valor.

VIII.8 Declaraciones: DO REPEAT; END REPEAT

Estas declaraciones permiten al usuario evitar la reducción tediosa de una secuencia de transformaciones iguales que se repiten con distintas variables.

Formato General

```
1          16
DO REPEAT  nombre de lista 1 = lista1 de variables/
           [nombre de lista 2 = lista2 de vars/]...
```

Nombre de lista: sigue las mismas reglas de la elaboración de nombre de variable (1 a 8 caracteres con el 1º alfabético); y su función en la declaración es hacer la equivalencia a una lista de variables.

Lista de variables: son una lista de variables que serán creadas y/o ya existente.

La mejor forma de ilustrar esta declaración es por medio de algunos ejemplos.

Ejemplo a):

```
1          16
VARIABLE LIST A1 TO A10, B1 TO B10
...
...
DO REPEAT  NUEVA = NUEVA1 TO NUEVA10/
           LISTA = A1 TO A10/LISTB = B1 TO B10/
COMPUTE    NUEVA = LISTA + LISTB*2
END REPEAT
```

Las 3 declaraciones DO REPEAT, COMPUTE; END REPEAT motivan que el sistema genere una serie de instrucciones que son las siguientes:

```
1          16
COMPUTE    NUEV1 = A1 + B1*2
COMPUTE    NUEV2 = A2 + B2*2
COMPUTE    NUEV3 = A3 + B3*2
...
...
COMPUTE    NUEV10 = A10 + B10*2
```

O sea en vez de tener que especificar 10 declaraciones COMPUTE el usuario con las tres declaraciones, usando DO REPEAT realiza las mismas transformaciones y de una manera más cómoda.

Ejemplo b):

```
1          16
DO REPEAT  XLIST1=A,B,C,D/LIST2 = Z1,Z2,Z3,Z4/
           LIST3 = Y1,Y2,Y3,Y4/XLIST4 = V1,V2,V3,V4/
COMPUTE    XLIST1 = 0
IF         (LIST2 GE LIST3) XLIST1 = LIST2 + 10
IF         (LIST2 GE 99)    XLIST1 = 99
COMPUTE    XLIST4 = LIST1 + LIST3
END REPEAT
```

El juego de instrucciones generado y equivalente sería:

```
COMPUTE    A = 0
IF         (Z1 GE Y1)  A = Z1 + 10
IF         (Z1 GE 99)  A = 99
COMPUTE    V1 = A + Y1

COMPUTE    B = 0
IF         (Z2 GE Y2)  B = Z2 + 10
IF         (Z2 GE 99)  B = 99
COMPUTE    V2 = B + Y2

COMPUTE    C = 0
IF         (Z3 GE Y3)  C = Z3 + 10
IF         (Z3 GE 99)  C = 99
COMPUTE    V3 = C + Y3

COMPUTE    D = 0
IF         (Z4 GE Y4)  D = Z4 + 10
IF         (Z4 GE 99)  D = 99
COMPUTE    V4 = D + Y4
```

Nótese que sólo para efecto de diferenciación se nombró XLIST1 y XLIST4 a las variables que se están creando y sin "X" adelante: LIST2 y LIST3 a las variables que sirven como base para los cálculos, o sea son variables ya existentes.

Normas para el uso del DO REPEAT

1. Declaraciones válidas entre la declaración DO REPEAT y END REPEAT.

COMPUTE	*COMPUTE
RECODE	*RECODE
IF	*IF
COUNT	*COUNT
SELECT IF	*SELECT IF
MISSING VALUES	
ASSIGN MISSING	

2. No existen declaraciones *DO REPEAT, *END REPEAT que serían temporales; pero el DO REPEAT se puede utilizar con las declaraciones *IF, *COMPUTE, etc. y su efecto será sólo sobre el procedimiento (tarea) que le sea subsiguiente.
3. Declaraciones DO REPEAT con transformaciones permanentes (IF, COMPUTE, etc. -sin asterisco-) deben aparecer antes de la primera tarjeta de procedimiento (1° tarea).

VIII.8.1 Limitaciones para el DO REPEAT

1. No se permite el DO anidado: o sea en la forma siguiente:

```
DO REPEAT
...
DO REPEAT
...
END REPEAT
END REPEAT
```

2. El número máximo de listas es de 50. El ejemplo b) posee 4 listas.
3. El número máximo de declaraciones dentro del ciclo; sin contar DO REPEAT y a END REPEAT; es de 50. En el ejemplo b) tenemos 4 declaraciones: 2 COMPUTE, 2 IF.

4. Las limitaciones de las declaraciones simples RECODE, COMPUTE, IF, COUNT, SELECT IF, MISSING VALUES y ASSIGN MISSING se aplican al DO REPEAT. Así, el ejemplo b) los 2 COMPUTE dentro del ciclo general 8 COMPUTE pues tenemos las listas de 4 variables y por lo tanto se debe contabilizar como 8 COMPUTE y no solamente 2.
5. El SPSS requiere espacio adicional de memoria para interpretar y generar el juego de instrucciones correspondiente.

La fórmula del espacio requerido es:

$$\text{WORKSPACE} = 8 * (\text{NS} + \text{NS} * \text{L} + \text{E}_1 + \text{E}_2 + \dots + \text{E}_i)$$

donde:

NS = n° de listas de variables declaradas

L = largo de las listas; o sea el número de variables de las listas.

E_i = n° de elementos en las i-ésimas declaraciones
Los delimitadores comunes: blanco y coma no se cuenta como elemento.

Si más de un DO REPEAT es utilizado para un proceso se debe tomar en cuenta el WORKSPACE de aquel DO que requiera mayor espacio y no la sumatoria de los WORKSPACE referente a cada DO.

VIII.9 La tarjeta "ALLOCATE": Asignación de memoria

Las declaraciones de modificación de datos son leídas, decodificadas y almacenadas hasta que el SPSS detecte una tarjeta de un procedimiento estadístico (una tarea).

Cada caso que es leído pasa por la secuencia de instrucciones de modificación de datos antes de ser entregado al procedimiento. Obviamente la secuencia de instrucciones requiere un espacio de memoria donde quedará almacenada para aplicarse sobre cada caso que se procese.

El SPSS utiliza el espacio de memoria en dos funciones:

- (a) **TRANSPACE**: para almacenar la secuencia de instrucciones de modificación de datos.
- (b) **WORKSPACE**: para crear, manipular y almacenar las tablas y matrices requeridas por los procedimientos estadísticos.

De esta forma el espacio total de memoria disponible:

$$\text{SPACE} = \text{TRANSPACE} + \text{WORKSPACE}$$

En forma estándar por omisión el SPSS asume para **TRANSPACE** el valor $\text{SPACE}/8$ y el sobrante para **WORKSPACE**.

Si disponemos de $\text{SPACE} = 24\ 000$ bytes, el SPSS asumirá $\text{TRANSPACE} = 3\ 000$ bytes y $\text{WORKSPACE} = 21\ 000$ bytes.

El parámetro **SPACE** es dependiente de la memoria disponible en cada equipo en particular, considerando que el SPSS requiere para las múltiples funciones un total aproximado de 140K bytes. Considerando la memoria requerida por el sistema operativo, el restante nos daría el valor del parámetro **SPACE**.

La proporción $1/8$ para **TRANSPACE** y $7/8$ para **WORKSPACE** puede ser modificada por el usuario por medio de la declaración **ALLOCATE**; la cual debe preceder la primera declaración de modificación de datos.

El cálculo de TRANSPACE

El cálculo del parámetro **TRANSPACE** se realiza por medio de 3 fórmulas independientes y se debe tomar el valor más alto.

1. $\text{TRANSPACE} = \text{NTRANS} * 100$

Donde

NTRANS = número de declaraciones (permanentes y temporales) **COMPUTE**, **RECODE** y **COUNT** sumado a 2 veces el número de declaraciones **IF** y **SELECT IF** que aparecen antes de una tarjeta de procedimiento.

2. $\text{TRANSPACE} = \text{NVALS} * 25$

NVALS = número de valores simples o rangos que aparecen en todas las declaraciones **RECODE** y **COUNT**. Se debe incluir en la cuenta las palabras claves **CONVERT**, **HIGHEST**, **LOWEST**, **BLANK**, **ELSE**. En las declaraciones **RECODE** sólo se necesita contabilizar los valores y rangos a la izquierda del signo "=" (igual).

Así en un RECODE (10 = 1) se contabiliza uno para el cálculo de NVALS; (10, 11, 12 = 2) se contabiliza 3; (10 THRU 20 = 1) se contabiliza 1; (10 THRU 20, 30 THRU 40 = 3) se contabiliza 2.

Ejemplos:

```
RECODE      (1,3 THRU 10,12 = 1) (2,11,13 THRU 20,22 = 2)
            (ELSE = 3)
```

La declaración anterior incrementa NVALS en 8

La siguiente declaración COUNT aportaría 6 al cálculo de NVALS.

```
COUNT      NUEVA = VAR001 (LOWEST THRU 10) VAR010 TO VAR020
            (1,2,3) VAR051 ('1','9')
```

3. TRANSPACE = NOPERS * 12.5

NOPERS = número total de operadores lógicos, aritméticos y de relación sumados al número total de referencias a funciones que aparecen en el conjunto de todas las declaraciones COMPUTE, IF, SELECT IF.

- Operadores aritméticos [******, *****, **/**, **+**, **-**] ⇒ cada ocurrencia incrementa UNO al valor de NOPERS.
- Funciones {**SQRT**, **LN**, **LG10**, **EXP**, **SIN**, **COS**, **ATAN**, **AND**, **ABS**, **TRUNC**, **MOD10**} - cada ocurrencia incrementa UNO A NOPERS
- Operadores de relación [**GE**, **GT**, **LE**, **LT**, **NE**, **EQ**] - cada ocurrencia incrementa en UNO A NOPERS (NOTA: los operadores implícitos deben ser contabilizados).
- Operadores lógicos (**AND**, **OR**, **NOT**) - cada ocurrencia incrementa UNO a NOPERS.

```
COMPUTE      A = SQRT (X + Y) + (B * C) ** 2
IF           (A EQ 0 OR 1) X = A + B = C + D
```

El COMPUTE aporta 5 al cálculo de NOPERS. El IF aporta 6 al cálculo de NOPERS -nótese que se debe contar el operador implícito "OR A EQ 1".

Observación:

El usuario deberá observar las limitaciones anteriormente descritas.

- Máximo de 250 elementos por declaración.
- Máximo de 400 variables pueden ser nombradas implícita o explícitamente en el conjunto de todas las declaraciones RECODE, *RECODE, COUNT y *COUNT.

EJEMPLO DE CALCULO DE TRANSPACE

		<u>NTRANS</u>	<u>NVALS</u>	<u>NOPERS</u>
1	16			
COMPUTE	NUEVA=(VAR1 + 20*VAR10)*SQRT(VAR20)	1	-	4
COMPUTE	NUEX=((VA + VB**2)*VC + TRUNC(VAR1)/10) + VAR2**5	1	-	8
IF	((VAR1 GE 10 AND LE 50) OR (VAR2 GE 20)) X = A + B + C	2	-	7
RECODE	INGRES (LOWEST THRU 100 = 1)(100.1 THRU 200 = 2)(200.1 THRU 1000 = 3)(ELSE = 4)	1	4	-
RECODE	CODX (BLANK = 7)(0,1 = 1)(2 THRU 10 = 2)(11,15 THRU 20 = 3)(12,13,14 = 4)(21 THRU 30,41 THRU 50 = 5)(ELSE = 6)	1	12	-
COUNT	INDIC1 = RAD1,RAD2 (1,2), RAD5(3)/ INDIC2 = RAD1 TO RAD5 (2,3)	1	5	-
		<hr/>	<hr/>	<hr/>
		7	21	19

Por lo tanto

NTRANS = 7

NVALS = 21

NOPERS = 19

Utilizando las fórmulas

1. TRANSPACE = NTRANS*100 = 7*100 = 700

2. TRANSPACE = NVALS*25 = 21*25 = 525

3. TRANSPACE = NOPERS*12.5 = 19*12.5 = 237.5 (debe ser redondeado a 238).

Se puede observar que el valor calculado por medio de NTRANS (700) es el más alto y por lo tanto sólo éste debe ser tomado en cuenta. Para dejar un margen de error el usuario puede especificar un valor más alto que 700 como por ejemplo 800 a 950 para TRANSPACE.

El cálculo de TRANSPACE debe tomar en consideración las declaraciones de modificación de datos permanentes y temporales (sin asterisco; con asterisco: COMPUTE, *COMPUTE, etc.). O sea - para cada procedimiento los valores de NVALS, NTRANS y NOPERS cambiaron según el número de declaraciones temporales ya que las permanentes afectan a todos los procedimientos.

El siguiente ejemplo ilustra este aspecto del cálculo; suponiendo que las transformaciones permanentes son las del ejemplo anterior o sea que aportan NTRANS = 7, NVALS = 21 y NOPERS = 19.

	1	16	NTRANS	NVALS	NOPERS
COMPUTE		NUEVA = ...	1	-	4
COMPUTE		NUEX = ...	1	-	8
IF		2	-	7
RECODE		INGRES ...	1	4	-
RECODE		CODX ...	1	12	-
COUNT		INDIC1	1	5	-
*COMPUTE		ZTEMP = VAR1 + VAR2 + VAR3	1	-	2
*RECODE		VAR50 (BLANK = 99)	1	1	
FREQUENCIES				
READ INPUT DATA					
*COMPUTE		TEMP2 = VAR1*VAR2 + VAR3	1	-	2
PEARSON CORR		TEMP2, VAR50, VAR1, VAR10			

Para el primer procedimiento FREQUENCIES tenemos:

$$NTRANS = 7 + 2 = 9 \quad \text{---} \quad TRANSPACE = \underline{900}$$

$$NVALS = 21 + 1 = 22 \quad \text{---} \quad TRANSPACE = 550$$

$$NOPERS = 19 + 2 = 21 \quad \text{---} \quad TRANSPACE = 262.5 = 263$$

Para el segundo procedimiento PEARSON CORR tenemos:

$$NTRANS = 7 + 1 = 8 \quad \text{---} \quad TRANSPACE = \underline{800}$$

$$NVALS = 21 + 0 = 21 \quad \text{---} \quad TRANSPACE = 525$$

$$NOPERS = 19 + 1 = 20 \quad \text{---} \quad TRANSPACE = 250$$

En este ejemplo las diferencias entre los espacios requeridos son mínimas y el propósito es ilustrar como influyen el cálculo de TRANSPACE las declaraciones temporales; pues por medio de la declaración ALLOCATE podemos cambiar la asignación de TRANSPACE para cada procedimiento en situaciones que la diferencia sea significativa.

Uso del ALLOCATE

Formato General

```
1          16
ALLOCATE   TRANSPACE = space
```

Las declaraciones ALLOCATE debe ser utilizada en las situaciones que el espacio requerido por las tarjetas de modificación de datos sea superior al estándar que fija el sistema ($TRANSPACE = SPACE/8$) o en la situación contraria en que casi no se usa tarjetas de modificación de datos y se desea por lo tanto disminuir el valor de TRANSPACE asumido por omisión, liberando así espacio para WORKSPACE.

La posición de la declaración ALLOCATE se ilustra en el siguiente ejemplo, que es el mismo que se describe anteriormente, sólo que se omitió el campo de especificación:

```
1          16
RUN NAME   ...
VARIABLE LIST ...
...
...
ALLOCATE   TRANSPACE = 1000
COMPUTE   NUEVA = ...
COMPUTE   NUEX = ....
IF
RECODE
RECODE
COUNT
*COMPUTE
*RECODE
FREQUENCIES ...
OPTIONS
STATISTICS
READ INPUT DATA
ALLOCATE   TRANSPACE = 900
*COMPUTE
PEARSON CORR
```

Una declaración ALLOCATE tiene efecto hasta que detecte otra ALLOCATE que modifique el espacio asignado.

Resumen

1. SPACE = WORKSPACE + TRANSPACE
2. TRANSPACE = SPACE/8 por omisión
3. TRANSPACE puede ser modificado para cada procedimiento por medio de ALLOCATE
4. SPACE es un parámetro dependiente de la memoria de cada equipo en particular.

IX. TARJETAS DE SELECCION DE DATOS

IX.1 La tarjeta SAMPLE, *SAMPLE: generación de muestra al azar

Formato General

1	16
SAMPLE	factor

Dónde factor debe ser un número menor que 1.0 y positivo.

1	16
SAMPLE	0.20

La declaración con factor = 0.20 indica al sistema que seleccione una muestra al azar del 20% del total de casos del archivo. El número de casos muestreados nunca será exactamente el 20% pero sí aproximadamente.

Todos los casos poseen una probabilidad igual de ser seleccionados para la muestra y esta probabilidad es igual al factor especificado.

Como la selección de la muestra es al azar, el proceso con un mismo factor para SAMPLE, aplicado sobre un mismo archivo, producirá muestras del 20% (cuando factor = 0.20) con características distintas si el procesamiento se realiza dos veces distintas. Para que el usuario se garantice el uso de la misma muestra deberá utilizar el SAVE FILE que salvará la parte del archivo muestreado.

Al igual que las declaraciones de modificación de datos tenemos el *SAMPLE que es el equivalente temporal que sólo tiene efecto sobre el procedimiento que le sigue en forma inmediata.

IX.2 SELECT IF, *SELECT IF: selección de casos

Formato General

1	16
SELECT IF	(expresión lógica)

Por medio de esta declaración el usuario puede seleccionar un subconjunto del universo. Por ejemplo, si deseamos seleccionar el subconjunto cuyo sexo es masculino (SEXO=1) la declaración sería:

1	16
SELECT IF	(SEXO EQ 1)

O sea, selecciona los casos para los cuales la expresión lógica es verdadera.

1	12
SELECT IF	(SEXO EQ 1 AND EDAD GE 12)

Este segundo ejemplo selecciona el subconjunto de hombres de 12 años y más.

También para esta declaración tenemos el *SELECT IF

IX.3 WEIGHT; *WEIGHT: Ponderación de casos

Formato General

1	16
WEIGHT	nombre de var

Las situaciones que requieren la ponderación de casos son variadas: por ejemplo al realizarse una encuesta se seleccionó una muestra para la parte urbana y una muestra para la parte rural; ambas representativas del universo. Por problema de costos la muestra de la zona rural se constituye de un menor número de casos. El tratamiento de las zonas geográficas en forma independiente no ofrecería problemas pero para el tratamiento de ambas para obtención de resultados totales requiere que los casos rurales tengan un peso superior a los casos urbanos. Supongamos que cada caso rural debe contar como 1.5 caso para que la muestra rural se transforme al mismo porcentaje de la muestra urbana.

La secuencia de instrucciones que crean la variable que será utilizada como variable de ponderación sería:

```

1                16
COMPUTE          PESO = 1.0
IF              (ZONA EQ 2) PESO = 1.5
WEIGHT          PESO

```

La ponderación, utilizando una variable que puede asumir valores con decimales, genera que el control de frecuencias también se realice con decimales: por ejemplo, para distribución de frecuencias o cruces un caso con PESO = 1.5 significa que se incremente el contador de frecuencias en 1.5, en la construcción de la tabla de frecuencias.

Esto motiva que el resultado final tenga valores decimales, que para efectos de mejor legibilidad serán redondeados al entero más próximo al momento de imprimir la tabla. Para los cálculos estadísticos, el SPSS utilizará los valores con decimales y no los valores redondeados para una mayor precisión.

El uso de la declaración permanente WEIGHT (sin asterisco) provocará que el archivo del sistema obtenido por medio de SAVE FILE quede permanentemente ponderado pues el SPSS para cada caso grabará la variable automática CASWGT que contiene el peso de cada caso (unidad de estudio).

Considerando que la ponderación no es acumulativa, el usuario podrá eliminar la ponderación existente en su archivo SPSS ponderando dicho archivo por una variable cuyo valor siempre será 1.0

Ejemplo

```

1                16
RUN NAME
GET FILE          ARCHPON
.....
COMPUTE          PESOFALS=1.0
WEIGHT          PESOFALS

```

Este ejemplo de otro uso de la ponderación es el siguiente: deseamos obtener el promedio de hijos por mujer de una muestra censal y de acuerdo a las características educacionales y grupos de edad decenales.

Las variables involucradas son:

<u>Nombre de variable</u>	<u>Código</u>	<u>Descripción del código</u>
SEXO	1	Hombre
	2	Mujer
HIJOS	00-25 ...	Número de hijos
	99	Ignorado
EDADEC	1	10-19 años
	2	20-29
	3	30-39
	4	40-49
	5	50 y +
NIVEL	0	Ninguno
	1	Primaria
	2	Secundaria
	3	Universitaria
	4	Otros

Debemos obtener en realidad 2 cruces:

- (a) Uno que nos informe el número de mujeres (frecuencia)
- (b) Uno que nos informe el número de hijos

y posteriormente en forma manual calcular el promedio de hijos.

La secuencia de instrucciones sería:

```
1          16
*SELECT IF    (SEXO EQ 2)
CROSSTABS    TABLES = EDADEC BY NIVEL
*SELECT IF    (SEXO EQ 2)
*IF          (HIJOS GT 25) HIJOS = 0
*WEIGHT      HIJOS
CROSSTABS    TABLES = EDADEC BY NIVEL
```

NOTA: tenemos que dejar HIJOS = 0 para los casos que traigan el valor HIJOS = 99 pues esto acarrearía un fuerte incremento en el número de hijos, distorsionando la realidad.

El primer crosstabs produciría

EDADEC	NIVEL	0	1	2	3	4
10-19	1					
20-29	2		107			
30-39	3					
40-49	4					
50 y +	5					

La casilla con el valor 107 ... indica el número de mujeres entre 20-29 años y nivel de instrucción primaria.

El segundo crosstabs:

EDADEC	NIVEL	0	1	2	3	4
10-19	1					
20-29	2		420			
30-39	3					
40-49	4					
50 y +	5					

O sea las 107 mujeres poseen en total 420 hijos y el promedio sería $420/107 = 3.9$ hijos por mujer.

IX.4 El orden de las tarjetas

<u>Declaración</u>	<u>Jerarquía</u>
SAMPLE	1 (más alto)
SELECT IF	2
RECODE	2
COMPUTE	2
IF	2
COUNT	2
WEIGHT	3
*SAMPLE	4
*SELECT IF	5
*RECODE	5
*COMPUTE	5
*IF	5
*COUNT	5
*WEIGHT	6 (más baja)

Así en forma permanente sólo se permite muestrear (SAMPLE), de la muestra, seleccionar casos (SELECT IF) y entonces ponderar (WEIGHT). Haciendo combinaciones entre declaraciones permanentes y temporales se permite seleccionar casos (SELECT IF) y entonces de los casos seleccionados obtener una muestra al azar (*SAMPLE) pues el orden de jerarquía debe ser obedecido.

X. LA TARJETA EDIT: CHEQUEO DE SINTAXIS

En proceso normal el SPSS realiza el chequeo de la sintaxis hasta la tarjeta del primer procedimiento y si no detecta error inicia el proceso. Los procedimientos subsiguientes son chequeados en la medida que son ingresados al sistema y al detectar un error se interrumpe el procesamiento. Esta situación puede provocar una pérdida de tiempo tanto hombre como de máquina bastante preciosa. Como ya se vio anteriormente al procesar el primer procedimiento el SPSS crea un archivo borrador (generalmente en disco) que sirve como input al proceso de los demás procedimientos y cuya lectura es varias veces más rápida que la lectura del archivo original.

Este aspecto puede ser solucionado por medio de la declaración EDIT que instruye el sistema para realizar el chequeo de sintaxis y el cálculo del WORKSPACE requerido por cada procedimiento. El chequeo se realiza hasta detectar la tarjeta FINISH.

Formato General

1	16
EDIT	

La declaración EDIT debe ser la primera tarjeta del programa

X.1 Funciones de la tarjeta EDIT

1. Chequear que los nombres de variables que se hacen referencia en las tarjetas de control hayan sido previamente definidos en VARIABLE LIST o creados en alguna declaración de modificación de datos. Así en VAR LABELS sólo podemos nombrar variables previamente definidas.

2. Chequear que la sintaxis de cada declaración en forma independiente sea correcta.
3. Cálculo del WORKSPACE para todas las tarjetas de procedimientos del programa hasta detectar la declaración FINISH.
4. Al detectar un error en el formato de la declaración no detecta que hay referencias a variables no definidas -este error aparecería en un próximo proceso EDIT.

Chequeos que no realiza un proceso EDIT

1. No chequea que el orden de las tarjetas sea correcto.
2. No detecta la falta de declaraciones obligatorias para el proceso de archivos del usuario (BCB) como son INPUT MEDIUM, N OF CASES, etc.
3. En VALUE LABELS no detecta que se perforó un valor alfa numérico en vez de numérico: "(P) SECUNDARIA" en vez de "(3) SECUNDARIA" no aparece como error pues la sintaxis está correcta.

X.2 El acceso a archivos:

1. Archivos del usuario:BCD

Para un proceso EDIT el SPSS no requiere acceso a los archivos que serán procesados por lo tanto:

- a) No se necesita especificar las tarjetas del sistema operativo que direccionan el archivo en el caso de INPUT MEDIUM ser TAPE o DISK.
- b) Para INPUT MEDIUM = CARD no se debe colocar el juego de datos después de READ INPUT DATA o READ MATRIX según sea el caso.

2. Archivos SPSS

Como el proceso de un archivo SPSS, los nombres de variables residen en el medio legible a máquina junto con los datos y son ingresados al sistema mediante el GET FILE, disponemos de 2 alternativas.

- a) Especificar las tarjetas del sistema operativo que dirreccionaron el archivo del sistema y mediante el GET FILE el proceso EDIT dispondrá de los nombres de variables existentes para el chequeo de referencias.

- b) Declarar en lugar de GET FILE una VARIABLE LIST que defina los nombres de las variables existentes en el archivo y por lo tanto no se necesita especificar las tarjetas del sistema operativo para direccionar donde reside el archivo SPSS que será procesado.

Obsérvese que: o declaramos GET FILE o VARIABLE LIST pero no ambos.

XI. ESTADISTICAS DESCRIPTIVAS Y DISTRIBUCION DE FRECUENCIAS

XI.1 Introducción

Normalmente el primer tratamiento sobre un juego de datos es la obtención de estadísticas descriptivas y/o de la distribución de frecuencias de cada variable para un análisis preliminar.

De este análisis preliminar puede derivar decisiones sobre técnicas a ser utilizadas para el posterior tratamiento de los datos; o sea, es una primera fotografía de los datos.

El SPSS dispone de dos subprogramas (procedimientos) para este trabajo: CONDESCRIPTIVE y FREQUENCIES

XI.2 CONDESCRIPTIVE: estadísticas descriptivas para variables continuas

Este subprograma (procedimiento) está orientado para producir estadísticas descriptivas para variables numéricas continuas. Las estadísticas disponibles son: media, error estándar, desviación estándar, varianza, kurtosis, asimetría, rango, mínimo y máximo. Posee además, la opción de generar un archivo cuyo contenido es variable normalizada Z.

Formato General

```
1          16
CONDESCRIPTIVE {lista de variables o
                {ALL
```

Ejemplos

1	16
CONDESCRIPTIVE VAR001 TO VAR010,VAR041,VAR081	
1	16
CONDESCRIPTIVE ALL	

Al especificar la palabra clave ALL el sistema procesará todas las variables definidas en el archivo y las definidas por medio de las tarjetas de transformación de datos.

1. Estadísticos disponibles

Tabla de equivalencia entre números y estadísticos

<u>N°</u>	<u>Estadístico</u>
1	Media
2	Error estándar
5	Desviación estándar
6	Varianza
7	Kurtosis
8	Coefficiente de Asimetría (skew ness)
9	Rango
10	Mínimo
11	Máximo

El usuario puede seleccionar los estadísticos que se desea por medio de la tarjeta STATISTICS, cuyo uso se ilustra con los siguientes ejemplos:

(a)

1	16
CONDESCRIPTIVE VARX,VARZ,VARC TO VARD	
STATISTICS 1,2,5,6	

En este caso sólo producirá:

- 1) Media; 2) Error estándar; 5) Desviación estándar
- 6) Varianza; para cada variable de la lista especificada.

(b)

1	16
CONDESCRIPTIVE VARX,VARZ,VARC	
STATISTICS ALL	

Producirá todos los estadísticos de la tabla.

(c)

1	16
CONDESCRIPTIVE VARX,VARZ,VARC	

También en este ejemplo se producirá todos los estadísticos o sea es equivalente a "STATISTICS ALL"

Nota: Como no es posible el cálculo de estadísticos, para variables alfanuméricas o numéricas que asuman valores superiores a 10¹³ el procedimiento enviará un mensaje de error en dichos casos.

2. Generación del archivo con la variable normal-Z-

El método más usado y universal de estandarizar una variable es por la normalización, o sea:

$$Z = \frac{X_i - \bar{X}}{SD_x}$$

- donde:
- Z = variable normal
 - \bar{X} = media de la variable X
 - X_i = valor de la variable X para el i-ésimo caso.
 - SD= desviación estándar de X.

Las características de la variable normal son:

- media = 0
- desviación estándar = 1

Uno de los usos de generar un archivo con las variables normalizadas es utilizarlo como entrada a los subprogramas BREAKDOWN y ANOVA (Análisis de Varianza) para efectos comparativos.

El archivo con las variables nombradas en el CONDESCRIPTIVE será generado si el usuario especifica "OPTIONS 3" y será grabado hacia la unidad lógica FT09 o hacia aquella especificada en "RAW OUTPUT UNIT".

El archivo será grabado en imagen de tarjeta en representación BCD (carácter) en el siguiente formato:

<u>Variable</u>	<u>Formato</u>	<u>Columna</u>
SEQNUM	F6.0	1-6
Clase tarjeta	F2.0	7-8
SUBFILE	F4	9-12
CASWGT	F4.2	13-16
variable ₁	F8.5	17-24
variable ₂	F8.5	25-32
....		
variable ₈	F8.5	73-80

SEQNUM, SUBFILE, CASWGT son las tres variables automáticas.

Si el número de variables en la lista del CONDESCRIPTIVE es superior a 8 el sistema grabará en el próximo registro, generando un número para la clase de tarjeta 1,2,3 ... etc.

Se grabará uno o más registros, dependiendo del número de variables, para cada caso en el archivo.

Para el formato F8.5 no se grabará punto decimal, motivo por el cual la lectura del archivo generado debe utilizar el mismo F8.5; el hecho de no grabar el punto decimal permite un rango entre -99.0 y 999.0.

Generalmente, el rango de la variable normal varía entre -99.0 y 99.0; si el cálculo en algún caso ultrapasa estos extremos el SPSS los hace igual al extremo.

Si la opción de tratamiento de los MISSING es la estándar o sea, exclusión de los missing en los cálculos estadísticos; el SPSS grabará el valor de 999.00000 para la variable normal de los casos cuyo valor de la variable original es missing.

3. Opciones del CONDESCRIPTIVE

Las opciones se especifican mediante la tarjeta OPTIONS que debe ser colocada a continuación de la tarjeta CONDESCRIPTIVE o a continuación de la tarjeta STATISTICS asociada. El orden de las tarjetas OPTIONS o STATISTICS puede ser cualquiera, pero ambas (si se especifican) deben estar a continuación del procedimiento al cual están asociadas.

Las opciones son:

- OPCION 1: informa al sistema que ignore la declaración de MISSING y trate los casos con valores declarados como missing como casos válidos. Si no se especifica Opción 1 el sistema asume el tratamiento estándar: excluye del cálculo los casos cuyos valores son missing (declarados en MISSING VALUES o ASSIGN MISSING).
- OPCION 2: informa al sistema para que no imprima los títulos de variables, los cuales son normalmente (sin opción = 2) impresos en los reportes de salida.
- OPCION 3: produce la generación del archivo con los valores de las variables normalizadas (Z)
- OPCION 4: provoca la impresión de un índice que indica para cada variable en qué página se encuentran las estadísticas producidas. Este índice (diccionario de referencia) trae dos listados:
- a) Uno ordenado en forma alfabética.
 - b) Otro ordenado en la forma que las variables se encuentran en el archivo.

El formato general para OPTIONS es:

1	16
OPTIONS	lista de números

Opciones estándares (si no se especifica OPTIONS)

- a) Excluye los missing de los cálculos.
- b) Imprime títulos extendidos
- c) No genera archivo de Z
- d) No produce el índice.

4. Limitaciones para CONDESCRIPTIVE

1) $WORKSPACE = NVARs * 50$

donde: WORKSPACE = espacio disponible y/o requerido, en bytes.

NVARs = número de variables nombradas implícitamente o explícitamente en la tarjeta de procedimiento.

- 2) Máximo de 250 elementos pueden aparecer en el campo de especificación.
Ejemplo: VAR1 TO VAR10 cuenta como 3 elementos.
Los delimitadores comunes (blanco y coma) no se contabilizan.
- 3) Máximo de 500 variables pueden ser nombradas implícitamente o explícitamente.

5. Ejemplos de uso del CONDESCRIPTIVE

A continuación se muestra un ejemplo de uso del CONDESCRIPTIVE con:

- a) El programa utilizado con OPTIONS 3,4

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES SPSSM - VERSION 6.01

08R25R76

PAGE

SPACE ALLOCATION FOR THIS RUN..

TOTAL AMOUNT REQUESTED 23552 BYTES

DEFAULT TRANSPLACE ALLOCATION 2944 BYTES

MAX NO OF TRANSFORMATIONS PERMITTED 29
 MAX NO OF RECODE VALUES 116
 MAX NO OF ARITHM.OR LOG.OPERATIONS 232

RESULTING WORKSPACE ALLOCATION 20608 BYTES

RUN NAME EJEMPLO DEL PROCEDIMIENTO -CONDESCRIPTIVE-
 VARIABLE LIST VAR001 TO VAR052
 INPUT FORMAT F1XED17X,F1.0,F4.0,F5.0,F4.2,F6.1,F5.1,9F4.1,2X,F4.1R
 7X,11F4.1,F7.0,2F5.1,F4.2R7X,F6.1,F5.1,14F4.1R7X,2F4.1,F4.2,
 F4.0,F4.2R

ACCORDING TO YOUR INPUT FORMAT, VARIABLES ARE TO BE READ AS FOLLOWS

VARIABLE	FORMAT	RECORD	COLUMNS
VAR001	F 7. 0	1	8- 14
VAR002	F 4. 0	1	15- 18
VAR003	F 5. 0	1	19- 23
VAR004	F 4. 2	1	24- 27
VAR005	F 6. 1	1	28- 33
VAR006	F 5. 1	1	34- 38
VAR007	F 4. 1	1	39- 42
VAR008	F 4. 1	1	43- 46
VAR009	F 4. 1	1	47- 50
VAR010	F 4. 1	1	51- 54
VAR011	F 4. 1	1	55- 58
VAR012	F 4. 1	1	59- 62
VAR013	F 4. 1	1	63- 66
VAR014	F 4. 1	1	67- 70
VAR015	F 4. 1	1	71- 74
VAR016	F 4. 1	1	77- 80
VAR017	F 4. 1	2	8- 11
VAR018	F 4. 1	2	12- 15
VAR019	F 4. 1	2	16- 19
VAR020	F 4. 1	2	20- 23
VAR021	F 4. 1	2	24- 27
VAR022	F 4. 1	2	28- 31
VAR023	F 4. 1	2	32- 35

ACCORDING TO YOUR INPUT FORMAT, VARIABLES ARE TO BE READ AS FOLLOWS

VARIABLE	FORMAT	RECORD	COLUMNS
VAR024	F 4. 1	2	36- 39
VAR025	F 4. 1	2	40- 43
VAR026	F 4. 1	2	44- 47
VAR027	F 4. 1	2	48- 51
VAR028	F 7. 0	2	52- 58
VAR029	F 5. 1	2	59- 63
VAR030	F 5. 1	2	64- 68
VAR031	F 4. 2	2	69- 72
VAR032	F 6. 1	3	8- 13
VAR033	F 5. 1	3	14- 18
VAR034	F 4. 1	3	19- 22
VAR035	F 4. 1	3	23- 26
VAR036	F 4. 1	3	27- 30
VAR037	F 4. 1	3	31- 34
VAR038	F 4. 1	3	35- 38
VAR039	F 4. 1	3	39- 42
VAR040	F 4. 1	3	43- 46
VAR041	F 4. 1	3	47- 50
VAR042	F 4. 1	3	51- 54
VAR043	F 4. 1	3	55- 58
VAR044	F 4. 1	3	59- 62
VAR045	F 4. 1	3	63- 66
VAR046	F 4. 1	3	67- 70
VAR047	F 4. 1	3	71- 74
VAR048	F 4. 1	4	8- 11
VAR049	F 4. 1	4	12- 15
VAR050	F 4. 2	4	16- 19
VAR051	F 4. 0	4	20- 23
VAR052	F 4. 2	4	24- 27

THE INPUT FORMAT PROVIDES FOR 52 VARIABLES. 52 WILL BE READ IT PROVIDES FOR 4 RECORDS (CARDS) PER CASE. A MAXIMUM OF 80 'COLUMNS' ARE USED ON A RECORD.

INPUT MEDIUM	CARD
N OF CASES	79
VAR LABELS	VAR001,POBLACION TOTALR
	VAR002,PCT POBLACION URBANAR
	VAR003,T.M.N.R
	VAR004,NATALIDADR
	VAR005,DENSIDADR
	VAR006,CONCENTR TIERRAR
	VAR007,PCT ASALARIADOSR
	VAR008,USO INTENS TIERRAR
	VAR009,PCT PROD EXPORTR
	VAR010,PCT MUJERES ACTIVASR
	VAR011,PCT ANALFABETOSR
	VAR012,PCT EMPL AGRICULTURAR
	VAR013,PCT EMPL INDUSTRA

VAR014,PCT SERV MODERNOSR
VAR015,PCT SERV TRADR
VAR016,PCT NIVEL SALARIALR
VAR017,PCT ASIST CENTROS ENSEÑANZAR
VAR018,PCT NO ASEG S PUBL ACT TOTALR
VAR019,PCT NO ASEG REN S PUBL REN TOTALR
VAR020,PCT NO ENERGIA ELECTRICAR
VAR021,PCT HOGARES CON TELEVR
VAR022,PCT HOGARES CON COCINA ELECTRA
VAR023,PCT FINCAS ENERGIA MECANICAR
VAR024,PCT FINCAS NO COMERCIALESR
VAR025,PCT FINCAS TRANSPORTE MECANICOR
VAR026,PCT FINCAS CONECTADAS CAM PAVR
VAR027,CONSULTAS MED ROR HABITR
VAR028,POBL TOTAL 73R
VAR029,PCT PUBL URBANA 73R
VAR030,T.M.N. 73R
VAR031,NATALIDAD 73R
VAR032,DENSIDAD 73R
VAR033,CONCENTR TIERRA 73R
VAR034,PCT ASALARIADOS 73R
VAR035,USO INTENS TIERRA 73R
VAR036,PCT PROD EXPORT 73R
VAR037,PCT MUJ ACTIVAS 73R
VAR038,PCT ANALFABETOS 73R
VAR039,PCT EMPL AGRICULT 73R
VAR040,PCT EMPL IND 73R
VAR041,PCT SERV MODERNOS 73R
VAR042,PCT SERV TRAD 73R
VAR043,PCT NIVEL SALARIAL 73R
VAR044,PCT ASIST CENTROS ENSEÑANZA 73R
VAR045,PCT ASEG S PUBL ACT TOTAL 73R
VAR046,PCT ASEG S PREN S PUBL REN TOTAL 73R
VAR047,PCT HOGARES CON TELEV 73R
VAR048,PCT HOGARES CON COCINA ELECT 73R
VAR049,PCT HOGARES CON TELLFONO 73R
VAR050,CONS MED POR HABIT 73R
VAR051,DISTANCIA CANTON CENTRAL R
VAR052,PLANIF FAM 73R
CONDESCRIPTIVE VAR028 TO VAR052

***** GIVEN WORKSPACE ALLOWS FOR 412 VARIABLES FOR CONDESCRIPTIVE PROBLEM *****

OPTIONS 3,4
READ INPUT DATA

b) La salida impresa de los estadísticos.

EJEMPLO DEL PROCEDIMIENTO -CONDESCRIPTIVE-
FILE NONAME (CREATION DATE = 08R25R76R

08R25R76

PAGE 5

VARIABLE VAR033 CONCENTR TIERRA 73

MEAN	39.694	STD ERROR	6.292	STD DEV	55.923
VARIANCE	3127.412	KURTOSIS	23.248	SKEWNESS	4.249
RANGE	406.500	MINIMUM	2.400	MAXIMUM	408.900

VALID OBSERVATIONS - 79 MISSING OBSERVATIONS - 0

VARIABLE VAR034 PCT ASALARIADOS 73

MEAN	67.889	STD ERROR	1.674	STD DEV	14.881
VARIANCE	221.437	KURTOSIS	-3.472	SKEWNESS	-3.698
RANGE	58.800	MINIMUM	28.600	MAXIMUM	87.400

VALID OBSERVATIONS - 79 MISSING OBSERVATIONS - 0

VARIABLE VAR035 USO INTENS TIERRA 73

MEAN	27.291	STD ERROR	2.365	STD DEV	21.017
VARIANCE	441.732	KURTOSIS	0.854	SKEWNESS	1.162
RANGE	91.400	MINIMUM	2.400	MAXIMUM	93.800

VALID OBSERVATIONS - 79 MISSING OBSERVATIONS - 0

VARIABLE VAR036 PCT PROD EXPORT 73

MEAN	17.753	STD ERROR	2.470	STD DEV	21.955
VARIANCE	482.010	KURTOSIS	1.508	SKEWNESS	1.495
RANGE	90.600	MINIMUM	0.0	MAXIMUM	90.600

VALID OBSERVATIONS - 79 MISSING OBSERVATIONS - 0

VARIABLE VAR037 PCT MUJ ACTIVAS 73

MEAN	34.178	STD ERROR	2.097	STD DEV	18.638
VARIANCE	347.379	KURTOSIS	0.503	SKEWNESS	0.972
RANGE	80.800	MINIMUM	8.800	MAXIMUM	89.600

VALID OBSERVATIONS - 79 MISSING OBSERVATIONS - 0

c) El diccionario de referencia (índice) producido por la opción 4.

EJEMPLO DEL PROCEDIMIENTO -CONDESCRIPTIVE-
FILE NONAME (CREATION DATE = 08R25R76I

08R25R76

PAGE 9

POSITIONAL INDEX

VARIABLE	PAGE	VARIABLE	PAGE	VARIABLE	PAGE	VARIABLE	PAGE
VAR028	4	VAR035	5	VAR042	6	VAR049	8
VAR029	4	VAR036	5	VAR043	7	VAR050	8
VAR030	4	VAR037	5	VAR044	7	VAR051	8
VAR031	4	VAR038	6	VAR045	7	VAR052	8
VAR032	4	VAR039	6	VAR046	7		
VAR033	5	VAR040	6	VAR047	7		
VAR034	5	VAR041	6	VAR048	8		

EJEMPLO DEL PROCEDIMIENTO -CONDESCRIPTIVE-
FILE NONAME (CREATION DATE = 08R25R76I

08R25R76

PAGE 10

ALPHABETIC INDEX

VARIABLE	PAGE	VARIABLE	PAGE	VARIABLE	PAGE	VARIABLE	PAGE
VAR028	4	VAR035	5	VAR042	6	VAR049	8
VAR029	4	VAR036	5	VAR043	7	VAR050	8
VAR030	4	VAR037	5	VAR044	7	VAR051	8
VAR031	4	VAR038	6	VAR045	7	VAR052	8
VAR032	4	VAR039	6	VAR046	7		
VAR033	5	VAR040	6	VAR047	7		
VAR034	5	VAR041	6	VAR048	8		

d) El mensaje producido por la opción 3 de generación del archivo con la variable Z que indica cómo fue grabado el archivo.

EJEMPLO DEL PROCEDIMIENTO -CONDESCRIPTIVE-
 FILE NONAME (CREATION DATE = 08R25R76)

25 Z-SCORES WERE WRITTEN ON LOGICAL UNIT 9 FOR 79 UNWEIGHTED CASES. 4 RECORDS OUTPUT PER CASE.
 FORMAT IS (F6.0,F2.0,A4,F4.2,8F8.5). TO INPUT Z-SCORES ONLY, WITHOUT SEQUENCING INFO, USE (I6X,8F8.50).
 MISSING Z-SCORES ARE OUTPUT AS 999.0. NON-MISSING BUT EXTREME Z-SCORES ARE TRUNCATED TO +99.0 OR -99.0

VARIABLE	OUTPUT RECORD NUMBER PER CASE	RECORD COLUMNS	UNWEIGHTED NUMBER OF MISSING CASES
SEONUM	1	1- 6	
RECORD N	1	7- 8	
SUBFILE	1	9-12	
CASMG1	1	13-16	
VAR028	1	17-24	0
VAR029	1	25-32	0
VAR030	1	33-40	0
VAR031	1	41-48	0
VAR032	1	49-56	0
VAR033	1	57-64	0
VAR034	1	65-72	0
VAR035	1	73-80	0
SEONUM	2	1- 6	
RECORD N	2	7- 8	
SUBFILE	2	9-12	
VAR036	2	17-24	0
VAR037	2	25-32	0
VAR038	2	33-40	0
VAR039	2	41-48	0
VAR040	2	49-56	0
VAR041	2	57-64	0
VAR042	2	65-72	0
VAR043	2	73-80	0

VARIABLE	OUTPUT RECORD NUMBER PER CASE	RECORD COLUMNS	UNWEIGHTED NUMBER OF MISSING CASES
SEONUM	3	1- 6	
RECORD N	3	7- 8	
SUBFILE	3	9-12	
VAR044	3	17-24	0
VAR045	3	25-32	0
VAR046	3	33-40	0
VAR047	3	41-48	0
VAR048	3	49-56	0
VAR049	3	57-64	0
VAR050	3	65-72	0
VAR051	3	73-80	0
SEONUM	4	1- 6	
RECORD N	4	7- 8	
SUBFILE	4	9-12	
VAR052	4	17-24	0

XI.3 FRECUENCIAS: Distribución de frecuencias

El subprograma FRECUENCIAS computa la distribución de frecuencias y los estadísticos descriptivos y se aplica a variables discretas o sea con un número limitado de categorías o valores.

El formato de la tarjeta que activa el subprograma FRECUENCIAS es:

1	16
FRECUENCIAS	{ GENERAL = o INTEGER = }

Bajo el modo GENERAL el usuario puede obtener distribución de frecuencias para cualquier tipo de variable:

- Alfanuméricas
- Numéricas enteras
- Numéricas con decimales

El modo INTEGER

Esta opción sólo permite el procesamiento de variables numéricas y enteras.

Selección entre: GENERAL, INTEGER

A pesar de que el modo GENERAL, como el propio nombre lo dice se aplica a cualquier tipo de variable y requiere menor esfuerzo del usuario al solicitar sus marginales es más recomendable el uso del modo INTEGER por las siguientes razones:

- a) Es más eficiente (veloz) en el procesamiento.
- b) Permite un uso óptimo de la memoria (WORKSPACE) disponible lo que significa permitir procesar un mayor número de variables en forma conjunta y por lo tanto un significativo ahorro en tiempo de máquina.

La comparación (obtenida por la U. de Chicago) entre GENERAL e INTEGER procesando un archivo del sistema en un equipo IBM 360/65 con las siguientes características:

- 244 variables
- 1012 casos
- 1200 value labels
- 200 var labels

para producir los marginales y estadísticos de 40 variables arrojó los siguientes resultados:

modo INTEGER 6.57 segundos de CPU
modo GENERAL17.10 segundos de CPU

Así que es recomendable el uso de las facilidades del RECODE para transformar variables alfanuméricas en numéricas; agrupar en categorías discretas las variables con decimales y utilizar el modo INTEGER para la obtención de marginales.

1. El modo GENERAL

Formato General

1	16	
FRECUENCIAS	GENERAL =	{ lista de variables o ALL }

Las especificación de la palabra clave "ALL" informa al sistema que elabore las distribuciones de frecuencias para todas las variables existentes.

El usuario deberá tener cuidado al utilizar "ALL" pues como se verá en la parte de limitaciones, para el modo GENERAL el sistema permitirá un número de categorías (valores) igual para todas las variables -si alguna variable ultrapasa el número de categorías permitidas el SPSS abandona el cálculo para dicha variable.

Ejemplos:

1	16	
FRECUENCIAS	GENERAL =	ALL
FRECUENCIAS	GENERAL =	VARI TO VAR9, VAR53, VARX, VARZ

2. El modo INTEGER

Formato General

1	16	
FRECUENCIAS	INTEGER =	{ nombre de var } (mínimo, máximo)/...
		o
		{ listas de vars }
	{ nombre de var } (mínimo, máximo)/...
		o
		{ lista de vars }

Los siguientes ejemplos ilustran la forma de uso:

1	16
FRECUENCIES	INTEGER = VAR1, VAR3 TO VAR8 (0,9), VAR10, VAR11 (0,89), VAR12 TO VAR40 (0,99)
FRECUENCIES	INTEGER = VAR1 (1,5), VAR3 TO VAR8 (0,9), VAR10 (1,30), VAR11 (0,89), VAR12 TO VAR40 (0,99)

Observaciones Importantes

- a) El rango especificado debe incluir los valores declarados como missing pues el sistema los detecta automáticamente, los toma en cuenta para la distribución relativa pero los excluye en la distribución ajustada y acumulada.
- b) Si el sistema detecta un valor fuera del rango especificado por (mínimo, máximo) será excluido de la tabla y el sistema sólo informará el número de casos con valores fuera del rango ("OUT OF RANGE") y no cuáles son esos valores.
- c) El rango especificado puede exceder el rango de valores reales con el único inconveniente de desaprovechamiento de la memoria disponible.
- d) En general como la obtención de los marginales es el primer tratamiento a que sometemos nuestro juego de datos es conveniente saber qué categorías no esperadas existen (o sea basura) por lo cual la fijación del rango está dada por el número de dígitos de cada variable: de 1 dígito → variación de 0-9, 2 dígitos → 0,99, etc. Por ejemplo, la variable SEXO sólo puede asumir dos valores válidos: 1,2. Pero si deseamos saber si hay basura, proveniente tanto de la etapa de codificación como de la etapa de digitación (perforación), la fijaríamos "SEXO (0,9)".

3. Estadísticos disponibles

<u>Número</u>	<u>Estadístico</u>
1	Media
2	Error estándar
3	Mediana
4	Moda
5	Desviación estándar
6	Varianza
7	Kurtosis
8	Coefficiente de Asimetría
9	Rango
10	Mínimo
11	Máximo

Al contrario del subprograma CONDESCRIPTIVE, si el usuario no especifica una tarjeta STATISTICS, ningún estadístico será calculado.

4. Opciones del FREQUENCIES

- Opción 1: el sistema ignora la declaración de los missing. Si no se usa esta opción el tratamiento de los casos con valores declarados missing es el estándar.
- Opción 2: suprime la impresión de los value labels. Esta opción incrementa el espacio de memoria disponible. Es recomendable su uso junto con la opción 5 ó 6 o para variables con un rango superior a 100 categorías.
- Opción 3: provoca la salida impresa en formato de 8 1/2 por 11 pulgadas, propio para fotocopiar o para la encuadernación.
- Opción 4: provoca que la salida se direcciona hacia un archivo en cinta o disco para permitir posteriores reproducciones sucesivas. El registro lógico contendrá el control de carro ASA y un largo, LRECL = 80.
Si la opción 4 es solicitada, la opción 3 es automáticamente utilizada.
- Opción 5: provoca que la impresión de todas las variables se realice en el formato condensado (ver ejemplos de salida al final de este capítulo)
- Opción 6: provoca que la impresión de las variables que ocuparían más de una página, se realice en formato condensado. Las demás variables serán impresas en formato normal.
- Opción 7: provoca sólo la impresión de los estadísticos solicitados en STATISTICS, eliminando las tablas de frecuencias. La función de esta opción es permitir al usuario obtener estadísticos que el CONDESCRIPTIVE no dispone: "MEDIANA" y "MODA".

Opción 8: provoca la producción de un histograma para cada variable de la lista especificada para FREQUENCIES. El histograma se imprime en la página siguiente a la tabla de frecuencias. Los estadísticos, si fueron solicitados por medio de STATISTICS serán impresos después del histograma.

Opción 9: Informa al sistema para que imprima un índice con nombre de variable y la página de la tabla de frecuencias. Serán producidos 2 índices (diccionario de referencia):

- a) uno por orden en que las variables fueran definidas.
- b) por orden alfabético.

Opciones estándares (si no se especifica OPTIONS)

- (a) Excluye los missing de todos los porcentajes para Modo INTEGER y de la frecuencia ajustada y acumulada para el modo GENERAL. En ambos modos los missing serán excluidos de los cálculos estadísticos.
- (b) Todos los var labels y value labels serán impresos (si fueron especificados).
- (c) La salida se hará por impresora y no crea un archivo con la salida.
- (d) Tablas de frecuencias serán impresas en el formato normal.
- (e) No se producen histogramas.
- (f) No produce diccionario de referencia (índice).

5. Limitaciones para FREQUENCIES

1) MODO GENERAL

$$\text{WORKSPACE} = 10 * \text{NVALS} + 12 * \text{NVARs}$$

donde:

WORKSPACE = espacio de memoria en bytes.

NVALS = número total de valores encontrados, sumado para todas las variables.

NVARs = número total de variables nombradas implícita o explícitamente.

En general disponemos del valor WORKSPACE y NVALS con lo cual podemos calcular NVALS. Este valor también puede ser obtenido mediante un proceso EDIT, lo cual es más recomendable.

MODO INTEGER

$$\text{WORKSPACE} = 4 * (\text{NVALS} + 1) + 20 * \text{MAXRANGE}$$

WORKSPACE = espacio de memoria en bytes
NVALS = número total de valores esperados que puede ser obtenido por

$$\sum (\max_i - \min_i + 1)$$

donde:

\max_i = valor máximo de la variable i-ésima

\min_i = valor mínimo de la variable i-ésima

MAXRANGE = rango máximo especificado en la lista, o sea es:

$$\underline{\max_i - \min_i + 1}$$

El componente 20 * MAXRANGE es el espacio requerido para almacenar los value labels.

Si se utiliza opción 2 o 5 o 7 sin la opción 8, la fórmula queda reducida a:

$$\text{WORKSPACE} = 4 * (\text{NVALS} + 1) \quad \dots \text{ OPCION } 2;5;7$$

2) Para ambos modos: máximo de 500 variables pueden ser nombradas en un procedimiento FREQUENCIES.

6. Ejemplos de salida del FREQUENCIES

Ejemplo a): El uso del formato condensado para esta misma variable (VAR082) se muestra en el ejemplo d)

EJEMPLO DEL PROCEDIMIENTO - FREQUENCIES -
FILE FESALEMB (CREATION DATE = 08R24R76)

08R25R76 PAGE 4

VAR082 AÑO DE NACIMIENTO DEL PRIMER NACIDO VIVO

CATEGORY LABEL	CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	ADJUSTED FREQ (PCT)	CUM FREQ (PCT)
	37.	1	0.0	0.1	0.1
	39.	1	0.0	0.1	0.1
	40.	1	0.0	0.1	0.2
	41.	7	0.3	0.4	0.5
	42.	10	0.5	0.5	1.1
	43.	16	0.8	0.9	1.9
	44.	19	0.9	1.0	2.9
	45.	19	0.9	1.0	3.9
	46.	22	1.1	1.2	5.1
	47.	30	1.5	1.6	6.7
	48.	27	1.3	1.4	8.2
	49.	42	2.1	2.2	10.4
	50.	52	2.6	2.8	13.2
	51.	47	2.3	2.5	15.7
	52.	58	2.9	3.1	18.8
	53.	59	2.9	3.1	21.9
	54.	41	2.0	2.2	24.1
	55.	51	2.5	2.7	26.8
	56.	55	2.7	2.9	29.7
	57.	54	2.7	2.9	32.6
	58.	64	3.2	3.4	36.0
	59.	66	3.3	3.5	39.6
	60.	81	4.0	4.3	43.9
	61.	69	3.4	3.7	47.5
	62.	68	3.4	3.6	51.2
	63.	79	3.9	4.2	55.4
	64.	86	4.3	4.6	60.0
	65.	74	3.7	3.9	63.9
	66.	80	4.0	4.3	68.2
	67.	70	3.5	3.7	71.9
	68.	69	3.4	3.7	75.6
	69.	93	4.6	5.0	80.5
	70.	90	4.5	4.8	85.3
	71.	108	5.4	5.8	91.1
	72.	114	5.7	6.1	97.2

EJEMPLO DEL PROCEDIMIENTO - FREQUENCIES -
FILE FESALEMB (CREATION DATE = 08R24R76m)

08R25R76 PAGE 5

	73.	53	2.6	2.8	100.0
	0.	129	6.4	MISSING	100.0
	99.	2	0.1	MISSING	100.0
	TOTAL	2007	100.0	100.0	

VALID CASES 1876 MISSING CASES 131

Ejemplo b): salida con el uso de opción 8 -producción del histograma. La opción 9 produce el diccionario de referencia que se muestra en el ejemplo c).

EJEMPLO DEL PROCEDIMIENTO - FREQUENCIES -

08R25R76

PAGE 6

FREQUENCIES GENERAL=VAR014,VAR015,VAR016,VAR019,VAR025,VAR078
OPTIONS 8,9

EJEMPLO DEL PROCEDIMIENTO - FREQUENCIES -

08R25R76

PAGE 15

FILE FESALEMB (CREATION DATE = 08R24R76)

VAR025 EDUCACION ESPOSO

CATEGORY LABEL	CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	ADJUSTED FREQ (PCT)	CUM FREQ (PCT)
	0.	716	35.7	35.7	35.7
	1.	73	3.6	3.6	39.3
	2.	544	27.1	27.1	66.4
	3.	186	9.3	9.3	75.7
	4.	256	12.8	12.8	88.4
	5.	108	5.4	5.4	93.8
	6.	29	1.4	1.4	95.3
	7.	67	3.3	3.3	98.6
	8.	25	1.2	1.2	99.9
	9.	3	0.1	0.1	100.0
	TOTAL	2007	100.0	100.0	

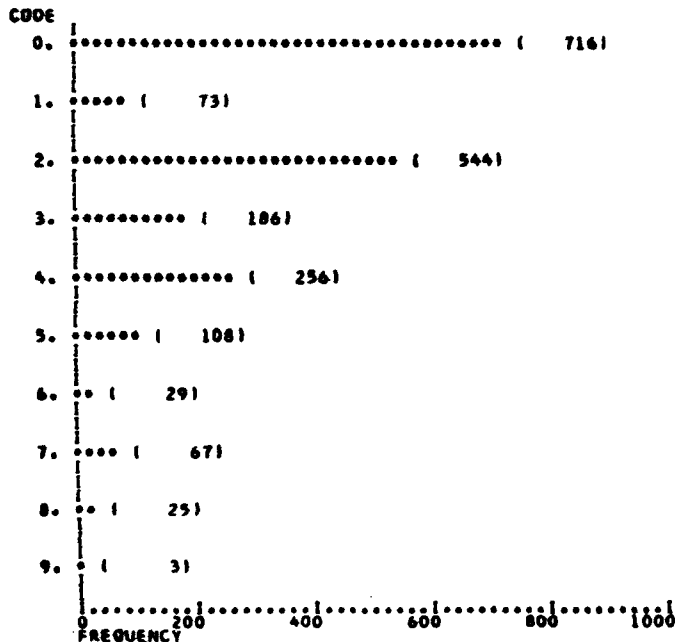
EJEMPLO DEL PROCEDIMIENTO - FREQUENCIES -

08R25R76

PAGE 16

FILE FESALEMB (CREATION DATE = 08R24R76)

VAR025 EDUCACION ESPOSO



VALID CASES 2007 MISSING CASES 0

Ejemplo c): Diccionario de referencia (INDICE) producido por el uso de la opción 9. La solicitud de este diccionario fue hecha por el procedimiento del ejemplo b).

EJEMPLO DEL PROCEDIMIENTO - FRECUENCIES -
FILE FESALEMB (CREATION DATE = 08R24R76)

08R25R76 PAGE 19

POSITIONAL INDEX

VARIABLE	PAGE	VARIABLE	PAGE	VARIABLE	PAGE	VARIABLE	PAGE
VAR014	7	VAR016	11	VAR025	15		
VAR015	9	VAR019	13	VAR078	17		

EJEMPLO DEL PROCEDIMIENTO - FRECUENCIES -
FILE FESALEMB (CREATION DATE = 08R24R76)

08R25R76 PAGE 20

ALPHABETIC INDEX

VARIABLE	PAGE	VARIABLE	PAGE	VARIABLE	PAGE	VARIABLE	PAGE
VAR014	7	VAR016	11	VAR025	15		
VAR015	9	VAR019	13	VAR078	17		

Ejemplo d): Formato condensado y estadísticos. Se utilizó opción 6 y STATISTICS ALL.

EJEMPLO DEL PROCEDIMIENTO - FRECUENCIES -
FILE FESALEMB (CREATION DATE = 08R24R76)

08R25R76 PAGE 2

VAR062 AÑO DE NACIMIENTO DEL PRIMER NACIDO VIVO

CODE	FREQ	ADJ PCT	CUM PCT	CODE	FREQ	ADJ PCT	CUM PCT	CODE	FREQ	ADJ PCT	CUM PCT
37.	1	0	0	50.	52	3	13	62.	68	4	51
39.	1	0	0	51.	47	3	16	63.	79	4	55
40.	1	0	0	52.	58	3	19	64.	86	5	60
41.	7	0	1	53.	59	3	22	65.	74	4	64
42.	10	1	1	54.	41	2	24	66.	80	4	68
43.	16	1	2	55.	51	3	27	67.	70	4	72
44.	19	1	3	56.	55	3	30	68.	69	4	76
45.	19	1	4	57.	54	3	33	69.	93	5	81
46.	22	1	5	58.	64	3	36	70.	90	5	85
47.	30	2	7	59.	66	4	40	71.	108	6	91
48.	27	1	8	60.	81	4	44	72.	114	6	97
49.	42	2	10	61.	8	4	48	73.	53	3	100

CODE	FREQ	MISSING DATA		CODE	FREQ
		CODE	FREQ		
0.	129	99.	2		

MEAN	61.088	STD ERR	0.190	MEDIAN	62.176
MODE	77.000	STD DEV	8.232	VARIANCE	67.764
KURTOSIS	-0.795	SKEWNESS	-0.430	RANGE	36.000
MINIMUM	37.000	MAXIMUM	73.000		
VALID CASES	1876	MISSING CASES	131		

XII. CROSSTABS: TABULACIONES CRUZADAS Y MEDIDAS DE ASOCIACION

XII.1 Subprograma CROSSTABS: modos - GENERAL, INTEGER

En forma similar al FRECUENCIAS, el subprograma CROSSTABS puede operar en dos modos: GENERAL e INTEGER y las diferencias son en términos generales las mismas.

El modo GENERAL:

- a) Permite procesar variables numéricas enteras o con decimales y alfanuméricas
- b) Requiere un menor esfuerzo en la preparación de las tarjetas de procedimiento. No requiere que el usuario especifique los rangos de cada variable.
- c) Permite un máximo de 250 valores únicos para cualquier tipo de variable involucrada en el cruce, no importando que el rango de la variable sea superior a 250.
- d) Hay un desaprovechamiento de la memoria disponible ya que reserva un tamaño igual de memoria para cada tabla solicitada -esto implica que reserva una cantidad X de bytes para un cruce de 2 x 2 y el mismo X para una tabla de 100 x 3.
- e) No es recomendable para archivos con un número de casos elevado.

El modo INTEGER

- a) La velocidad de procesamiento es bastante superior a la del modo general.
- b) Permite la producción de medidas de asociación (solicitudes por STATISTICS) que no están disponibles en el otro modo.
- c) Permite un aprovechamiento óptimo de la memoria disponible: o sea, en un solo paso nos permite elaborar un mayor número de tablas que el modo general; pues cada cruce ocupará el número de bytes que realmente requiere.
- d) Requiere un mayor esfuerzo para la preparación de las tarjetas de procedimiento y conocimiento previo sobre el rango de las variables.
- e) Sólo opera con variables numéricas enteras.
- f) Es más recomendable para archivos con un número de casos elevado, pues sus características implicarán en una significativa reducción del tiempo de máquina comparado con el tiempo que se requeriría para el modo general.

La selección de uso entre un modo u otro, depende por ende de cada aplicación en particular.

Los elementos para la decisión están mencionados y cada usuario deberá decidir cuál modo utilizará.

XII.2 CROSSTABS: modo GENERAL

Formato General

<u>1</u>	<u>16</u>
CROSSTABS	TABLES= $\left[\begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right] \text{ BY } \left[\begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right] \text{ BY } \dots \text{ BY } \left[\begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right] \left[/ \right.$ $\left. \left[\begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right] \text{ BY } \left[\begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right] \text{ BY } \dots \text{ BY } \left[\begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right] \right] \left[/ \dots \right.$

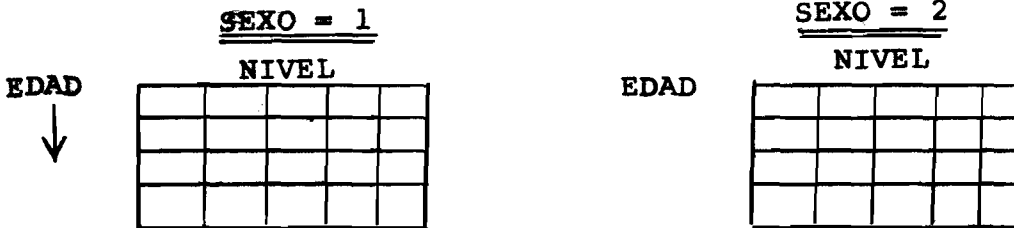
Ejemplos:

(a)

1	16
CROSSTABS	TABLES = EDAD BY NIVEL BY SEXO / EDAD BY NIVEL BY REGION

En este ejemplo la primera lista de especificación de cruces es: "EDAD BY NIVEL BY SEXO" y la segunda lista, que debe estar separada de la primera por el delimitador especial "/" es: "EDAD BY SEXO BY REGION". El primer BY nos delimita cuáles son las variables de fila y variables de columna y a partir del 2º y subsiguiente BY, cuáles son las variables de control.

La primera lista de cruces nos producirá dos tablas que serían de la siguiente forma:



Así:

EDAD ----- variable de fila
 NIVEL ----- variable de columna
 SEXO ----- variable de control

La segunda lista de cruces "EDAD BY NIVEL BY REGION" nos produce una tabla de EDAD x NIVEL para cada código de REGION.

NUMERO DE CELDAS: Si conocemos los valores que pueden asumir las variables involucradas podemos calcular el número de celdas que requiere esta solicitud de cruces. Supongamos:

EDAD --- varía de 0 - 99 años
 NIVEL -- varía de 0 - 7 categorías
 SEXO --- varía de 1 - 2 categorías
 REGION - varía de 1 - 4 categorías

Así: lista "EDAD BY NIVEL BY SEXO" --- $100 * 7 * 2 = 1400$
lista "EDAD BY NIVEL BY REGION" = $100 * 7 * 4 = 2800$

El total de celdas (no bytes) es = $1400 + 2800 = 4200$

(b)

1	16
CROSSTABS	TABLES = VAR001,VAR002,VAR005 TO VAR010 BY VAR021,VAR022.

Este ejemplo producirá la siguiente serie de tablas:

VAR001 BY VAR021
VAR001 BY VAR022
VAR002 BY VAR021
VAR002 BY VAR022
VAR005 BY VAR021
VAR005 BY VAR022

...
...

VAR010 BY VAR021
VAR010 BY VAR022

De esta forma al lado de cualquier "BY" podemos especificar una lista de variables y los cruces que serán calculados será la combinación (no redundante) de las variables especificadas.

NOTA: El único cuidado que el usuario debe tener al utilizar el modo general es saber si la cantidad de tablas que solicita no excederá el espacio de memoria disponible.

En el caso de que esto ocurra, la construcción de la tabla que excede el espacio disponible será abandonada (obviamente al intermedio del proceso) y el sistema seguirá con las otras tablas.

Un mensaje informando al respecto será enviado por el sistema.

XII.3 CROSSTABS: modo INTEGER

Formato General

1	16
CROSSTABS	$\text{VARIABLES} = \left\{ \begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right\} (\text{minimo,maximo}) , \dots \left\{ \begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right\} (\text{min,max}) /$
	$\text{TABLES} = \left\{ \begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right\} \text{BY} \left\{ \begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right\} \text{BY} \dots \text{BY} \left\{ \begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right\} \left[/ \right.$
	$\left. \left\{ \begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right\} \text{BY} \left\{ \begin{array}{c} \text{nombre de var} \\ \cdot \\ \text{lista de vars} \end{array} \right\} \dots \dots \dots \right]$

Ejemplo:

1	16
RUN NAME	EJEMPLO DE CROSSTABS
VARIABLE LIST	VARX01 TO VARX20
...	
...	
CROSSTABS	VARIABLES = VARX01(1,3),VARX05 TO VARX08 (0,20), VARX11 (0,99), VARX12 (1,5), VARX18 (0,9) / TABLES = VARX05 TO VARX18 BY VARX01 / VARX05 TO VARX11, VARX18 BY VARX12 BY VARX01
STATISTICS	ALL
OPTIONS	9
READ INPUT DATA	

En este ejemplo la especificación VARIABLE LIST define una lista de variables VARX01, VARX02, VARX03, ... VARX19, VARX20. En la especificación del CROSSTABS la lista "VARIABLES =" define que las siguientes variables serán utilizadas para la elaboración de cruces: VARX01,VARX05,VARX06,VARX07,VARX08,VARX11,VARX12,VARX18. O sea, el uso de la convención "TO" en la parte "VARIABLES =" se refiere al orden en que las variables fueron definidas. Por otro lado en la parte "TABLES =" el uso de la convención "TO" se refiere al orden en que las variables fueron declaradas en la parte "VARIABLES =". De esta forma la lista (en TABLES =):

"VARX05 TO VARX18" = VARX05,VARX06,VARX07,VARX08,VARX11,
VARX12,VARX18.

Se debe además hacer notar que en la parte "TABLES =" sólo se permite hacer referencia a variables declaradas en "VARIABLES:". La especificación de la parte "VARIABLES =" es que define que se seleccionó el modo integer y esta parte de be preceder a la especificación de "TABLES = ".

NUMERO DE CELDAS:

Para el modo integer, como exige la declaración de los rangos de cada variable involucrada en los cruces, nos permite calcular exactamente el número total de celdas requerido.

Para el ejemplo en análisis el cruce de VARX05 BY VARX01 requiere $21 * 3 = 63$ celdas; el cruce de VARX05 BY VARX12 BY VARX01 requiere $21 * 5 * 3 = 315$ celdas. La sumatoria de las celdas requeridas por cada cruce nos dará el total. El cálculo en detalle se muestra a continuación:

CRUCE

CELDAS

... primera lista de cruces ...

VARX05 BY VARX01	21 * 3 = 63
VARX06 BY VARX01	21 * 3 = 63
VARX07 BY VARX01	21 * 3 = 63
VARX08 BY VARX01	21 * 3 = 63
VARX11 BY VARX01	100 * 3 = 300
VARX12 BY VARX01	5 * 3 = 15
VARX18 BY VARX01	10 * 3 = 30

... segunda lista de cruces ...

VARX05 BY VARX12 BY VARX01	21 * 5 * 3 = 315
VARX06 BY VARX12 BY VARX01	21 * 5 * 3 = 315
VARX07 BY VARX12 BY VARX01	21 * 5 * 3 = 315
VARX08 BY VARX12 BY VARX01	21 * 5 * 3 = 315
VARX11 BY VARX12 BY VARX01	100 * 5 * 3 = 1500
VARX18 BY VARX12 BY VARX01	10 * 5 * 3 = 150

$$\begin{aligned} \text{TOTAL CELDAS} &= (4 * 63) + 300 + 15 + 30 + (4 * 315) + 1500 + 150 \\ &= \underline{3407 \text{ celdas}} \end{aligned}$$

Como se verá en las limitaciones, el número de celdas no podrá exceder un valor específico para un WORKSPACE dado; y por otro lado cual sería el WORKSPACE para un dado número de celdas.

XII.4 Los VALUE LABELS e inclusión de los MISSING en los cruces

A excepción que el usuario utilice opción 2, el sistema imprimirá los value labels si fueron especificados. A pesar de que la declaración VALUE LABELS permite hasta 20 caracteres para el título de cada categoría, en este procedimiento, por requerimientos de espacio, solamente los primeros 16 caracteres serán truncados en dos parte de 8 caracteres cada una. Para una mejor visualización véase el punto XII.8 de ejemplos de la salida del CROSSTABS.

La selección de la opción 7, sólo para el modo integer, permite al usuario incluir los casos que traen valores missing en el cruce sin que esto afecte los porcentajes ni tampoco los tomará en cuenta para el cálculo de los estadísticos. El subprograma además agregará la letra "M" a continuación de la frecuencia para explicitar que son casos Missing. En la sección de ejemplos hay una ilustración de este tipo de opción.

XII.5 Los ESTADISTICOS disponibles

<u>Número</u>	<u>Estadístico</u>	<u>Observaciones</u>
1	Chi-cuadrado	Para tablas de 2 x 2 y menos de 21 casos aplica el test exacto de Fisher. Para tablas de 2 x 2 con 21 o más casos aplica el chi-cuadrado <u>corregido</u> de Yates.
2	Phi para tablas 2 x 2; V de CRamer para tablas mayores	
3	Coefficiente de contig <u>en</u> cia	
4	LAMBDA, simétrico y asi <u>m</u> étrico	
5	Coefficiente de incerti- dumbre, simétrico y asi <u>m</u> étrico	
6	TAU B de Kendall	
7	TAU C de Kendall	
8	GAMMA GAMMA Parcial y Gamma de orden cero ... disponible sólo para tablas de 3 y + dimensiones en el modo INTEGER	
9	D de Somer, simétrico y asimétrico	
10	ETA	Sólo para variables numéricas

Los estadísticos deben ser solicitados en forma explícita, ya que no serán calculados si no se especifica la tarjeta STATISTICS.

1	16
STATISTICS	{ ALL o lista de números }

Al contrario de CONDESCRIPTIVE y FREQUENCIES, en los cuales el orden de las tarjetas OPTIONS y STATISTICS puede ser cualquiera; el CROSSTABS exige que la tarjeta STATISTICS sea puesta después de OPTIONS si es que se pide alguna opción.

Los siguientes ejemplos ilustran las distintas posibilidades:

1	16
{ CROSSTABS OPTIONS	
{ CROSSTABS STATISTICS	
{ CROSSTABS OPTIONS STATISTICS	

XII.6 OPCIONES disponibles para CROSSTABS

- OPCION 1: informa al sistema para que ignore la declaración MISSING -esto significa que los valores declarados como missing serán tratados como códigos válidos. Si no se especifica esta opción el sistema automáticamente excluye los casos con valores declarados como missing tanto del cruce como de los estadísticos.
- OPCION 2: informa al sistema para que elimine la impresión de títulos. Tanto los VAR LABELS como VALUE LABELS no saldrán impresos.
- OPCION 3: suprime la impresión de los porcentajes de fila (ROW PCT)
- OPCION 4: suprime la impresión de los porcentajes de columna (COL PCT)
- OPCION 5: suprime la impresión de los porcentajes totales (TOT PCT)

.... opciones 6, 7, 8 disponibles sólo para el modo INTEGER

OPCION 6: elimina la impresión de value labels, pero imprime los VAR LABELS.

OPCION 7: informa al sistema para que incluya los casos con valores declarados como missing en la tabla cruzada sin que la frecuencia de estos casos influya en los porcentajes; ni tampoco sean incluidos en el cálculo de los estadísticos. La frecuencia de los valores missing va acompañada de la letra "M" para indicar que se refiere a valores missing.

OPCION 8: informa al sistema para que los valores de la variable de fila sean impresos en el orden inverso; o sea, la primera fila de la tabla contendrá las frecuencias del valor máximo de la variable de fila y la última fila contendrá el valor (código) mínimo. El orden de la variable de columna permanece inalterada o sea de menor a mayor.

...

OPCION 9: produce la impresión de un índice informando los cruces producidos y el número de la página donde comienza cada cruce.

El formato general de la tarjeta OPTIONS es

1	16
OPTIONS	lista de números

XII.7 LIMITACIONES del subprograma CROSSTABS

a) Limitaciones para el modo GENERAL

- 1) El número máximo de nombres de variables es de 200. Cada nombre de variable cuenta como "1" y todas las variables especificadas por un "TO" cuenta como uno.

Así:

TABLES = VAR001, VAR002 TO VAR090

tendría un total de 2

- 2) El máximo número de valores individuales (categorías) que puede asumir una variable es de 250.
- 3) Requerimiento de espacio de memoria.

$$\text{WORKSPACE} = 4 * (\text{NC} + 1) * (\text{D} + 2)$$

NC = número total de celdas requerido por el total de tablas solicitado.

D = número de dimensiones de la tabla con la mayor dimensión.

WORKSPACE = espacio de memoria en byte.

Así teniendo 10 cruces de 2 x 2 y 10 cruces de 2 x 2 x 2 tenemos que:

$$\begin{aligned} \text{NC} = \text{TOTAL CELDAS} &= 10 * (2 * 2) + 10 * (2 * 2 * 2) = \\ &= 40 + 80 = 120 \end{aligned}$$

$$\text{D} = 3$$

$$\text{WORKSPACE} = 4 * (120 + 1) * (3 + 2) = 2420 \text{ bytes}$$

Por otro lado si disponemos de WORKSPACE = 20.000 y D = 3

$$\begin{aligned} \text{NC} &= \frac{\text{WORKSPACE}}{4 * (\text{D} + 2)} - 1 = \frac{20.000}{4 * (3 + 2)} - 1 \\ &= \frac{20.000}{20} - 1 = 1.000 - 1 = 999 \end{aligned}$$

O sea, el máximo de celdas para cada tabla es 999

- 4) Se permite un máximo de 20 listas de cruces. Así:

"TABLES = VAR1 TO VAR5 BY VARX / VAR2, VAR3 BY VARX, VARZ BY VAR7"

trae dos listas de cruces delimitadas por el "/".

- 5) La dimensión máxima permitida es 10 (esto es 9 BY); con un máximo de 8 variables de control.

b) Limitaciones para el modo INTEGER

- 1) El máximo de nombres de variables en la parte "VARIABLES =" es 100. VAR1 TO VAR10 se cuenta como 1.
- 2) El máximo de variables que pueden ser nombrados explícita o implícitamente ("TO") en la parte "TABLES" es de 100.
- 3) No acepta variables alfanuméricas o numéricas con decimales -sólo se aplica a variables numéricas enteras.
- 4) Se permite un máximo de 20 listas de especificación de cruces, en la parte "TABLES"=". Cada lista está delimitada por un "/".
- 5) Requerimientos de memoria.

$$\text{WORKSPACE} = 4 * \text{NC} + 20 * \text{MAXLAB}$$

NC = número total de celdas requerido por todas las cruces.

MAXLAB = número máximo de value labels involucrado en alguna tabla.

WORKSPACE = espacio de memoria en bytes.

Ejemplo: 10 cruces de 2 x 2 y 10 cruces de 2 x 2 x 2

NC = $10 * (2 * 2) + 10 * (2 * 2 * 2) = 120$ celdas.

MAXLAB = $2 * 2 * 2 = 8$ (la tabla que requiere el máximo de value labels).

Luego

$$\text{WORKSPACE} = 4 * 120 + 20 * 8 = 480 + 160 = \underline{640 \text{ bytes}}$$

Obsérvese que el mismo ejemplo del modo GENERAL donde se requiere 2420 bytes contra 640 del modo integer.

- 6) La dimensión máxima es 8 (o sea: 7 BY), pero no más de 6 variables de control pueden estar involucradas en cualquier juego de tablas.

XII.8 Ejemplos de salida del CROSSTABS

Ejemplo a): Listado que produce el sistema del programa utilizado, recuperando un archivo SPSS (GET FILE) y la primera solicitud de cruces.

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES SPSSM - VERSION 6.01

08R27076

PAGE 1

SPACE ALLOCATION FOR THIS RUN..

TOTAL AMOUNT REQUESTED 23552 BYTES

DEFAULT TRANSPOSE ALLOCATION 2944 BYTES

MAX NO OF TRANSFORMATIONS PERMITTED 29
MAX NO OF RECODE VALUES 116
MAX NO OF ARITHM.OR LOG.OPERATIONS 232

RESULTING WORKSPACE ALLOCATION 20608 BYTES

RUN NAME EJEMPLO DEL PROCEDIMIENTO - CROSSTABS -
GET FILE FESALEMB

FILE FESALEMB HAS 81 VARIABLES

THE SUBFILES ARE..

NAME	NO OF CASES
FESALEMB	2007

PAGESIZE 83
VAR LABELS VAR013, EDAD DECLARADA
VALUE LABELS VAR013 (1) RURAL (2) URBANA (3) METROPOLITANA R
VAR014 (1) NO SE APLICA (1) DENTRO HOGAR (2) FUERA HOGAR
VAR015 (3) DENTRO Y FUERA (4) LUGAR N.D. (5) EMPL. DOMESTICA
(9) NO RESPONDE R
VAR016 (1) IGNORADO (1) VENEDORES (2) OBREROS NO CALIF.
(3) SERV. DOMESTICO (4) CUENTA PROPIA (5) OBRERO CALIF.
(6) EMPLACC-1 (7) EMPLACC-2 (8) PEO. EMPRESARIO
(9) JEFES ADMINIST. R
VAR025 (0) NINGUNA-ANALFAB (1) NING.-ALFABETO (2) PRIMARIA 1-3
(3) PRIMARIA 4-5 (4) PRIM. COMPLETA (5) SECUNCIARIA 1-3
(6) SECUNCIARIA 4-5 (7) SECUNC. COMPLETA (8) UNIVERSITARIA
(9) NO SABE-N.R. R
MISSING VALUES VAR013(0,9), VAR016(0), VAR025(9)
CROSSTABS TABLES= VAR016 BY VAR025
STATISTICS ALL

***** GIVEN WORKSPACE ALLOWS FOR 1287 CELLS AND 2 DIMENSIONS FOR CROSSTAB PROBLEM *****

Ejemplo b): Formato de salida estándar producido por la solicitud con modo general de página 151. En el rincón superior izquierdo sale la explicación: COUNT = frecuencia; ROW PCT = porcentaje de fila; COL PCT = porcentaje de columna; TOT PCT = porcentaje del total.

EJEMPLO DEL PROCEDIMIENTO - CROSSTABS -

C8R27R76

PAGE 2

FILE FESALEMB (CREATION DATE = 08R24R76)

VAR016 OCUPACION ENTREVISTADA CROSSTABULATION OF EDUCACION ESPOSO BY VAR025

VAR016	COUNT ROW PCT COL PCT TOT PCT	VAR025										ROW TOTAL
		NINGUNA- ANALFAB 0.	NING.-AL FABETO 1.	PRIMARIA 1-3 2.	PRIMARIA 4-5 3.	PRIM.COM PLETA 4.	SECUNDAR IA 1-3 5.	SECUNDAR IA 4-5 6.	SECUNC.C OMPLETA 7.	UNIVERSI TARIA 8.		
VENEDORES 1.	19 33.6 13.8 3.0	0 C.0 C.0	0 C.0 C.0	21 39.6 16.0 1.9	9 17.0 13.0 1.7	5 9.4 4.7 0.9	2 3.8 3.1 0.4	0 0.0 0.0 0.0	0 C.0 C.0 C.0	0 C.0 C.0 C.0	53 9.8	
OBREROS NO CALIF 2.	20 44.4 17.2 3.7	0 C.0 C.0	3 6.7 21.4 C.6	15 33.3 11.5 2.8	1 2.2 1.4 0.2	5 11.1 4.7 0.9	1 2.2 2.6 0.2	0 0.0 0.0 0.0	0 C.0 C.0 C.0	0 C.0 C.0 C.0	45 8.3	
SERV.DOMESTICO 3.	17 21.3 14.3 3.2	0 C.0 C.0	2 2.5 14.3 C.4	15 18.8 11.5 2.8	12 15.0 17.4 2.2	29 36.3 27.4 5.4	3 3.8 7.7 0.6	1 1.3 6.7 0.2	1 1.3 3.2 C.2	0 C.0 C.0 C.0	80 14.8	
CUENTA PROPIA 4.	52 23.7 53.4 11.5	0 C.0 C.0	9 3.4 64.3 1.7	74 28.2 56.5 13.7	35 13.4 50.7 6.5	51 19.7 48.1 9.5	18 6.9 46.2 3.3	8 3.1 53.3 1.5	0 C.0 C.0 C.0	2 C.0 C.0 C.0	262 48.6	
OBREO CALIF. 5.	0 0.0 0.0 0.0	0 C.0 C.0 C.0	0 C.0 C.0 C.0	2 6.7 1.5 0.4	6 20.0 8.7 1.1	11 36.7 10.4 2.0	7 23.3 17.9 1.3	2 6.7 13.3 0.4	2 C.0 C.0 C.0	0 C.0 C.0 C.0	30 5.6	
EMPLEADO-1 6.	11 0.9 0.2 0.2	0 C.0 C.0 C.0	0 C.0 C.0 C.0	33 33.3 26.3 C.6	11 11.4 1.4 0.2	11 33.3 3.9 0.2	2 3.1 3.1 0.4	0 0.0 0.0 0.0	1 1.1 3.2 C.2	0 C.0 C.0 C.0	9 1.7	
EMPLEADO-2 7.	0 0.0 0.0 0.0	0 C.0 C.0 C.0	0 C.0 C.0 C.0	1 1.8 0.8 C.2	5 8.9 7.2 0.9	4 7.1 3.8 0.7	6 10.7 15.4 1.1	4 7.1 26.7 0.7	25 44.6 75.8 4.6	11 15.6 28.8 2.0	56 10.4	
COLUMN TOTAL	116 21.5	14 2.6	131 24.3	69 12.8	106 19.7	39 7.2	15 2.8	33 6.1	16 3.0	535 100.0		

EJEMPLO DEL PROCEDIMIENTO - CROSSTABS -

C8R27R76

PAGE 3

FILE FESALEMB (CREATION DATE = 08R24R76)

VAR016 OCUPACION ENTREVISTADA CROSSTABULATION OF EDUCACION ESPOSO BY VAR025

VAR016	COUNT ROW PCT COL PCT TOT PCT	VAR025										ROW TOTAL
		NINGUNA- ANALFAB 0.	NING.-AL FABETO 1.	PRIMARIA 1-3 2.	PRIMARIA 4-5 3.	PRIM.COM PLETA 4.	SECUNDAR IA 1-3 5.	SECUNDAR IA 4-5 6.	SECUNC.C OMPLETA 7.	UNIVERSI TARIA 8.		
PEQ.EMPRESARIO 8.	0 0.0 0.0 0.0	0 C.0 C.0 C.0	0 C.0 C.0 C.0	0 0.0 0.0 C.0	0 0.0 0.0 0.0	0 0.0 0.0 0.0	0 0.0 0.0 0.0	0 0.0 0.0 0.0	33 33.3 3.2 C.2	66 12.5 C.4	3 0.6	
JEFES ADMINIST. 9.	0 0.0 0.0 0.0	0 C.0 C.0 C.0	0 C.0 C.0 C.0	0 0.0 0.0 C.0	0 0.0 0.0 C.0	0 0.0 0.0 0.0	0 0.0 0.0 0.0	0 0.0 0.0 0.0	100 100.0 3.2 C.2	0 C.0 C.0 C.0	1 0.2	
COLUMN TOTAL	116 21.5	14 2.6	131 24.3	69 12.8	106 19.7	39 7.2	15 2.8	33 6.1	16 3.0	535 100.0		

CHI SQUARE = 406.23371 WITH 64 DEGREES OF FREEDOM SIGNIFICANCE = 0.0
 CRAMER'S V = 0.30694
 CONTINGENCY COEFFICIENT = 0.65557
 LAMBDA (ASYMMETRIC) = 0.11191 WITH VAR016 DEPENDENT. = 0.13480 WITH VAR025 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.12255
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.17940 WITH VAR016 DEPENDENT. = 0.14725 WITH VAR025 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.16176
 KENDALL'S TAU B = 0.36375 SIGNIFICANCE = 0.0
 KENDALL'S TAU C = 0.31399 SIGNIFICANCE = 0.0
 GAMMA = 0.46519
 SOMERS'S D (ASYMMETRIC) = 0.33699 WITH VAR016 DEPENDENT. = 0.39264 WITH VAR025 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.36269
 ETA = 0.61613 WITH VAR016 DEPENDENT. = 0.62332 WITH VAR025 DEPENDENT.
 NUMBER OF MISSING OBSERVATIONS = 1468

Ejemplo c): Solicitud de cruces con opciones 3,4,5, supresión de todos los porcentajes y la salida producida.

EJEMPLO DEL PROCEDIMIENTO - CROSSTABS -

08R27A76

PAGE 4

CROSSTABS TABLES VAR016 BY VAR025
 OPTIONS 3,4,5
 STATISTICS ALL

EJEMPLO DEL PROCEDIMIENTO - CROSSTABS -

08R27A76

PAGE 5

FILE PESALEMB (CREATION DATE = 08R24R76)

***** C R O S S T A B U L A T I O N O F *****
 VAR016 OCUPACION ENTREVISTADA BY VAR025 ECUCACION ESPCSG
 ***** PAGE 1 OF 1

VAR016	COUNT	VAR025										ROW TOTAL
		NINGUNA-ANALFAB	NING.-AL FABETO	PRIMARIA 1-3	PRIMARIA 4-5	PRIM.COM PLETA	SECUNCAR IA 1-3	SECUNCAR IA 4-5	SECUN.C OMPLETA	UNIVERSI TARIA	8.	
VENEDORES	1.	14	0	21	9	5	2	0	0	0	0	53
OBBEROS NO CALIF	2.	20	3	15	1	5	1	0	0	0	0	49
SERV.DOMESTICO	3.	17	2	15	12	29	3	1	1	0	0	80
CUENTA PROPIA	4.	62	9	74	35	51	18	8	2	3	0	262
OBBERO CALIF.	5.	0	0	2	6	11	7	2	2	0	0	30
EMPLEADO-1	6.	1	0	3	1	1	2	0	1	0	0	9
EMPLEADO-2	7.	0	0	1	5	4	6	4	25	11	0	56
PEO.EMPRESARIO	8.	0	0	0	0	0	0	0	1	2	0	3
JEFES ADMINIST.	9.	0	0	0	0	0	0	0	1	0	0	1
COLUMN TOTAL		116	14	131	69	106	39	15	33	16	0	539
		21.5	2.6	24.3	12.8	19.7	7.2	2.8	6.1	3.0	0	100.0

CHI SQUARE = 406.23271 WITH 64 DEGREES OF FREEDOM SIGNIFICANCE = 0.0
 CRAMER'S V = 0.30694
 CONTINGENCY COEFFICIENT = 0.65557
 LAMBDA (ASYMMETRIC) = 0.11191 WITH VAR016 DEPENDENT. = 0.13480 WITH VAR025 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.12255
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.17940 WITH VAR016 DEPENDENT. = 0.14729 WITH VAR025 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.16176
 KENDALL'S TAU B = 0.36375 SIGNIFICANCE = 0.0
 KENDALL'S TAU C = 0.31399 SIGNIFICANCE = 0.0
 GAMMA = 0.46519
 SOMERS'S D (ASYMMETRIC) = 0.33659 WITH VAR016 DEPENDENT. = 0.39264 WITH VAR025 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.36269
 ETA = 0.61613 WITH VAR016 DEPENDENT. = 0.62332 WITH VAR025 DEPENDENT.
 NUMBER OF MISSING OBSERVATIONS = 1468

Ejemplo d): Modo INTEGER con opciones 3,7,9: 3 = elimina porcentajes de fila; 7 = incluye los missing en las tablas -véase última fila y última columna de los cruces con la frecuencia marcada con la letra "M"; 9 = que produzca un índice.

En este ejemplo se muestra la salida de un cruce tridimensional o sea con una variable de control.

EJEMPLO DEL PROCEDIMIENTO - CROSSTABS -

08R27R76

PAGE

CROSSTABS VARIABLES= VAR004(1,3),VAR015,VAR016,VAR025(0,9), VARC13(0,99) R
 TABLES= VAR016 BY VAR025 R VAR016 BY VAR025 BY VAR004
 3,7,9
 STATISTICS ALL

EJEMPLO DEL PROCEDIMIENTO - CROSSTABS -

08R27R76

PAGE 7

FILE FESAEMB (CREATION DATE = 08R24R76)

***** C R O S S T A B U L A T I O N C F E D U C A C I O N E S P C S C *****
 VAR016 O C U P A C I O N E N T R E V I S T A D A B Y V A R 0 2 5 E D U C A C I O N E S P C S C
 ***** PAGE 1 CF 1 *****

VAR016	COUNT COL PCT TOT PCT	VAR025												ROW TOTAL						
		NINGUNA- ANALFAB	NINGUN- FABETO	AL 1-3	PRIMARIA 2	PRIMARIA 4-5	3	PRIM.COM PLETA	4	SECUNAR IA 1-3	5	SECUNAR IA 4-5	6		SECUNAR C MPLETA	7	UNIVERSI TARIA	8	NC SABA- N.R.	9
VENEDORES	1	16 13.8 3.C	0 C.O C.C	21 16.0 3.9	9 13.C 1.7	5 4.7 0.9	2 5.1 0.4	0 0.C 0.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	53 5.8
OBROEROS NO CALIF	2	20 17.2 3.7	3 21.4 C.6	15 11.5 2.8	1 1.4 C.2	5 4.7 0.9	1 2.6 0.2	0 0.C 0.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	45 8.3
SERV.DOMESTICO	3	14 14.7 3.2	2 14.2 C.4	15 11.5 2.8	2 17.4 2.2	29 27.4 3.4	3 7.3 0.6	1 6.1 0.2	1 2.1 C.2	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	14 14.6
CUENTA PROPIA	4	62 53.4 11.5	9 64.3 1.7	74 56.5 13.7	35 50.7 6.5	51 48.1 9.5	18 46.2 3.3	8 53.3 1.5	2 6.7 C.4	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	262 48.6
OBROERO CALIF.	5	0 0.C 0.C	0 C.O C.C	2 1.9 C.4	6 8.7 1.1	11 10.4 2.0	7 17.9 1.3	2 13.2 0.4	2 6.1 C.4	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	30 5.6
EMPLEADO-1	6	0 0.2 0.C	0 C.O C.C	3 2.9 C.2	1 1.4 C.2	1 0.9 C.2	4 5.1 0.4	0 0.C 0.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	1.5 1.5
EMPLEADO-2	7	0 0.C 0.C	0 C.O C.C	1 C.8 C.2	5 7.2 C.9	4 3.8 C.7	6 15.4 1.1	4 26.7 C.7	2 75.2 4.6	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	10.4 10.4
PEO.EMPRESARIO	8	0 0.C 0.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	3 C.6
JEFES ADMINIST.	9	0 0.C 0.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	0 C.O C.C	3 C.6
IGNORADO	0	6 6.C 0.C	5 5.M C.O	4 413M C.O	7 117M C.O	15 150M C.O	6 69M C.O	14 14M C.O	34 34M C.O	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	0 C.C C.C	146 C.C
COLUMN TOTAL		116 21.5	14 2.6	131 24.3	65 12.8	106 19.7	39 7.2	15 2.8	33 6.1	16 3.C	0 C.C	0 C.C	0 C.C	0 C.C	0 C.C	0 C.C	0 C.C	0 C.C	0 C.C	835 100.C

RAW CHI SQUARE = 406.23071 WITH 64 DEGREES OF FRECCM. SIGNIFICANCE = 0.0
 CRAMER'S V = 0.30694
 CONTINGENCY COEFFICIENT = 0.65557
 LAMBDA (ASYMMETRIC) = 0.11191 WITH VAR016 DEPENDENT. = 0.13480 WITH VAR025 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.12655
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.17940 WITH VAR016 DEPENDENT. = 0.14725 WITH VAR025 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.16176
 KENDALL'S TAU B = 0.36375. SIGNIFICANCE = 0.0
 KENDALL'S TAU C = 0.31395. SIGNIFICANCE = 0.0
 GAMMA2 = 0.46519
 SOMER'S D (ASYMMETRIC) = 0.33659 WITH VAR016 DEPENDENT. = 0.39264 WITH VAR025 DEPENDENT.
 SOMER'S D (SYMMETRIC) = 0.36265
 ETA = 0.6113 WITH VAR016 DEPENDENT. = 0.62332 WITH VAR025 DEPENDENT.
 NUMBER OF MISSING OBSERVATIONS = 1468

EJEMPLO DEL PROCEDIMIENTO - CROSSTABS -
FILE FESALEMB (CREATION DATE = 08R24R76)

CR27R76 PAGE 8

VAR016 OCUPACION ENTREVISTADA CROSS TABULATION CF
CONTROLLING FOR: BY VAR025 EDUCACION ESPCS
VAR004 AREA VALUE.. 1 RURAL PAGE 1 CF 1

VAR016	COUNT COL PCT TOT PCT	VAR025											RCH TOTAL
		NINGUNA- ANALFAB C	NINGUN- FABETO 1	AL 1-2 2	PRIMARIA 1-3 2	PRIMARIA 4-5 3	PRIM.COM PLETA 4	SECUNCAR IA 1-3 5	SECUNCAR IA 4-5 6	SECLNC.C CMPLETA 7	LAUNERSI TARIA 8	NC SARE- N.R. 5	
VENDEDORES	1	10.8 4.6	C.O C.O	11.7 4.0	C.O C.O	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	C.C C.C	C.C C.C	C.C C.C	18 6.8
OBROS NO CALIF	2	15 20.3	C.O C.O	13 22.0	C.O C.O	6.3 11.1	18.8 1.7	0.0 0.0	0.0 0.0	C.C C.C	C.C C.C	C.C C.C	18.4 16.4
SERV.DOMESTICO	3	11 14.9	C.O C.O	3 5.1	C.O C.O	18.3 1.7	18.8 1.7	0.0 0.0	0.0 0.0	C.C C.C	C.C C.C	C.C C.C	20 11.5
CUENTA PROPIA	4	4 54.1	C.C 23.0	3 20.1	C.C C.C	35 59.3	12 6.9	2 5.2	66.7 1.1	0.0 0.0	100.0 1.1	100.0 C.C	104 56.8
OBRO CALIF.	5	C.O C.O	C.O C.O	C.O C.O	C.O C.O	C.O C.O	6.3 0.6	33.3 0.6	0.0 0.0	C.C C.C	C.C C.C	C.C C.C	1 0.5
EMPLEADO-1	6	C.O C.O	C.O C.O	1 C.6	C.O C.O	C.O C.O	C.O C.O	0.0 0.0	0.0 0.0	C.C C.C	C.C C.C	C.C C.C	C 0.5
IGNORADO	0	511M 0.0	43M C.O	291M C.O	C.O C.O	48M C.O	C.O C.O	43M 0.0	12M 0.0	0.0 0.0	C.C C.C	C.C C.C	551M C.C
COLUMN TOTAL		74 42.5	1.7	33.9	16 9.2	16 9.2	1.7	0.0	1.1	C.C	C.C	100.0	

RAW CHI SQUARE = 50.60884 WITH 35 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0426
 CRAMER'S V = 0.24116
 CONTINGENCY COEFFICIENT = 0.47468
 LAMBDA (ASYMMETRIC) = 0.0 WITH VAR016 DEPENDENT. = 0.02000 WITH VAR025 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.01176
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.08668 WITH VAR016 DEPENDENT. = 0.02232 WITH VAR025 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.07885
 KENDALL'S TAU B = 0.19293. SIGNIFICANCE = 0.0104
 KENDALL'S TAU C = 0.11661. SIGNIFICANCE = 0.0104
 CONDITIONAL GAMMA = 0.24415
 SOMERS' D (ASYMMETRIC) = 0.16155 WITH VAR016 DEPENDENT. = 0.16523 WITH VAR025 DEPENDENT.
 SOMERS' D (SYMMETRIC) = 0.15247
 ETA = 0.24790 WITH VAR016 DEPENDENT. = 0.25113 WITH VAR025 DEPENDENT.

EJEMPLO DEL PROCEDIMIENTO - CROSSTABS -
FILE FESALEMB (CREATION DATE = 08R24R76)

CR27R76 PAGE 9

VAR016 OCUPACION ENTREVISTADA CROSS TABULATION CF
CONTROLLING FOR: BY VAR025 EDUCACION ESPCS
VAR004 AREA VALUE.. 2 URBANA PAGE 1 CF 1

VAR016	COUNT COL PCT TOT PCT	VAR025											RCH TOTAL
		NINGUNA- ANALFAB C	NINGUN- FABETO 1	AL 1-2 2	PRIMARIA 1-3 2	PRIMARIA 4-5 3	PRIM.COM PLETA 4	SECUNCAR IA 1-3 5	SECUNCAR IA 4-5 6	SECLNC.C CMPLETA 7	LAUNERSI TARIA 8	NC SARE- N.R. 5	
VENDEDORES	1	17.6 3.0	C.O C.O	16.8 4.1	C.O C.O	8.2 1.0	4.3 1.0	0.0 0.0	0.0 0.0	C.C C.C	C.C C.C	C.C C.C	18 9.1
OBROS NO CALIF	2	14.3 2.5	50.3 1.5	4.2 1.0	C.O C.O	0.0 0.0	2.1 0.5	0.0 0.0	0.0 0.0	C.C C.C	C.C C.C	C.C C.C	11 5.1
SERV.DOMESTICO	3	14.3 2.5	16.3 C.3	20.0 5.1	25.6 10.0	21.0 5.3	6.3 1.5	0.0 0.0	0.0 0.0	C.C C.C	C.C C.C	C.C C.C	24 17.5
CUENTA PROPIA	4	19 54.3	2 9.6	26 53.1	11 13.2	25 5.6	25 12.7	6 3.0	80.4 2.0	C.C C.C	C.C C.C	C.C C.C	63 47.8
OBRO CALIF.	5	C.O C.O	C.O C.O	2 C.5	C.O C.O	4 0.5	12.8 3.0	26.7 2.0	0.0 0.0	C.C C.C	C.C C.C	C.C C.C	13 6.8
EMPLEADO-1	6	C.O C.O	C.O C.O	2 C.5	C.O C.O	4 0.5	C.O C.O	0.0 0.0	0.0 0.0	C.C C.C	C.C C.C	C.C C.C	1 0.5
EMPLEADO-2	7	C.O C.O	C.O C.O	1 C.5	12.3 1.5	3 0.6	6.4 1.5	26.7 2.0	20.1 0.5	4.1 0.6	100.0 1.1	C.C C.C	24 13.2
IGNORADO	0	75M 0.0	11M C.O	95M C.O	36M C.O	58M 0.0	23M 0.0	0.0 0.0	0.0 0.0	C.C C.C	C.C C.C	C.C C.C	214M C.C
COLUMN TOTAL		38 17.8	3.0	24.9	12.2	47 23.9	15 7.6	2.3	6.6	C.C	C.C	100.0	

RAW CHI SQUARE = 158.07175 WITH 48 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.36569
 CONTINGENCY COEFFICIENT = 0.66722
 LAMBDA (ASYMMETRIC) = 0.14423 WITH VAR016 DEPENDENT. = 0.12162 WITH VAR025 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.13095
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.21268 WITH VAR016 DEPENDENT. = 0.17221 WITH VAR025 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.19032
 KENDALL'S TAU B = 0.41093. SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.36711. SIGNIFICANCE = 0.0000
 CONDITIONAL GAMMA = 0.32334
 SOMERS' D (ASYMMETRIC) = 0.38241 WITH VAR016 DEPENDENT. = 0.44071 WITH VAR025 DEPENDENT.
 SOMERS' D (SYMMETRIC) = 0.40484
 ETA = 0.61301 WITH VAR016 DEPENDENT. = 0.62067 WITH VAR025 DEPENDENT.

EJEMPLO DEL PROCEDIMIENTO - CROSSTABS -

CBR27676

PAGE 1C

FILE FESALEMB (CREATION DATE = 08R24R76)

VAR016 OCCUPACION ENTREVISTADA BY VAR025 EDUCACION ESPSC
 CONTROLLING FOR VAR004 AREA VALUE.. 3 METROCLITANA

PAGE 1 CF 1

VAR016	COUNT		VAR025										RCW TOTAL
	COL TOT	PCT PCT	NINGUNA- ANALFAB	NING.-AL FABETC	PRIMARIA 1-3	PRIMARIA 4-5	PRIM.COM PLETA	SECUNCAR IA 1-3	SECUNCAR IA 4-5	SECLAC.C CMPLTA	UNIVERSI TARIA		
VENEDORES	1	28.2 11.2	C.C C.C	C.C C.C	26.6 3.6	24.7 4.1	7.3 1.8	9.2 1.2	0.0 0.0	C.C C.C	C.C C.C	20 11.9	
OBREROS NU CALIF	2	0.0 0.0	C.C C.C	C.C C.C	0.0 0.0	0.0 0.0	2.3 0.6	4.8 0.6	0.0 0.0	C.C C.C	C.C C.C	1.2	
SERV.DOMESTICO	3	14.3 0.6	2C.1 C.6	8.2 1.2	10.3 1.8	37.2 9.5	9.5 1.2	10.1 0.6	C.C C.C	C.C C.C	C.C C.C	15.2 6.5	
CUENTA PROPIA	4	42.3 1.8	BC.4 2.4	56.5 7.7	41.4 7.1	39.5 10.1	4.7 0.6	40.4 2.4	C.C C.C	C.C C.C	16.2 1.2	38.7	
OBRERO CALIF.	5	0.0 0.0	C.C C.C	C.C C.C	4.3 0.6	17.2 3.0	9.3 2.4	9.5 1.2	20.0 1.2	C.C C.C	C.C C.C	15 8.9	
EMPLEADO-1	6	14.3 0.6	C.C C.C	C.C C.C	4.3 0.6	C.C C.C	2.3 0.6	9.5 1.2	C.C C.C	C.C C.C	C.C C.C	3.6	
EMPLEACO-2	7	0.0 0.0	C.C C.C	C.C C.C	0.0 0.0	6.2 1.2	2.3 0.6	9.5 1.2	30.3 1.8	77.6 8.1	66.8 7.6	17.9	
PEQ.EMPRESARIO	8	0.0 0.0	C.C C.C	C.C C.C	0.0 0.0	C.C C.C	C.C C.C	0.0 0.0	C.C C.C	C.C C.C	16.2 1.2	1.8	
JEFES ADMINIST.	9	C.C C.C	C.C C.C	C.C C.C	C.C C.C	C.C C.C	C.C C.C	C.C C.C	C.C C.C	C.C C.C	C.C C.C	C.C C.C	
IGNORADO	0	14M C.C	5M C.C	27M C.C	33M C.C	49M C.C	34M C.C	5M C.C	15M C.C	7M C.C	157M C.C		
COLUMN TOTAL		4.2	3.0	13.7	17.3	25.6	12.5	6.0	10.7	7.1	100.0		

RAW CHI SQUARE = 160.50433 WITH 64 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.34558
 CONTINGENCY COEFFICIENT = 0.65865
 LAMBDA (ASYMMETRIC) = 0.19417 WITH VAR016 DEPENDENT. = 0.17600 WITH VAR025 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.18421
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.26010 WITH VAR016 DEPENDENT. = 0.22012 WITH VAR025 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.23845
 KENDALL'S TAU B = 0.41720. SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.37930. SIGNIFICANCE = 0.0000
 CONDITIONAL GAMMA = 0.50307
 SOMERS'S D (ASYMMETRIC) = 0.39776 WITH VAR016 DEPENDENT. = 0.43756 WITH VAR025 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.41672
 ETA = 0.68681 WITH VAR016 DEPENDENT. = 0.67592 WITH VAR025 DEPENDENT.

SUMMARY GAMMAS FOR CROSSTABLATION OF VAR016 BY VAR025
 ZERC-ORDER GAMMA = 0.46515
 FIRST-ORDER PARTIAL GAMMA = 0.45455

NUMBER OF MISSING OBSERVATIONS = 1468

El índice producido por opción 9.

EJEMPLO DEL PROCEDIMIENTO - CROSSTABS -
 FILE FESALEMB (CREATION DATE = 08R24R76)

CBR27676

PAGE 11

CROSSTABLATION INDEX

PAGE	TABLE
7	VAR016 BY VAR025
8	VAR016 BY VAR025 BY VAR004

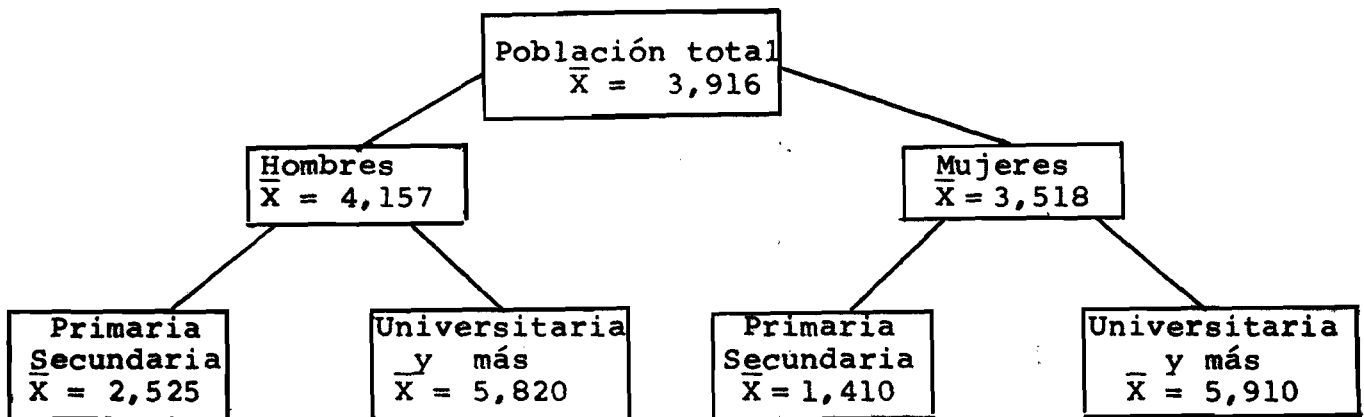
XIII. BREAKDOWN: Descripción de subpoblaciones

XIII.1 El subprograma BREAKDOWN

Dada una variable dependiente para la cual la media es una buena medida de la tendencia central, el subprograma BREAKDOWN produce para esta variable (dependiente) la suma, media, varianza y desviación estándar para los subconjuntos del universo, delimitados por las variables independientes.

Por ejemplo si del universo en estudio tomamos como variable dependiente el INGRESO y deseamos saber la media del ingreso por sexo y educación, Así para la subpoblación con SEXO = hombres la media del ingreso será una y para la subpoblación SEXO = mujeres, la media del ingreso será otra.

Así si llamamos de \bar{X} = media del ingreso, la estructura de los resultados que produce el BREAKDOWN es la siguiente; por sexo y educación (los valores de \bar{X} no son reales).



A pesar de que el del subprograma BREAKDOWN sólo permite procesar variables numéricas, posee dos modos de operación:

GENERAL = acepta variables numéricas enteras y/o con decimales como variables independientes.

INTEGER = sólo acepta variables numéricas enteras como variables independientes.

En ambos modos la variable dependiente puede ser discreta o continua.

Resumen del BREAKDOWN:

a) Produce en forma estándar:

- suma
- media
- varianza
- desviación estándar

para la población total y subpoblaciones.

b) Permite en forma opcional:

- análisis de varianza en una dirección
- test de linealidad

c) Permite un máximo de cinco variables independientes

d) La forma de especificación del BREAKDOWN es idéntica a la del CROSSTABS, inclusive el modo integer es seleccionado si se especifica la parte "VARIABLES = ".

XIII.2 BREAKDOWN: modo GENERAL

Formato General

<u>1</u>	<u>16</u>
BREAKDOWN	TABLES= { variable o lista de vars dependientes } BY { variable o lista de vars independientes } BY....BY { variable o lista de vars independientes } /
	{ variable o lista de vars dependientes } BY { variable o lista de vars independientes } BY....BY { variable o lista de vars independientes } /....

Nótese que la variable o lista de variables especificada a la izquierda del primer "BY" delimita cuál o cuáles son las variables dependientes para la cual se calculará la suma, media, varianza y desviación estándar.

Para el ejemplo que se muestra la estructura en el párrafo XIII.1, la tarjeta de procedimiento sería:

```

1           16
BREAKDOWN  TABLES = INGRESO BY SEXO BY EDUC
    
```

Los siguientes ejemplos ilustran la forma de especificación para el modo general; las variables dependientes están subrayadas.

```

1           16
BREAKDOWN  TABLES = VARX, VARZ BY VAR1 TO VAR10/
                VARX, VARZ BY VAR1 TO VAR5 BY VAR10

BREAKDOWN  TABLES = VARX BY VARZ BY VAR1 BY VAR2/
                VARZ BY VAR1 BY VAR2 BY VAR3
    
```

XIII.3 BREAKDOWN: modo INTEGER

Formato General

```

1           16
BREAKDOWN    VARIABLES= { variable o } (min,max) , . . . . , { variable o } (min,max) /
                    { lista de vars }
                    { lista de vars }

                TABLES= { variable o } BY { variable o } BY . . . . BY { variable o }
                    { lista de vars } { lista de vars } { lista de vars }
                    { dependientes } { independientes } { independientes }

                [ { variable o } BY { variable o } BY . . . . BY { variable o } ] [ . . . . ]
                    { lista de vars } { lista de vars } { lista de vars }
                    { dependientes } { independientes } { independientes }
    
```

Al igual que el procedimiento CROSSTABS, modo Integer, para el BREAKDOWN tenemos:

- a) En la parte "VARIABLES = " el uso de la convención "TO" se refiere al orden en que las variables están definidas al paso que en la parte "TABLES = " el uso de la convención "TO" se refiere sólo a las variables declaradas en "VARIABLES =" y al orden en que fueron declaradas.
- b) En la parte "TABLES = " sólo se puede hacer referencia a variables que están declaradas en VARIABLES =. Esto significa que inclusive las variables dependientes deben ser declaradas en VARIABLES =.

Especificación de rangos para variables dependientes

Al utilizar el modo integer el usuario debe especificar en "VARIABLES =" la(s) variable(s) dependiente(s) y su(s) respectivo(s) rango(s). Como generalmente las variables dependientes son continuas y quizás con un rango no conocido, el SPSS permite que para dichas variables se utilice las palabras claves LOWEST, HIGHEST (LO,HI) para la especificación del rango. No se puede por ende utilizar esta facilidad para las variables independientes.

Ejemplo:

```
1          16
BREAKDOWN  VARIABLES = INGRESO (LOWEST, HIGHEST)
```

El siguiente ejemplo ilustra la forma de especificación para el modo Integer.

```
1          16
BREAKDOWN  VARIABLES = INGRESO (LO,HI), SEXO(1,2), EDUC(0,9),
                VAR01 TO VAR10(0,7)/
                TABLES = INGRESO BY SEXO BY EDUC/ INGRESO, VAR01
                BY SEXO, EDUC BY VAR05.
```

XIII.4 Estadísticas disponibles para BREAKDOWN

El subprograma BREAKDOWN calcula automáticamente la suma media, varianza, desviación estándar y número de casos para cada subpoblación. Además de estos estadísticos, el usuario puede solicitar por medio de STATISTICS, dos cálculos estadísticos adicionales:

ESTADISTICO 1: provoca el cálculo e impresión de análisis de varianza en una dirección y el estadístico ETA-cuadrado.

ESTADISTICO 2: cálculo de test de linealidad entre la var de pendiente y la primera var independiente. Incluye el cálculo del F estadístico, del coeficiente de correlación (R) de PEARSON y el R^2 .

Si se desea el estadístico 2, el 1 también debe ser solicitado.

Los estadísticos deben ser solicitados por medio de la tarjeta STATISTICS que debe ser puesta después de la tarjeta OPTIONS (si se utiliza).

El usuario debe tener cuidado en el uso de la solicitud de los estadísticos ya que tanto el análisis de varianza como el test de linealidad se hacen en una dirección. Así, si se especifica un BREAKDOWN con dos BY, los estadísticos (si son solicitados) sólo tomarán en cuenta las variables ligadas por el primer BY.

Ejemplo:

1	16
BREAKDOWN	TABLES = INGRESO BY SEXO BY EDUC
STATISTICS	ALL

En este caso, el análisis de varianza y test de linealidad sólo tomará en cuenta INGRESO BY SEXO y sería exactamente lo mismo que un BREAKDOWN sin el "BY EDUC"

XIII.5 Las opciones disponibles para BREAKDOWN

El subprograma BREAKDOWN posee tres formas de manipular los casos con valores missing:

Opción estándar: eliminación por tabla: un caso es omitido del cálculo de la tabla cuando el valor de cualquiera de las variables involucradas trae un valor missing.

En otras palabras, un caso entra al cálculo de la tabla si los valores de todas las variables involucradas son no-missing. Esta opción es utilizada a menos que se utilice opción 1 ó 2.

OPCION 1: inclusión de los missing. Esta opción informa al sistema para que ignore la declaración de MISSING y por lo tanto todos los casos serán tratados como casos válidos.

OPCION 2: Exclusión de missing sólo para variable dependiente: esta opción provoca la exclusión de los casos en que la variable dependiente trae un valor declarado como MISSING.

OPCION 3: elimina impresión de los labels.

OPCION 4: salida en forma de árbol. Esta opción está disponible para el modo GENERAL. Véase en el párrafo XIII. 8, de ejemplo de salida, el formato de impresión cuando se utiliza esta opción.

La tarjeta OPTIONS debe preceder a la de STATISTICS: las distintas posibilidades se ilustran a continuación.

1 16

BREAKDOWN
OPTIONS

BREAKDOWN
OPTIONS
STATISTICS

BREAKDOWN
STATISTICS

XIII.6 Limitaciones para el BREAKDOWN

a) Para el modo GENERAL

- 1) Un máximo de 200 variables pueden estar nombradas o involucradas en una solicitud de BREAKDOWN
- 2) El máximo número de tablas que puede ser solicitado es de 250.
- 3) La dimensión máxima es 6, o sea se permite hasta 5 "BY".
- 4) Un máximo de 30 listas de tablas pueden ser solicitadas. Una lista está delimitada por "/", a excepción de la primera y última lista.
- 5) Requerimientos de espacio de memoria:

$$\text{WORKSPACE} = 4 * (\text{NCELLS} + 1) * (\text{MAXDIMS} + 4)$$

NCELLS = es el número total de celdas, contabilizado en todas las tablas. Una celda está definida por una combinación única de valores de las variables independientes.

MAXDIMS = el número máximo de dimensiones obtenido entre todos los juegos de tablas.

WORKSPACE = espacio de memoria en bytes

Ejemplo:

1	16
BREAKDOWN	TABLES = INGRESO BY SEXO BY EDUC / INGRESO BY EDUC BY ZONA BY SEXO

Conociendo que:

<u>Variable</u>	<u>Códigos</u>
SEXO	1, 2
ZONA	1, 2
EDUC	1, 2, 3, 4, 5, 6, 7, 8, 9

tenemos:

$$\text{NCELLS} = (2 * 9) + (9 * 2 * 2) = 54 \text{ celdas}$$

$$\text{MAXDIMS} = 4 \text{ (de la 2a. lista de tablas = 3 "BY" + 1)}$$

$$\text{WORKSPACE} = 4 * (54 + 1) * (4 + 4) = 1760 \text{ bytes}$$

- 6) El máximo número de value labels que pueden ser impresos por una solicitud de BREAKDOWN es de 200. Si se excede esta limitación, no serán impresos los value labels que excedieron.

b) Para el modo INTEGER

- 1) Un máximo de 100 variables pueden ser nombradas en la parte VARIABLES =.
- 2) El número de variables especificados en "TABLES =" no puede exceder de 100.
- 3) La dimensión máxima es 6: una variable dependiente y 5 independiente.
- 4) Máximo de 30 listas de tablas es permitido en "TABLES = ".

5) Requerimientos de memoria.

WORKSPACE = 12 * NCELLS + 20 * MAXLABS

NCELLS = número total de celdas para el total de tablas.

MAXLABS = el máximo número de value labels requerido; seleccionado de la tabla que requiere el máximo y no la sumatoria. Si el **WORKSPACE** no es suficiente para almacenar los value labels, sencillamente estos no serán impresos.

Ejemplo:

1	16
BREAKDOWN	VARIABLES = INGRESO (LO,HI), SEXO (1,2), EDUC (1,9), ZONA (1,2) / TABLES = INGRESO BY SEXO BY EDUC / INGRESO BY EDUC BY ZONA BY SEXO

NCELLS = 54 celdas (igual al ejemplo del modo **GENERAL**)

MAXLABS = 9 * 2 * 2 = 36 (para la 2° tabla).

Luego:

WORKSPACE = 12 * 54 + 20 * 36 = 1368 bytes

Nótese que el modo general requiere 1760 bytes

XIII.7 BREAKDOWN: impresión en forma tabular-CROSSBREAK

Formato General

1	16
BREAKDOWN	VARIABLES= { variable o lista de vars } (min,max) , . . . , { variable o lista de vars } (min,max) /
	CROSSBREAK= { variable o lista de vars dependientes } BY ... BY { variable o lista de vars independientes }
	[{ variable o lista de vars dependientes } BY ... BY { variable o lista de vars independientes }] [...]

En el formato la única diferencia con el modo Integer es que se especifica "CROSSBREAK =" en lugar de "TABLES = "

El formato de impresión para CROSSBREAK es en forma tabular similar al formato del CROSSTABS a diferencia de que en vez de imprimir la frecuencia los porcentajes de fila, columna y total, imprime la suma, media, desviación estándar, varianza para la variable dependiente para la subpoblación especificada por las variables independientes.

A. OPCIONES PARA CROSSBREAK

OPCION 1,2,3: iguales a las ya descritas para BREAKDOWN. Nótese que si se usa opción 2, los missing impresos en la tabla estarán marcados con la letra "M", pero no influyen en los cálculos.

OPCION 4: No está disponible para el CROSSBREAK.

OPCION 5: Elimina la impresión de las celdas de frecuencias.

OPCION 6: Elimina impresión de las sumas.

OPCION 7: Elimina la impresión de la desviación estándar.

OPCION 8: Elimina la impresión de los value labels, pero no de los VAR LABELS.

B. ESTADISTICOS PARA CROSSBREAK

Los estadísticos 1,2 del BREAKDOWN (general, integer) no están disponibles para el CROSSBREAK.

Los estadísticos disponibles para el CROSSBREAK son los de medidas de relación entre las variables de fila y columna, o sea sólo consideran las variables independientes, y son prácticamente los mismos estadísticos del CROSSTABS.

<u>Número</u>	<u>Estadístico</u>
3	Chi-cuadrado
4	Phi para tablas de 2 x 2; V de CRAMES para otras tablas.
5	Coefficiente de contingencia
6	LAm bolo (simétrico y asimétrico)
7	Coefficiente de incertidumbre
8	TAU B de Kendall
9	TAU C de Kendall
10	GAMMA
11	D de Sommers (simétrico, asimétrico)
12	ETA

C. LIMITACIONES PARA EL CROSSBREAK

Las limitaciones son las mismas que las del modo INTEGER anteriormente descritas.

XIII.8 EJEMPLOS DE SALIDA DEL BREAKDOWN

Ejemplo a): salida estándar del subprograma BREAKDOWN sin OPTIONS, con STATISTICS utilizando el modo general.

CELADE - Banco de Datos

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES SPSSM - VERSION 6.01 09R03R76 PAGE 1

STA

RUN NAME GET FILE	EJEMPLO DEL PROCEDIMIENTO - BREAKDOWN - FESALEMB
FILE FESALEMB HAS 81 VARIABLES	
THE SUBFILES ARE..	
NAME	NO OF CASES
FESALEMB	2007
PAGESIZE	80
VAR LABELS	VAR013, EDAD DECLARADA
VALUE LABELS	VAR004 (1) RURAL (2) URBANA (3) METROPOLITANA R VAR015 (1) NO SE APLICA (1) DENTRO HOGAR (2) FUERA HOGAR (3) DENTRO Y FUERA (4) LUGAR N.O. (5) EMPL.DOMESTICA (6) NO RESPONDE R VAR016 (1) IGNORADO (1) VENDEDORES (2) OBREROS NO CALIF (3) SERV.DOMESTICO (4) CUENTA PROPIA (5) OBRERO CALIF. (6) EMPLEADO-1 (7) EMPLEADO-2 (8) PEG.EMPRESARIO (9) JEFS ADMINIST. R VAR025 (1) NINGUNA-ANALFAB (1) NING.-ALFABETO (2) PRIMARIA 1-3 (3) PRIMARIA 4-5 (4) PRIM.COMPLETA (5) SECUNDARIA 1-3 (6) SECUNDARIA 4-5 (7) SECUND.COMPLETA (8) UNIVERSITARIA (9) NO SABE-N.R. R
MISSING VALUES	VAR015(10,9), VAR016(10), VAR025(19)
BREAKDOWN TABLES	VAR013 BY VAR015 BY VAR004
STATISTICS	ALL

CELADE - Banco de Datos

EJEMPLO DEL PROCEDIMIENTO - BREAKDOWN - 09R03R76 PAGE 2

FILE FESALEMB (CREATION DATE = 08R24R76)

CRIERION VARIABLE BROKEN DOWN BY	VAR013 VAR015 BY VAR004	EDAD DECLARADA TRABAJO AREA	DESCRIPTION OF SUBPOPULATIONS			
VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE
FOR ENTIRE POPULATION			17596.0000	32.4052	8.0706	65.1347
VAR015	1.	DENTRO HOGAR	8384.000	32.671	8.541	72.948
VAR004	1.	RURAL	3416.000	33.490	9.044	81.798
VAR004	2.	URBANA	2095.000	34.034	8.218	67.528
VAR004	3.	METROPOLITANA	1973.000	33.441	8.238	67.872
VAR015	2.	FUERA HOGAR	8759.000	31.171	7.464	55.714
VAR004	1.	RURAL	2092.000	32.185	7.433	55.248
VAR004	2.	URBANA	3383.000	31.037	7.880	62.092
VAR004	3.	METROPOLITANA	3284.000	30.492	7.048	49.669
VAR015	3.	DENTRO Y FUERA	453.000	34.846	7.680	58.975
VAR004	1.	RURAL	216.000	30.857	8.513	72.477
VAR004	2.	URBANA	156.000	39.000	1.155	1.333
VAR004	3.	METROPOLITANA	81.000	40.500	4.950	24.500

TOTAL CASES = 2007
MISSING CASES = 1466 OR 72.9 PCT.

Ejemplo b): Análisis de varianza y test de linearidad producido por BREAKDOWN con "STATISTICS ALL". La solicitud de esta salida se muestra en el ejemplo a).

CELADI - Banco de Datos

EJEMPLO DEL PROCEDIMIENTO - BREAKDOWN -

09603R76

PAGE 3

CRITERION VARIABLE VAR013

ANALYSIS OF VARIANCE							
VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	SUM OF SQ	N
VAR013	1.	DENTRO HOGAR	6384.000	33.671	8.541	18091.125	(249)
VAR013	2.	FUERA HOGAR	8759.700	31.171	7.464	15599.937	(281)
VAR013	3.	DENTRO Y FUERA	453.000	34.846	7.680	707.699	(13)
TOTAL			17596.000	32.405	8.071	34398.762	(543)

ANOVA TABLE			
	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE
BETWEEN GROUPS	904.2383	(2)	452.1191
WITHIN GROUPS	34398.7617	(540)	63.7014
TOTAL	35303.0000	(542)	
F = 7.0975 ETA SQRD = 0.0256			

TEST OF LINEARITY			
	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE
REGRESSION	503.6794	(1)	503.6754
DEV FROM LINEARITY	400.5588	(1)	400.5588
F = 6.2881 R = -0.1194 R SQUARED = 0.0143			

Ejemplo c): Salida del modo general con OPTIONS 4: estructura de árbol

CELADE - Banco de Datos

EJEMPLO DEL PROCEDIMIENTO - BREAKDOWN -

09803R76

PAGE 4

DATA TRANSFORMATION DONE UP TO THIS POINT..

NO OF TRANSFORMATIONS	0
NO OF RECORD VALUES	0
NO OF ARITH. OR LOG. OPERATIONS	0
THE AMOUNT OF TRANSPACE REQUIRED IS	0 BYTES

BREAKDOWN	TABLES= VAR013 BY VAR015 BY VAR004
OPTIONS	4
STATISTICS	ALL

***** GIVEN WORKSPACE ALLOWS FOR 857 CELLS AND 2 DIMENSIONS FOR SUBPROGRAM BREAKDOWN *****

CELADE - Banco de Datos

EJEMPLO DEL PROCEDIMIENTO - BREAKDOWN -

09G03R76

PAGE 5

FILE FESALEMB (CREATION DATE = 08R24R76D

----- DESCRIPTION OF SUBPOPULATIONS -----
 CRITERION VARIABLE VAR013 EQUAD OLCCLARADA
 BROKEN DOWN BY VAR015 TRABAJO
 BY VAR004 AREA

FOR ENTIRE POPULATION
 SUM 17596.000
 MEAN 32.405
 STD DEV 8.071
 VARIANCE 65.135
 N (543)

VARIABLE	VAR015	VARIABLE	VAR004
CODE	1.	CODE	1.
DENTRO HOGAR		RURAL	
SUM	8384.000	SUM	3416.000
MEAN	33.671	MEAN	33.490
STD DEV	8.541	STD DEV	9.044
VARIANCE	(72.948	VARIANCE	(81.798
N	(249)	N	(102)
		CODE	2.
		URBANA	
		SUM	2995.000
		MEAN	34.034
		STD DEV	8.218
		VARIANCE	(67.524
		N	(88)
		CODE	3.
		METROPOLITANA	
		SUM	1973.000
		MEAN	33.441
		STD DEV	8.238
		VARIANCE	(67.877
		N	(59)
CODE	2.	CODE	1.
FUERA HOGAR		RURAL	
SUM	8759.000	SUM	2092.000
MEAN	31.171	MEAN	32.185
STD DEV	7.464	STD DEV	7.433
VARIANCE	(55.714	VARIANCE	(55.248
N	(281)	N	(65)
		CODE	2.
		URBANA	
		SUM	3383.000
		MEAN	31.037
		STD DEV	7.860
		VARIANCE	(62.092
		N	(109)
		CODE	3.
		METROPOLITANA	
		SUM	3284.000
		MEAN	30.692
		STD DEV	7.048
		VARIANCE	(49.669
		N	(107)
CODE	3.	CODE	1.
DENTRO Y FUERA		RURAL	
SUM	453.000	SUM	216.000
MEAN	34.846	MEAN	30.857
STD DEV	7.680	STD DEV	8.513
VARIANCE	(58.975	VARIANCE	(72.477
N	(13)	N	(7)

EJEMPLO DEL PROCEDIMIENTO - BREAKDOWN -

09R03R76

PAGE 6

VARIABLE	VAR015	VARIABLE	VAR004
		CODE	2.
		URBANA	
		SUM	156.000
		MEAN	39.000
		STD DEV	1.155
		VARIANCE	(1.333
		N	(4)
		CODE	3.
		METROPOLITANA	
		SUM	81.000
		MEAN	40.500
		STD DEV	4.950
		VARIANCE	(24.500
		N	(2)

TOTAL CASES = 2007
 MISSING CASES = 1464 OR 72.9 PCT.

Ejemplo d): Salida del CROSSBREAK con STATISTICS ALL

EJEMPLD DEL PROCEDIMIENTO - BREAKDOWN -

C9R03R76 PAGE 8

BREAKDOWN VARIABLES= VAR004(1,3), VAR013(0,99), VAR015(0,9) R
 STATISTICS CROSSBREAK= VAR013 BY VAR015 BY VAR004
 ALL

EJEMPLD DEL PROCEDIMIENTO - BREAKDOWN -

09R03R76 PAGE 9

FILE FESALEMB (CREATION DATE = 08R24R76)

..... C R O S S B R E A K D O W N O F
 VAR015 TRABAJO BY VAR004 AREA
 VARIABLE AVERAGED... VAR013 EDAD DECLARADA
 PAGE 1 OF 1

		VAR004			
		RURAL	URBANA	METROPOL (TANA)	ROW TOTAL
VAR015	MEAN COUNT STD DEV	1	2	3	
DENTRO HOGAR	1	33.49 10.2 3416.00 9.04	34.03 8.8 2995.00 8.22	33.44 5.9 1973.00 8.24	33.67 24.9 8384.00 8.54
FUERA HOGAR	2	32.18 6.5 2092.00 7.43	31.04 10.9 3383.00 7.88	30.69 10.7 3284.00 7.05	31.17 28.1 8759.00 7.46
DENTRO Y FUERA	3	30.86 7 216.00 8.51	39.00 4 156.00 1.15	40.50 2 81.00 4.95	34.85 13 453.00 7.68
COLUMN TOTAL		32.90 174 5724.00 8.44	32.51 201 6534.00 8.12	31.77 168 5338.00 7.61	32.41 543 17596.00 8.07

RAW CHI SQUARE = 25.51003 WITH 4 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.15326
 CONTINGENCY COEFFICIENT = 0.21183
 LAMDA (ASYMMETRIC) = 0.14122 WITH VAR015 DEPENDENT. = 0.04921 WITH VAR004 DEPENDENT.
 LAMDA (ASYMMETRIC) = 0.08940
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.03007 WITH VAR015 DEPENDENT. = 0.02162 WITH VAR004 DEPENDENT.
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.02515
 KENDALL'S TAU B = 0.15733 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.13892 SIGNIFICANCE = 0.0000
 GAMMA = 0.26238
 SOMERS'S D (ASYMMETRIC) = 0.13935 WITH VAR015 DEPENDENT. = 0.17764 WITH VAR004 DEPENDENT.
 SOMERS'S D (ASYMMETRIC) = 0.15618
 ETA = 0.02346 WITH VAR015 DEPENDENT. = 0.04551 WITH VAR004 DEPENDENT.
 NUMBER OF MISSING OBSERVATIONS = 1464

XIV. ANALISIS DE CORRELACION BIVARIADA: PEARSON CORR, NONPAR
CORR, SCATTERGRAM

XIV.1 El subprograma PEARSON CORR

Este subprograma calcula el coeficiente de correlación del producto momento de Pearson, entre un par de variables y la fórmula utilizada es:

$$r = \frac{\sum X_i Y_i - (\sum X_i)(\sum Y_i)/N}{\sqrt{\left[\sum X_i^2 - (\sum X_i)^2/N \right] \left[\sum Y_i^2 - (\sum Y_i)^2/N \right]}}$$

Donde N = número de observaciones

X_i = i-ésima observación de la variable X

Y_i = i-ésima observación de la variable Y

r = coeficiente de correlación

\sum = sumatoria para $i = 1, 2, \dots, N$

XIV.1.1 La tarjeta de procedimiento

Formato General

```

1                16
PEARSON CORR    {variable o [WITH {variable o /.....
                  lista de   lista de
                  variables}  variables}

```

Los paréntesis cuadrados [], indican que la palabra clave "WITH" es opcional. En el caso que no se especifique "WITH" o sea sólo tenemos una lista de variables el subprograma calculará los coeficientes de correlación no redundantes entre todos los pares de variables posibles, y esta forma es denominada forma matricial (sin el WITH).

Los siguientes ejemplos ilustran las dos formas de especificación:

Ejemplo a):

```

1                16
PEARSON CORR    A,B,C,D,E

```

En este caso serán computados coeficientes para los siguientes pares de variables:

- A - B B - C C - D D - E
- A - C B - D C - E
- A - D B - E
- A - E

ya que $r_{AB} = r_{BA}$

Ejemplo b):

```

1                16
PEARSON CORR    A,B WITH C,D,E / X1, X2, X3

```

En este ejemplo tendremos:

- Para la primera lista:

- A - C B - C
- A - D B - D
- A - E B - E

- Para la segunda lista (similar al ejemplo a)):

X1 - X2 Y2 - Y3
X1 - X3

Ejemplo c):

1	16
PEARSON CORR	X1 TO X5 WITH X1 TO X5

Producirá:

X1 - X1	X2 - X1	X3 - X1	X4 - X1	X5 - X1
X1 - X2	X2 - X2	X3 - X2	X4 - X2	X5 - X2
X1 - X3	X2 - X3	X3 - X3	X4 - X3	X5 - X3
X1 - X4	X2 - X4	X3 - X4	X4 - X4	X5 - X4
X1 - X5	X2 - X5	X3 - X5	X4 - X5	X5 - X5

Este ejemplo ilustra que el uso de WITH provoca el cálculo de los coeficientes de las variables especificadas a la izquierda con las variables a la derecha del WITH; y además ilustra que una serie de coeficientes redundantes serán producidos, dependiendo de las listas que sean declaradas.

XIV.1.2 Opciones para el subprograma PEARSON CORR

OPCION ESTANDAR. Eliminación de missing por pares.
La opción normal que utiliza el SPSS en el tratamiento de los valores missing es de eliminar el caso del cálculo si el valor de una de las variables del par es missing. Esto es ventajoso desde el punto de vista que cada coeficiente es calculado en base al máximo de casos válidos. Por otro lado, la desventaja reside en que el número de casos para el cómputo de cada coeficiente será distinto. El tratamiento de los missing puede ser modificado por medio de opciones 1,2.

OPCION 1: inclusión de los missing. Informa al sistema que ignore la declaración de MISSING y por lo tanto que incluya todos los casos en el cálculo.

OPCION 2: eliminación por lista. Esta opción provoca que un caso sea eliminado del cálculo de todos los coeficientes si el valor de alguna de las variables de la lista es un valor missing. Esta opción necesariamente reducirá el número total de casos en base a los cuales son calculados los coeficientes pero por otro lado garantiza que el número de casos siempre será el mismo para todos los coeficientes de una misma lista.

OPCION 3: test de significancia de dos colas. Esta opción produce que el test de significancia estadística de dos colas sea aplicado para cada coeficiente de correlación, en lugar del test de un cola que produce normalmente. O sea el test de 1 cola es aplicado en forma estándar a menos que se especifique la opción 3.

OPCION 4: Perforación (grabación) de matriz de coeficientes. Esta opción produce la perforación, grabación o impresión de las matrices de coeficientes de correlación para todas las listas especificadas en forma matricial (sin WITH), para usos futuros, tales como para el cálculo de regresión múltiple. El uso de esta opción para: "PEARSON CORR A,B,C", produce la grabación de la siguiente matriz de 3 x 3; simétrica:

AA	AB	AC
BA	BB	BC
CA	CB	CC

esto es con todos los coeficientes redundantes (AB, BA, etc.). La grabación se realiza por la unidad lógica FT09 o por la unidad deseada que el usuario quiera especificar por la declaración "RAW OUTPUT UNIT", en un registro o imagen de tarjeta con formato 8F10.7.

El formato 8F10.7 permite un máximo de 8 coeficientes por tarjeta. La grabación (perforación) se realiza por filas de la matriz y utilizará tantas tarjetas como sea necesario; siendo que cada fila se inicia en una tarjeta nueva.

OPCION 5: Omisión de número de casos y significancia. Solamente serán impresos los coeficientes de correlación en formato matricial; eliminando la impresión del número de casos y del test de significancia de 1 cola.

OPCION 6: Impresión sólo de coeficientes no redundantes en formato de hilera.

Opciones estándares:

- a) Los missing serán eliminados por pares.
- b) Test de 1 cola es aplicado a todos los coeficientes.
- c) No produce grabación (perforación) de la matriz de coeficientes.
- d) El número de casos y la significancia serán impresos.
- e) La impresión de los coeficientes es en formato matricial.

En el párrafo XIV.1.5, de ejemplos de salida se muestra la salida con el uso de cada opción y la opción estándar. Si el usuario especifica todas las opciones, el sistema la ignora y asume la opción estándar, pues hay opciones excluyentes.

XIV.1.3 Estadísticos disponibles para PEARSON CORR

ESTADISTICO 1: provoca el cálculo y la impresión de la media y desviación estándar de cada variable nombrada en las listas. La media y desviación estándar serán calculadas en base a todos los casos válidos (no missing) para una variable dada, a menos que opción 1 sea usada. La opción estándar y opción 2 no tienen ningún efecto en estos cálculos.

ESTADISTICO 2: produce la impresión de la co-varianza y producto cruzado de las desviaciones para cada par de variables. Al contrario del estadístico 1, estos cálculos se realizan sobre el mismo número de casos utilizado para el cálculo del coeficiente de correlación.

XIV.1.4 Limitaciones para PEARSON CORR

- 1) Un máximo de 40 listas en una solicitud de procedimiento.
- 2) Se permite un máximo de 500 nombres de variables (explícitamente o implícitamente). De esta forma si una variable aparece en 4 listas será contada como 4 veces para el total de 500 nombres.
- 3) El número de elementos máximo es de 250. Cada ocurrencia de nombre de variable, palabra clave y delimitadores especiales, cuenta como un elemento.
- 4) Requerimientos de memoria

- a) Para la opción estándar de tratamiento de los missing.

$$\text{WORKSPACE} = 48 * \text{NCOEFF}$$

- b) Para opción 1 ó 2

$$\text{WORKSPACE} = 8 * \text{NCOEFF}$$

Donde:

NCOEFF = número de coeficientes que serán calculados, resultado de la sumatoria de todas las listas.

Para listas de variables con 'WITH

$$\text{NCOEFF} = \text{NL} * \text{NR}$$

donde NL = número de variables a la izquierda del WITH

NR = número de variables a la derecha

Para listas de variables en forma matricial:

$$\text{NCOEFF} = \frac{\text{NVAR} * (\text{NVAR} - 1)}{2}$$

donde NVAR = número de variables especificadas en la lista.

XIV.1.5 Ejemplos de salida de "PEARSON CORR"

Ejemplo a): Se muestra la salida producida con STATISTICS ALL - los estadísticos estan en páginas 4 y 5. El uso de la opción 4 provoca la grabación de la matriz cuyo listado se muestra.

TASK NAME SALIDA ESTANDAR, ESTADISTICOS Y OPCION 4 - GRABA MATRIZ DE COEF.
 PEARSON CORR VAR028 TO VAR032
 OPTIONS 4
 STATISTICS ALL

CELADE - Banco de Datos

EJEMPLO DEL PROCEDIMIENTO - PEARSON CORR -
 SALIDA ESTANDAR, ESTADISTICOS Y OPCION 4 - GRABA MATRIZ DE COEF.
 FILE NONAME (CREATION DATE = 09R03R76)

09R03R76 PAGE 4

VARIABLE	CASES	MEAN	STD DEV
VAR028	79	23693.7969	28445.6484
VAR029	79	26.7886	22.5618
VAR030	79	13.0127	74.6128
VAR031	79	0.2724	0.3603
VAR032	79	391.6328	999.3093

EJEMPLO DEL PROCEDIMIENTO - PEARSON CORR -
 SALIDA ESTANDAR, ESTADISTICOS Y OPCION 4 - GRABA MATRIZ DE COEF.
 FILE NONAME (CREATION DATE = 09R03R76)

09R03R76 PAGE 5

VARIABLES	CASES	CROSS-PROD DEV	VARIANCE-COVAR	VARIABLES	CASES	CROSS-PROD DEV	VARIANCE-COVAR
VAR028 VAR029	79	25879088.0	331783.125	VAR028 VAR030	79	-13431911.0	-172203.93
VAR028 VAR031	79	-67844.5000	-869.8013	VAR028 VAR032	79	991116544.	12706622.0
VAR029 VAR030	79	-21675.3906	-277.8894	VAR029 VAR031	79	-127.3044	-1.63
VAR029 VAR032	79	886171.875	11361.1758	VAR030 VAR031	79	1973.0405	25.29
VAR030 VAR032	79	186779.875	2394.6138	VAR031 VAR032	79	-2398.6797	-30.75

EJEMPLO DEL PROCEDIMIENTO - PEARSON CORR -
 SALIDA ESTANDAR, ESTADISTICOS Y OPCION 4 - GRABA MATRIZ DE COEF.
 FILE NONAME (CREATION DATE = 09R03R76)

09R03R76 PAGE 6

----- PEARSON CORRELATION COEFFICIENTS -----

VAR028	VAR029	VAR030	VAR031	VAR032	
VAR028	1.0000 (.01) \$=0.001	0.5170 (.79) \$=0.001	-0.0811 (.79) \$=0.239	-0.0849 (.79) \$=0.229	0.4470 (.79) \$=0.001
VAR029	0.5170 (.79) \$=0.001	1.0000 (.01) \$=0.001	-0.1651 (.79) \$=0.073	-0.2008 (.79) \$=0.038	0.5039 (.79) \$=0.001
VAR030	-0.0811 (.79) \$=0.239	-0.1651 (.79) \$=0.073	1.0000 (.01) \$=0.001	0.9409 (.79) \$=0.001	0.0321 (.79) \$=0.389
VAR031	-0.0849 (.79) \$=0.229	-0.2008 (.79) \$=0.038	0.9409 (.79) \$=0.001	1.0000 (.01) \$=0.001	-0.0854 (.79) \$=0.227
VAR032	0.4470 (.79) \$=0.001	0.5039 (.79) \$=0.001	0.0321 (.79) \$=0.389	-0.0854 (.79) \$=0.227	1.0000 (.01) \$=0.001

(COEFFICIENT R (CASES) R SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

Listado de la matriz grabada en FTO9, con el uso de la opción 4

1.0000000	0.5169703	-0.0811360	-0.0848672	0.4470070
0.5169703	1.0000000	-0.1650769	-0.2007760	0.5039073
-0.0811360	-0.1650769	1.0000000	0.9409447	0.0321161
-0.0848672	-0.2007760	0.9409447	1.0000000	-0.0854110
0.4470070	0.5039073	0.0321161	-0.0854110	1.0000000

Ejemplo b): Ilustra la salida producida con el uso de la opción 5, la cual informa al sistema que sólo imprima los coeficientes de correlación. Obsérvese que el sistema imprime un asterisco (*) y (**) si la significancia es menor o igual a 0.01 y 0.001, respectivamente.

TASK NAME USO DE OPCION 5 - ELIMINA IMPRESION DE FRECUENCIAS Y SIGNIFICANC.
 PEARSON CORR VAR028 TO VAR032
 OPTIONS 5

EJEMPLO DEL PROCEDIMIENTO -PEARSON CORR -
 USO DE OPCION 5 - ELIMINA IMPRESION DE FRECUENCIAS Y SIGNIFICANC.
 FILE NONAME (CREATION DATE = 09R03R76)

09R03R76 PAGE 10

----- PEARSON CORRELATION COEFFICIENTS -----

	VAR028	VAR029	VAR030	VAR031	VAR032
VAR028	1.0000	0.5170**	-0.0811	-0.0849	0.4470**
VAR029	0.5170**	1.0000	-0.1651	-0.2008	0.5039**
VAR030	-0.0811	-0.1651	1.0000	0.9409**	0.0321
VAR031	-0.0849	-0.2008	0.9409**	1.0000	-0.0854
VAR032	0.4470**	0.5039**	0.0321	-0.0854	1.0000

* - SIGNIF. LE .01 ** - SIGNIF. LE .001 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED

Ejemplo c): Este formato, llamado formato de hilera es producido con el uso de la opción 6:

TASK NAME USO DE OPCION 6 - FORMATO DE HILERA CON COEF. NO REDUNDANTES
 PEARSON CORR VAR028 TO VAR032
 OPTIONS 6

EJEMPLO DEL PROCEDIMIENTO -PEARSON CORR -
 USO DE OPCION 6 - FORMATO DE HILERA CON COEF. NO REDUNDANTES
 FILE NONAME (CREATION DATE = 09R03R76)

09R03R76 PAGE 12

----- PEARSON CORRELATION COEFFICIENTS -----

VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR
VAR028	0.5170	VAR028	-0.0811	VAR028	-0.0849
WITH N(79)		WITH N(79)		WITH N(79)	
VAR029	SIG .001	VAR030	SIG .239	VAR031	SIG .229
VAR028		VAR029		VAR028	0.4470
				WITH N(79)	
				VAR032	SIG .001
				VAR029	-0.1651
				WITH N(79)	
				VAR030	SIG .073
				VAR029	-0.2008
				WITH N(79)	
				VAR031	0.9409
				WITH N(79)	
				VAR032	SIG .001
				VAR030	0.0321
				WITH N(79)	
				VAR031	-0.0854
				WITH N(79)	
				VAR032	SIG .227

A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

XIV.2 El subprograma NONPAR CORR

Este subprograma calcula los coeficientes de correlación no paramétricos de SPEARMAN y/o KENDALL.

El procedimiento de cálculo de ambos coeficientes, que requieren el uso de rangos en vez de los valores absolutos de cada variable, motiva que NONPAR CORR y SCATTERGRAM sean prácticamente los únicos subprogramas del SPSS en los cuales hay una limitación del número de casos que pueden procesar en función del espacio de memoria disponible.

El coeficiente r_s de Spearman es formalmente definido como:

$$r_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N^3 - N}$$

El SPSS, para propósitos computacionales y para corregir la ocurrencia de rangos empatados utiliza la siguiente re-definición para r_s :

$$r_s = \frac{T_x + T_y - \sum d_i^2}{2(T_x \cdot T_y)^{1/2}}$$

Donde:

d_i = diferencia entre rangos de 2 variables para el caso i .

T_x o T_y es definido como $\frac{N(N^2 - 1) - R(R^2 - 1)}{12}$

Donde R es el número de empates a un cierto rango para X o Y respectivamente.

El coeficiente TAU B de Kendall es definido como:

$$TAU = \frac{S}{1/2N(N - 1)}$$

Para corregir los empates de rango la re-definición es:

$$TAU = \frac{S}{\sqrt{\frac{1}{2} N(N - 1) - T_x} \sqrt{\frac{1}{2} N(N - 1) - T_y}}$$

$T_x = \frac{1}{2} t(t-1)$ donde t es el número de observaciones empatadas en cada grupo de empate de la variable X; T_y es similar para la variable Y.

XIV.2.1 La tarjeta de procedimiento

Formato general

1	16			
NONPAR CORR	{ variable o lista de variables }	[WITH	{ variable o lista de variables }	/...

Las mismas reglas para PEARSON CORR son válidas para NONPAR CORR.

XIV.2.2 Opciones disponibles para NONPAR CORR

OPCION ESTANDAR: igual que PEARSON CORR

OPCION 1, 2, 3, 4: igual a PEARSON CORR

OPCION 5: impresión sólo de los coeficientes de Kendall

OPCION 6: produce los dos coeficientes -de Kendall y Spearman. Si se usa OPTIONS 4,6 la grabación se realiza primero para la matriz de coeficientes de Kendall y posteriormente la matriz de los coeficientes de Spearman.

Si no se usa ni opción 5, ni la 6, la salida estándar es con los coeficientes de Spearman.

El formato de impresión es siempre estándar y en formato de hilera.

XIV.2.3. Estadísticos para NONPAR CORR

El subprograma NONPAR CORR es uno de los procedimientos SPSS que no posee la tarjeta STATISTICS.

XIV.2.4 Limitaciones para NONPAR CORR

- 1) Un máximo de 100 variables pueden ser especificadas en una solicitud de NONPAR CORR.

2) El número de listas máximo = 25

3) Requerimientos de memoria

a) Para la opción estándar de tratamiento de missing.

$$\text{WORKSPACE} = 4 * \text{NVAR} * \text{NCASE} + 10 * \text{NCASE}$$

b) Si se usa opción 1 ó 2:

$$\text{WORKSPACE} = 4 * \text{NVAR} * \text{NCASE} + 2 * \text{NCASE}$$

donde:

NVAR = total de variables nombradas explícita o implícitamente en las listas. Cada ocurrencia de una misma variable incrementa UNO al valor de NVAR.

NCASE = número de casos en el archivo, o en el subarchivo en el caso que se está procesando cada subarchivo por separado.

En el caso del uso de la opción 4 (grabación de matrices) y que las matrices contienen más de 30 variables se requiere el siguiente espacio de memoria adicional al anteriormente especificado:

$$\text{WORKAD} = 2 * \text{MAXLIST} * (\text{MAXLIST} - 1)$$

WORKAD = espacio adicional en bytes

MAXLIST = largo de la mayor lista de variables para la cual se desea la grabación de la matriz.
(largo = número de variables en la lista).

XIV.2.5 Ejemplos de salida de NONPAR CORR

Ejemplo a): Salida producida con las opciones estándares - sólo calcula e imprime los coeficientes de Spearman y en formato de hilera.

EJEMPLO DEL PROCEDIMIENTO - NONPAR CORR - 09R03R76 PAGE 2
 TASK NAME SALIDA CON OPCIONES ESTANDARES
 NONPAR CORR VAR008,VAR002,VAR004,VAR006,VAR007,VAR011

EJEMPLO DEL PROCEDIMIENTO - NONPAR CORR - 09R03R76 PAGE 3
 SALIDA CON OPCIONES ESTANDARES
 FILE NONAME (CREATION DATE = 09R03R76)
 SUBFILE SUBF1 SUBF2

----- S P E A R M A N C O R R E L A T I O N C O E F F I C I E N T S -----

VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR	
VAR008 WITH VAR002	0.0624 N(70) SIG .303	VAR008 WITH VAR004	0.0902 N(70) SIG .229	VAR008 WITH VAR006	-0.1660 N(70) SIG .085	VAR008 WITH VAR007	-0.2292 N(70) SIG .028	VAR008 WITH VAR011	0.1772 N(70) SIG .071	VAR002 WITH VAR004	0.5865 N(70) SIG .001
VAR002 WITH VAR006	-0.5759 N(70) SIG .001	VAR002 WITH VAR007	-0.5729 N(70) SIG .001	VAR002 WITH VAR011	0.6082 N(70) SIG .001	VAR004 WITH VAR006	-0.8430 N(70) SIG .001	VAR004 WITH VAR007	-0.8361 N(70) SIG .001	VAR004 WITH VAR011	0.8143 N(70) SIG .001
VAR006 WITH VAR007	0.7794 N(70) SIG .001	VAR006 WITH VAR011	-0.7828 N(70) SIG .001	VAR007 WITH VAR011	-0.8460 N(70) SIG .001						

A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Ejemplo b): El uso de la opción 5 provee la salida solamente de los coeficientes de Kendall como se muestra:

TASK NAME USO DE OPCION 5 - SOLO COEFICIENTES DE KENDALL
 NONPAR CORR VAR008,VAR002,VAR004,VAR006,VAR007,VAR011
 OPTIONS 5

EJEMPLO DEL PROCEDIMIENTO - NONPAR CORR - 09R03R76 PAGE 5
 USO DE OPCION 5 - SOLO COEFICIENTES DE KENDALL
 FILE NONAME (CREATION DATE = 09R03R76)
 SUBFILE SUBF1 SUBF2

----- K E N D A L L C O R R E L A T I O N C O E F F I C I E N T S -----

VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR	
VAR008 WITH VAR002	0.0482 N(70) SIG .284	VAR008 WITH VAR004	0.0453 N(70) SIG .298	VAR008 WITH VAR006	-0.1096 N(70) SIG .097	VAR008 WITH VAR007	-0.1498 N(70) SIG .038	VAR008 WITH VAR011	0.1065 N(70) SIG .103	VAR002 WITH VAR004	0.4406 N(70) SIG .001
VAR002 WITH VAR006	-0.4191 N(70) SIG .001	VAR002 WITH VAR007	-0.4048 N(70) SIG .001	VAR002 WITH VAR011	0.4387 N(70) SIG .001	VAR004 WITH VAR006	-0.6760 N(70) SIG .001	VAR004 WITH VAR007	-0.6732 N(70) SIG .001	VAR004 WITH VAR011	0.6473 N(70) SIG .001
VAR006 WITH VAR007	0.5922 N(70) SIG .001	VAR006 WITH VAR011	-0.5968 N(70) SIG .001	VAR007 WITH VAR011	-0.6689 N(70) SIG .001						

A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Ejemplo e): El uso de la opción 6, informa al sistema que calcule e imprima los 2 coeficientes como se muestra en este ejemplo:

TASK NAME USD DE OPCION 6 - COEFICIENTES DE SPEARMAN Y KENDALL
 NONPAR CORR VAR008,VAR002,VAR004,VAR006,VAR007,VAR011
 OPTIONS 6

EJEMPLO DEL PROCEDIMIENTO - NONPAR CORR -
 USO DE OPCION 6 - COEFICIENTES DE SPEARMAN Y KENDALL
 FILE NONAME (CREATION DATE = 09R03R76)
 SUBFILE SUBF1 SUBF2

09R03R76

PAGE 7

----- KENDALL CORRELATION COEFFICIENTS -----

VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR	
VAR008	0.0482	VAR008	0.0453	VAR008	-0.1096	VAR008	-0.1498	VAR008	0.1065	VAR002	0.4466
WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)	
VAR002	SIG .284	VAR004	SIG .298	VAR006	SIG .097	VAR007	SIG .038	VAR011	SIG .103	VAR004	SIG .001
VAR002	-0.4191	VAR002	-0.4048	VAR002	0.4387	VAR004	-0.6760	VAR004	-0.6732	VAR004	0.6473
WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)	
VAR006	SIG .001	VAR007	SIG .001	VAR011	SIG .001	VAR006	SIG .001	VAR007	SIG .001	VAR011	SIG .001
VAR006	0.5922	VAR006	-0.5968	VAR007	-0.6689						
WITH N(70)		WITH N(70)		WITH N(70)							
VAR007	SIG .001	VAR011	SIG .001	VAR011	SIG .001						

A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

EJEMPLO DEL PROCEDIMIENTO - NONPAR CORR -
 USO DE OPCION 6 - COEFICIENTES DE SPEARMAN Y KENDALL
 FILE NONAME (CREATION DATE = 09R03R76)
 SUBFILE SUBF1 SUBF2

09R03R76

PAGE 8

----- SPEARMAN CORRELATION COEFFICIENTS -----

VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR	
VAR008	0.0629	VAR008	0.0902	VAR008	-0.1660	VAR008	-0.2292	VAR008	0.1772	VAR002	0.5865
WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)	
VAR002	SIG .303	VAR004	SIG .229	VAR006	SIG .085	VAR007	SIG .028	VAR011	SIG .071	VAR004	SIG .001
VAR002	-0.5759	VAR002	-0.5729	VAR002	0.6082	VAR004	-0.8430	VAR004	-0.8361	VAR004	0.8143
WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)		WITH N(70)	
VAR006	SIG .001	VAR007	SIG .001	VAR011	SIG .001	VAR006	SIG .001	VAR007	SIG .001	VAR011	SIG .001
VAR006	0.7794	VAR006	-0.7828	VAR007	-0.8460						
WITH N(70)		WITH N(70)		WITH N(70)							
VAR007	SIG .001	VAR011	SIG .001	VAR011	SIG .001						

A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

XIV.3 El subprograma SCATTERGRAM

Este subprograma produce una nube de puntos, bidimensional, que permite la visualización gráfica de la correlación existente o no entre dos variables.

El gráfico producido por SCATTERGRAM posee en la coordenada vertical 51 unidades y 101 en la horizontal. La tarjeta de control PAGESIZE no afecta la salida de SCATTERGRAM.

La escala estándar (si el usuario no especifica otra) es construida en base al valor mínimo y máximo detectado para cada variable y estos valores serán impresos en los extremos de los ejes. En forma alternativa, el usuario puede fijar rangos (min,max) para las variables en la lista con lo cual delimita la escala ya que estos valores serán utilizados para los extremos de los ejes. Si se especifican los rangos y además opción 7, la opción 7 será ignorada.

Los puntos son representados por un asterisco (*) si solamente un caso trae la coordenada (x,y). Si 2,3, ...8, casos poseen las mismas coordenadas (x,y) el programa imprimirá en vez de (*) el número 2,3, ... 8, indicando el número de casos que coinciden en el mismo punto. El número 9 será impreso si el número de ocurrencias en un mismo punto es de 9 y + casos.

XIV.3.1 La tarjeta de procedimiento de SCATTERGRAM

En forma similar a PEARSON CORR y NONPAR CORR la especificación para SCATTERGRAM puede utilizar o no la palabra clave WITH.

Formato General 1

<u>1</u>	<u>16</u>		
SCATTERGRAM	{ variable o lista de variables }	{ (minimo,maximo) (LOWEST,maximo) (minimo,HIGHEST) }	, ... , { variable o lista de variables } { (minimo,maximo) (LOWEST,maximo) (minimo,HIGHEST) } / ...

Ejemplos:

```

1           16
SCATTERGRAM  A,B,C,D,E/X1,X2,X3
SCATTERGRAM  VAR001 TO VAR004 (0,10)/
              X1,X2 (0,100), X4,X5 (LOWEST, 20)
SCATTERGRAM  X1 (7,HIGHEST), X2 (0,20), X3 (LOWEST, 10)
    
```

El número de gráficos es determinado por la siguiente fórmula:

$$NT = \frac{1}{2} (n^2 - n) \text{ donde } n = n^\circ \text{ de variables de la lista.}$$

Así la lista "A,B,C,D,E" producirá $\frac{1}{2} (5^2 - 5) = \frac{1}{2} (25 - 5) = 10$

gráficos, que serían los siguientes:

	<u>Variable vertical</u>	<u>Variable horizontal</u>
1.	A	B
2.	A	C
3.	A	D
4.	A	E
5.	B	C
6.	B	D
7.	B	E
8.	C	D
9.	C	E
10.	D	E

Formato General 2

```

1           16
SCATTERGRAM  { variable o [ (minimo,maximo)
                  lista de  (LOWEST,maximo)
                  variables } (minimo,HIGHEST) WITH { variable o [ (minimo,maximo)
                  lista de  (LOWEST,maximo)
                  variables } (minimo,HIGHEST) /...
    
```

Ejemplos

```

1           16
SCATTERGRAM  X1,X2(0,10),X3(1,20) WITH X4
SCATTERGRAM  X1,X2 WITH X3,X4 (LOWEST,20) /
              X5,X6 WITH A,B,C
    
```

Si se especifica el rango (min,max), el sistema sólo aceptará los casos con valores dentro del rango e informará con el mensaje EXCLUDED VALUES los valores excluidos del cálculo.

XIV.3.2 Opciones disponibles en SCATTERGRAM

- OPCION 1: inclusión de missing -ignoraré la declaración MISSING y considerará todos los casos como válidos tanto para el gráfico como para los estadísticos.
- OPCION 2: exclusión de missing por lista. Un caso será excluido del gráfico y de los estadísticos si el valor de una de las variables de la lista es missing.
- OPCION 3: elimina impresión de VAR LABELS
- OPCION 4: suprime la impresión de la grilla. Esta consiste en líneas paralelas (horizontales y verticales) que dividen el gráfico en 9 rectángulos iguales y permite una mejor localización de los puntos.
- OPCION 5: Imprime grillas diagonales.
- OPCION 6: Aplica el test de significancia de dos colas si se escoge el estadístico 3.
- OPCION 7: Elabora escalas en forma automática de manera tal que las coordenadas resulten números enteros. Esta opción es ignorada si el usuario especifica un rango.
- OPCION 8: informa al sistema que si el espacio de memoria no es suficiente para el número de casos en el archivo, que produzca el gráfico para los n primeros casos, donde n es el número máximo de casos que puede ser procesado con el espacio de memoria disponible.

Si no se especifican opciones el sistema asume:

- a) Eliminación de missing por pares.
- b) VAR LABELS son impresos
- c) Grilla (líneas horizontales y cverticales) es producida.
- d) Test de un cola si se pide el estadístico 3.
- e) No hace ajuste automático de escala para producir coordenadas enteras.

f) No produce ningún gráfico si se agota el espacio de memoria.

XIV.3.3 Estadísticos disponibles para SCATTERGRAM

<u>Número</u>	<u>Estadístico</u>
1	r de PEARSON (coeficiente de correlación)
2	r^2
3	significancia de r
4	error estándar
5	punto de intersección con el eje vertical
6	regresión lineal simple.

XIV.3.4 Limitaciones para SCATTERGRAM

1) Número máximo de variables = 100

2) Número máximo de listas = 25

3) Requerimientos de memoria

$$\text{WORKSPACE} = 4 * (\text{NV} + 1) * (\text{NCASE} + 2)$$

donde NV = número de variables

NCASE = número de casos a ser procesado

XIV.3.5 Ejemplos de salida del SCATTERGRAM

Ejemplo a): Opciones estándares y solicitud de todos los estadísticos.

EJEMPLO DEL PROCEDIMIENTO - SCATTERGRAM -

09R04R76

PAGE 2

TASK NAME USU DE OPCIONES ESTANDARES Y TODOS LOS ESTADISTICOS
SCATTERGRAM VAR001 WITH VAR012
STATISTICS ALL

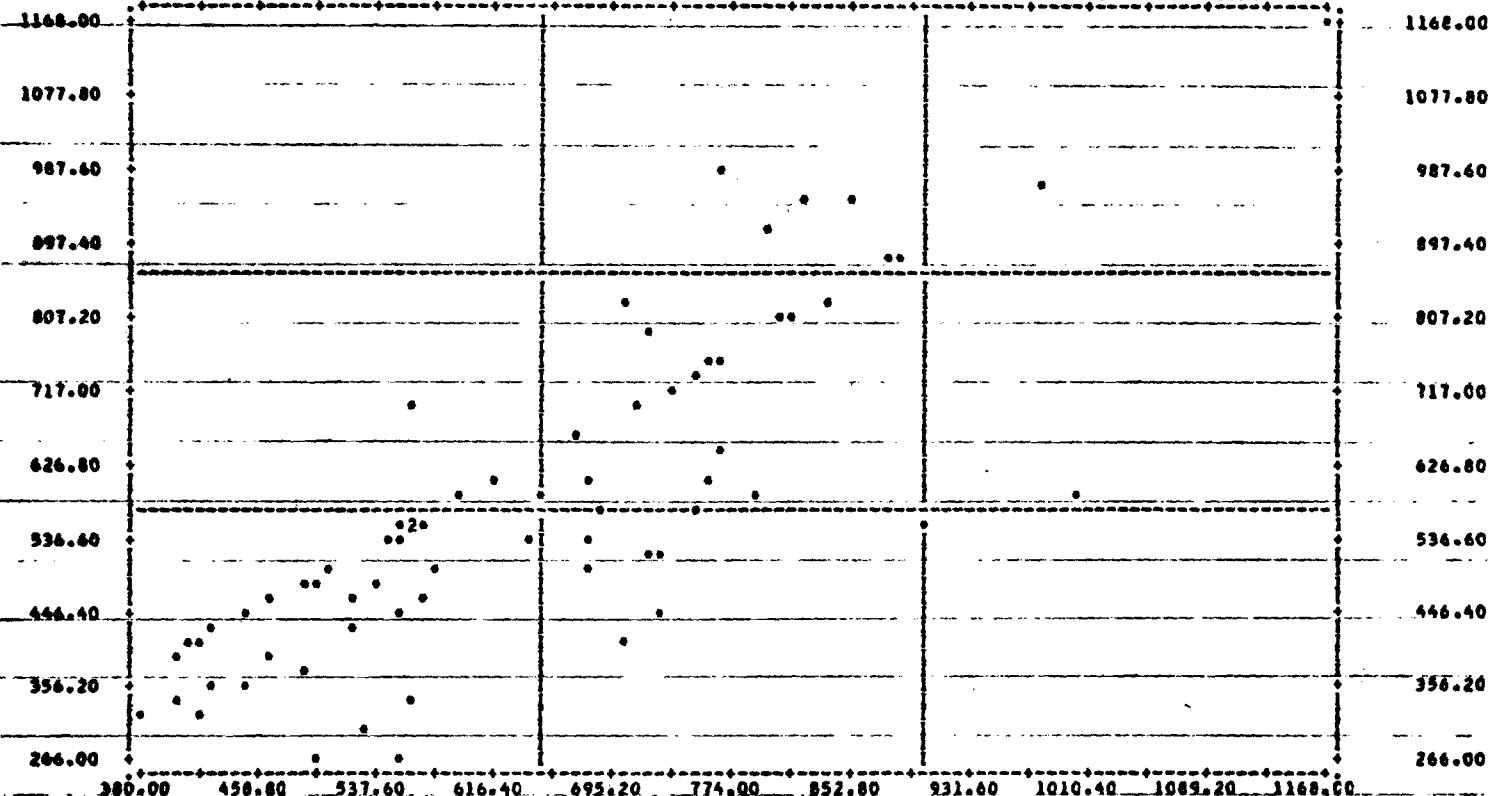
EJEMPLO DEL PROCEDIMIENTO - SCATTERGRAM -

09R04R76

PAGE 3

USU DE OPCIONES ESTANDARES Y TODOS LOS ESTADISTICOS
FILE NONAME (CREATION DATE = 09R04R76)

SCATTERGRAM OF (DOWN) VAR001 TASA DE MORTALIDAD INFANTIL (ACROSS) VAR012 TASA DE MORTALIDAD INFANTIL CORR
SUBJ1 SUBJ2
419.40 498.20 577.00 655.80 734.60 813.40 892.20 971.00 1049.80 1128.60



EJEMPLO DEL PROCEDIMIENTO - SCATTERGRAM -
USU DE OPCIONES ESTANDARES Y TODOS LOS ESTADISTICOS

09R04R76

PAGE 4

STATISTICS..

CORRELATION (R) -	0.81372	R SQUARED -	0.66214	SIGNIFICANCE -	0.00001
STD ERR OF EST -	114.81989	INTERCEPT (A) -	-34.38376	SLOPE (B) -	0.98800
PLOTTED VALUES -	70	EXCLUDED VALUES -	0	MISSING VALUES -	0

***** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

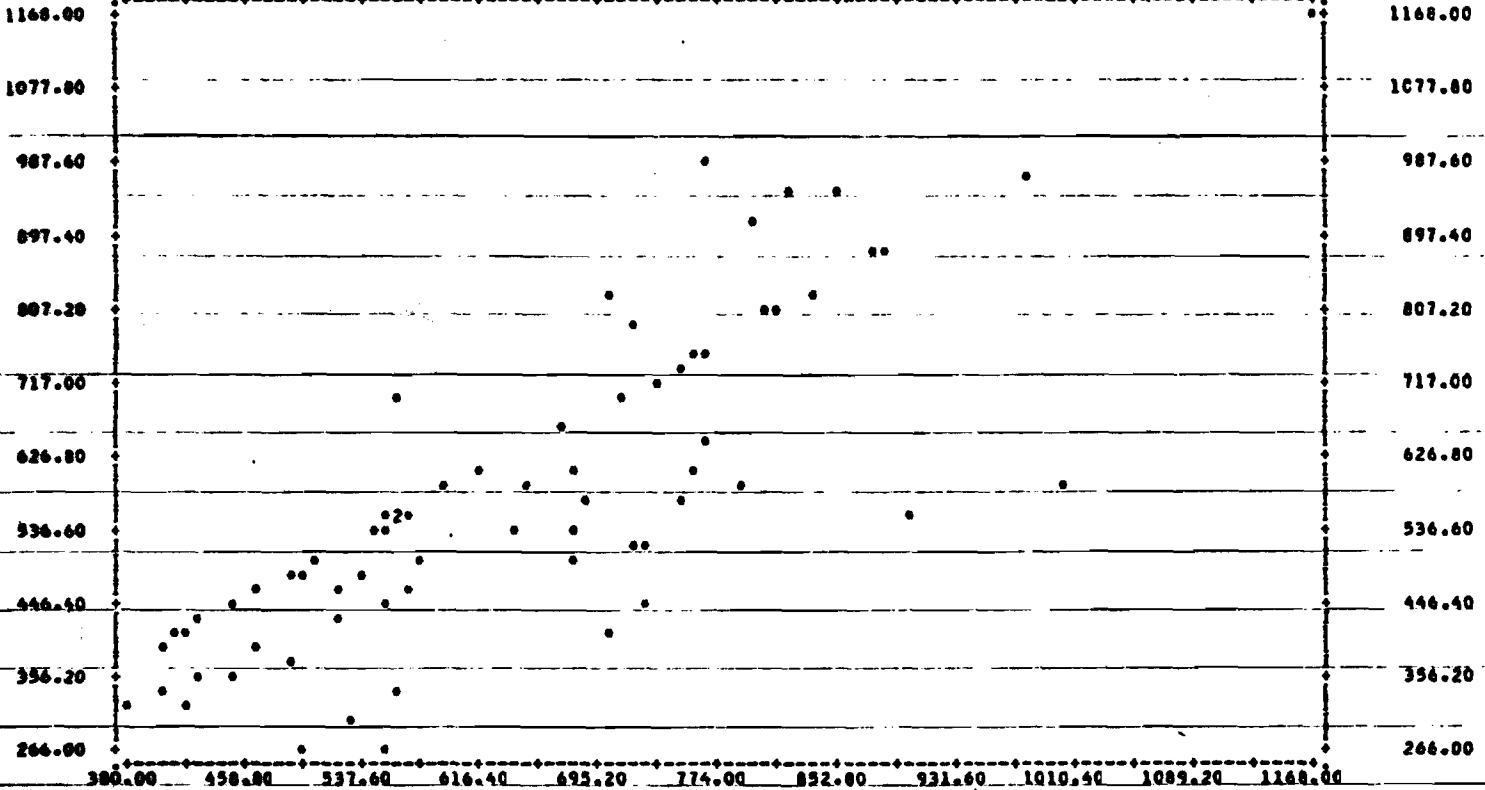
Ejemplo b): Uso de opción 4 que provoca la no impresión de la grilla.

TASK NAME USO DE OPCION 4 - SUPRIME IMPRESION DE LA GRILLA
SCATTERGRAM VAR001 WITH VAR012
OPTIONS 4

EJEMPLO DEL PROCEDIMIENTO - SCATTERGRAM -
USO DE OPCION 4 - SUPRIME IMPRESION DE LA GRILLA
FILE NAME (CREATION DATE = 09R04R76)
SUBFILE SUBF1 SUBF2

09R04R76 PAGE 6

SCATTERGRAM OF (DOWN) VAR001 TASA DE MORTALIDAD INFANTIL (ACROSS) VAR012 TASA DE MORTALIDAD INFANTIL CCRR
419.40 498.20 577.00 655.80 734.60 813.40 892.20 971.00 1049.80 1128.60



EJEMPLO DEL PROCEDIMIENTO - SCATTERGRAM -
USO DE OPCION 4 - SUPRIME IMPRESION DE LA GRILLA

09R04R76 PAGE 7

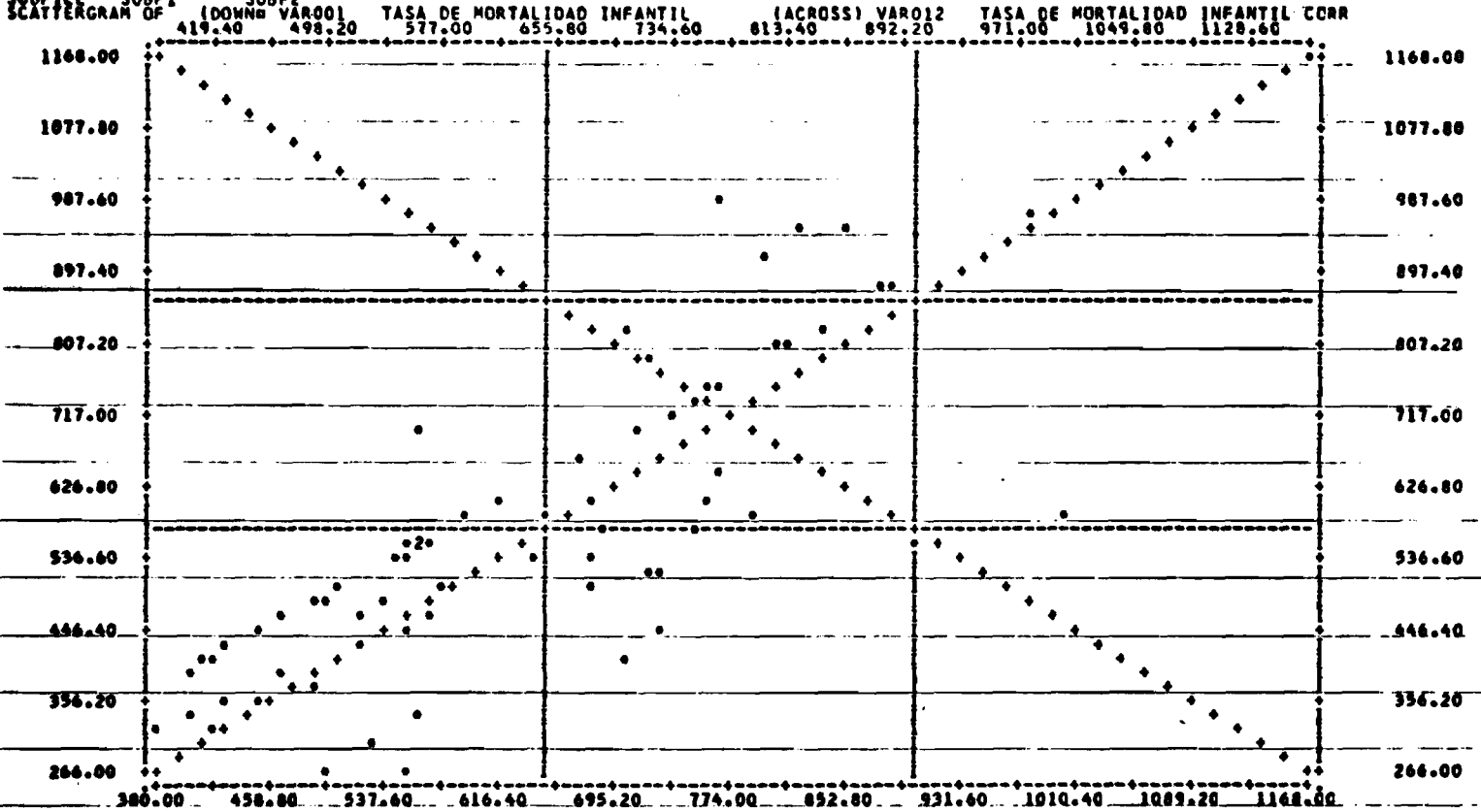
PLOTTED VALUES - 70 EXCLUDED VALUES - 0 MISSING VALUES - 0

Ejemplo e): Use de opción 5: impresión de grilla diagonal.

TASK NAME USO DE OPCION 5 - IMPRIME GRILLA DIAGONAL
SCATTERGRAM VAR001 WITH VAR012
OPTIONS 1

09R04R76 PAGE 9

EJEMPLO DEL PROCEDIMIENTO - SCATTERGRAM -
USO DE OPCION 5 - IMPRIME GRILLA DIAGONAL
FILE NOVARE (CREATION DATE = 09R04R76)
SUBFILE SUBF1 SUBF2
SCATTERGRAM OF



EJEMPLO DEL PROCEDIMIENTO - SCATTERGRAM -
USO DE OPCION 5 - IMPRIME GRILLA DIAGONAL

09R04R76 PAGE 10

PLOTTED VALUES - 70 EXCLUDED VALUES - 0 MISSING VALUES - 0

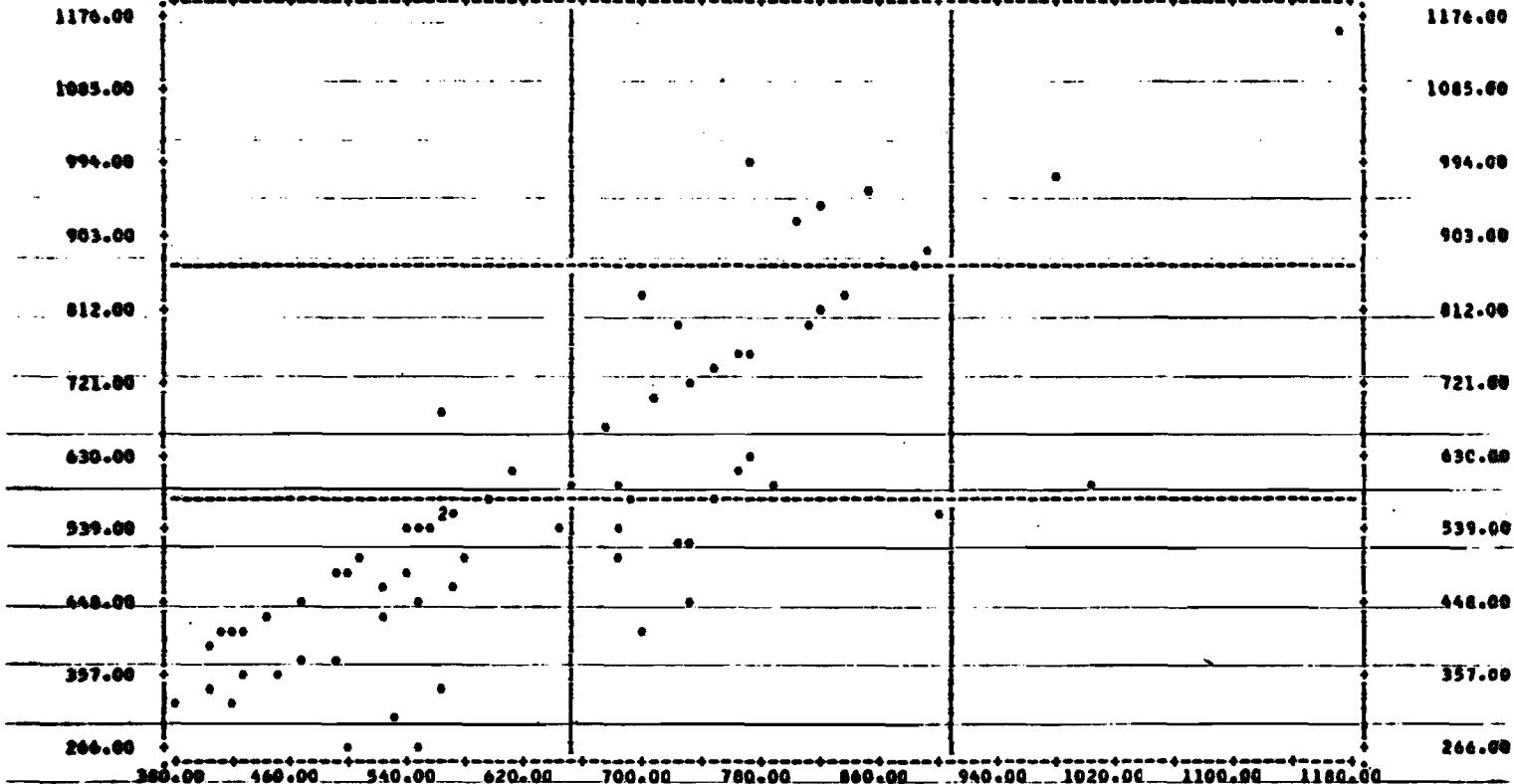
Ejemplo d): Use de la opción 7: construye las escalas en forma automática de tal forma que las coordenadas resulten números enteros.

TASK NAME USO DE OPCION 7 - ELABORACION DE ESCALA AUTOMATICA
SCATTERGRAM VAR001 WITH VAR012
OPTIONS 7

EJEMPLO DEL PROCEDIMIENTO - SCATTERGRAM -
USO DE OPCION 7 - ELABORACION DE ESCALA AUTOMATICA
FILE NAME (CREATION DATE = 09R04R76)
SUBFILE SUBF1 SUBF2
SCATTERGRAM OF (DOWN) VAR001 TASA DE MORTALIDAD INFANTIL (ACROSS) VAR012 TASA DE MORTALIDAD INFANTIL CORR
420.00 500.00 580.00 660.00 740.00 820.00 900.00 980.00 1060.00 1140.00

09R04R76

PAGE 15



EJEMPLO DEL PROCEDIMIENTO - SCATTERGRAM -
USO DE OPCION 7 - ELABORACION DE ESCALA AUTOMATICA

09R04R76

PAGE 16

PLOTTED VALUES -

70

EXCLUDED VALUES -

0

MISSING VALUES -

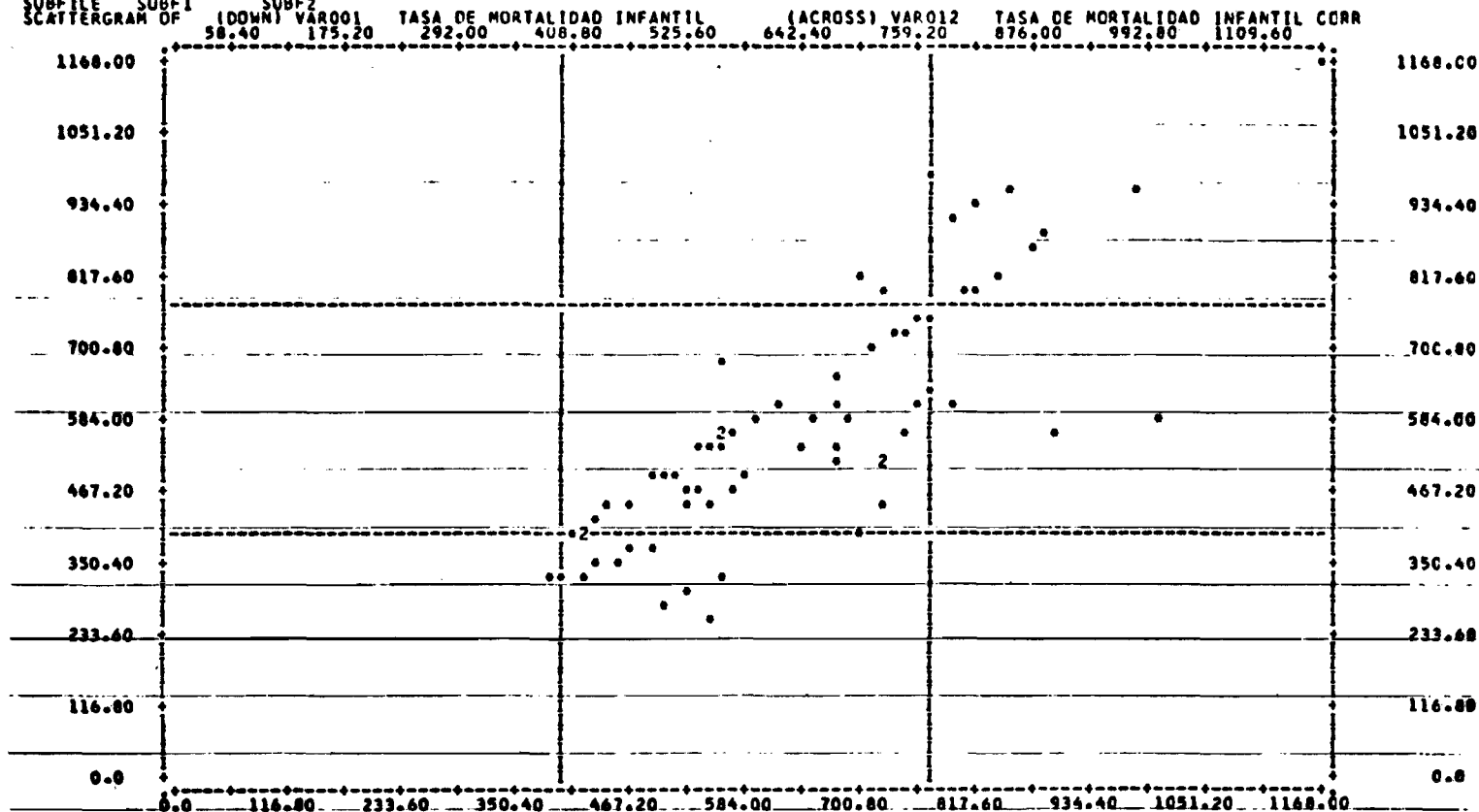
0

Ejemplo e): En este ejemplo la escala fue fijada por el rango de las variables.

TASK NAME USO DE ESCALAS FIJADAS POR EL RANGO
 SCATTERGRAM VAR001(0,HIGHEST) WITH VAR012 (0,HIGHEST)

EJEMPLO DEL PROCEDIMIENTO - SCATTERGRAM -
 USO DE ESCALAS FIJADAS POR EL RANGO
 FILE NONAME (CREATION DATE = 09R04R76H
 SUBFILE SUBF1 SUBF2

09R04R76 PAGE 18



EJEMPLO DEL PROCEDIMIENTO - SCATTERGRAM -
 USO DE ESCALAS FIJADAS POR EL RANGO

09R04R76 PAGE 19

PLOTTED VALUES - 70 EXCLUDED VALUES - 0 MISSING VALUES - 0

XV. PARTIAL CORR: CORRELACION PARCIAL

XV.1. La tarjeta de procedimiento:

Formato general

<u>1</u>	<u>16</u>		
PARTIAL CORR	lista de correlación	BY	lista de control (valores de orden) /
	lista de correlación	BY	lista de control (valores de orden) /...

A. Lista de correlación

En la lista de correlación especificamos las variables que serán correlacionadas, con el control de las variables que se especifiquen en la lista de control. La forma en que debemos especificar la lista de correlación es totalmente si milar a la forma que utiliza los subprogramas PEARSON CORR y NONPAR CORR: 1) puede ser una lista con la palabra clave WITH o 2) puede ser una lista en forma matricial sin "WITH".

Ejemplos de Listas de correlación:

"A,B,C,D,E" forma matricial

"X1 TO X5 WITH A,B"

B. Lista de control y valores de orden

La lista de variables de control es una lista de variables simples al final de la cual especificamos entre paréntesis los valores de orden que informan al sistema el orden de correlaciones parciales que estamos solicitando. Se permite especificar un máximo de 5 valores de orden.

Ejemplos de especificación de valores de orden.

(1,3,6) ... válido

(1,2,3,5,7,8) ... inválido pues trae 6 valores

(1,2,7) ... válido

Antes de los ejemplos, veamos el formato de la tarjeta de procedimiento nuevamente, ahora especificando con mayor detalle:

1	16		
PARTIAL CORR	{	lista de vars WITH lista de vars	BY
		o	
	}	lista de variables en formato matricial	
		lista de variables (valores de orden) / ...	

Ejemplos de especificación:

(a)

1	16		
PARTIAL CORR	A	WITH Z BY X1,X2 (1)	

Serán calculados dos parciales de 1er. orden que son

A - Z controlado por X1

A - Z controlado por X2

(b)

1	16		
PARTIAL CORR	A	WITH Z BY X1,X2 (2)	

Aquí se calculará sólo un coeficiente de correlación parcial de 2° orden entre A y Z controlado simultáneamente por X1 y X2.

(c)

Los dos ejemplos (a,b) pueden ser fundidos en una sola solicitud que sería:

1	16		
PARTIAL CORR	A	WITH Z BY X1,X2 (1,2)	

En este ejemplo la especificación de los valores de orden (1,2) informan al subprograma que calcule parciales de 1er. orden y de 2° orden que serían:

- A - Z ---- controlado por X1 - de primer orden
- A - Z ---- controlado por X2 - de primer orden
- A - Z ---- controlado por X1,X2 - de 2° orden

(d)

```

1           16
PARTIAL CORR A WITH Z BY VAR001 TO VAR004 (1,2,3)

```

Esta solicitud de PARTIAL CORR produciría el siguiente conjunto de 14 coeficientes de correlación parcial:

<u>Primer orden</u>	<u>Segundo Orden</u>	<u>Tercer orden</u>
A con Z	A con Z	A con Z
BY	BY	BY
VAR001	VAR001,VAR002	VAR001,VAR002,VAR003
VAR002	VAR001,VAR003	VAR001,VAR002,VAR004
VAR003	VAR001,VAR004	VAR001,VAR003,VAR004
VAR004	VAR002,VAR003	VAR002,VAR003,VAR004
	VAR002,VAR004	
	VAR003,VAR004	
4	6	4

Las solicitudes de los ejemplos c,d, podrían ser realizadas con una sola especificación, que sería:

```

1           16
PARTIAL CORR A WITH Z BY X1,X2 (1,2) /
A WITH Z BY VAR001 TO VAR004 (1,2,3)

```

XV.2 Convenciones para ingresar matriz de coeficientes de correlación a PARTIAL CORR

Tanto el subprograma PARTIAL CORR como REGRESSION y FACTOR (Análisis Factorial) permiten al usuario ingresar con datos crudos (primarios) o con coeficientes de correlación simples. Estos

coeficientes pueden ser calculados por el subprograma PEARSON CORR o por el mismo PARTIAL CORR haciendo uso de las opciones que provocan la perforación (grabación) de la matriz de coeficientes de correlación.

El uso de las opciones 4 o 6 de PARTIAL CORR que informan al subprograma para que lea matrices de coeficientes es una forma de permitir al usuario una gran disminución en el tiempo de computador ya que no requiere la lectura de todo un archivo de datos, principalmente cuando el volumen de datos es relativamente alto. O sea, el usuario sólo requiere una lectura del archivo de datos crudos con la cual puede producir los coeficientes de correlación que utilizará para posteriores procesos.

Si no se especifica opción 4 o 6 el subprograma espera que el proceso se realice sobre un archivo de datos crudos.

Especificaciones para la matriz de correlación

1. Los coeficientes de correlación deben estar en formato 8F10.7; en registros lógicos en imagen de tarjeta.
2. La matriz de coeficientes puede estar almacenada en tarjetas, cintas o discos. En el caso que residan en cinta o disco, el usuario debe direccionar el archivo a través de la unidad lógica FT08.
3. Cada fila de la matriz debe comenzar en una nueva tarjeta.
4. La máxima precisión permitida es de 7 decimales.
5. La matriz completa de los coeficientes debe ser especificada -o sea, coeficientes redundantes y no redundantes.

XV.2.1 Métodos para especificar el orden de las variables para lectura de matriz.

El subprograma ofrece al usuario dos métodos para especificar la estructura de la matriz de correlación; que son:

- a) Método estándar - seleccionado con el uso de la opción 4.
- b) Por VARIABLE LIST - seleccionado con la opción 6.

A. El método estándar: opción 4

El orden de los coeficientes en la matriz debe ser el mismo orden que las variables son declaradas en la tarjeta de

procedimiento de PARTIAL CORR, tomando en cuenta la lista de correlación y la lista de control. El uso de esta opción requiere que se especifique una tarjeta VARIABLE LIST cuya función única es definir los nombres de las variables.

El siguiente ejemplo ilustra el uso del método estándar:

```

1           16
RUN NAME      EJEMPLO DE ENTRADA CON MATRIZ
VARIABLE LIST D,E,F,A,C,B,X1,X2,X3,Z1,Z2,Z3
N OF CASES    3157
INPUT MEDIUM CARD
PARTIAL CORR  A,B WITH C,D BY E,F(1,2)
OPTIONS       4
READ MATRIX
.....
..... matriz 1 de coeficientes de correlación ...
.....
PARTIAL CORR  Z1 WITH Z2,Z3,X1 BY X2,X3 (2)
OPTIONS       4
READ MATRIX
.....
..... matriz 2 de coeficientes de correlación
.....
FINISH

```

La matriz 1 de coeficientes debe ser:

AA	AB	AC	AD	AE	AF
BA	BB	BC	BD	BE	BF
CA	CB	CC	CD	CE	CF
DA	DB	DC	DD	DE	DF
EA	EB	EC	ED	EE	EF
FA	FB	FC	FD	FE	FF

cuyo orden no guarda ninguna relación con el orden que fueron definidos en VARIABLE LIST pero sí en el orden que están especificados en la tarjeta de procedimiento.

La matriz 2 debe ser:

Z1,Z1	Z1,Z2	Z1,Z3	Z1,X1	Z1,X2	Z1,X3
Z2,Z1	Z2,Z2	Z2,Z3	Z3,X1	Z2,X2	Z2,X3
Z3,Z1	Z3,Z2	Z3,Z3	Z3,X1	Z3,X2	Z3,X3
X1,Z1	X1,Z2	X1,Z3	X1,X1	X1,X2	X1,X3
X2,Z2	X2,Z2	X2,Z3	X2,X1	X2,X2	X2,X3
X3,Z1	X3,Z2	X3,Z3	X3,X1	X3,X1	X3,X3

El READ MATRIX desarrolla una función similar a READ INPUT DATA, con la observación de que cada solicitud de procedimiento requiere después de la tarjeta OPTIONS (o después de STATISTICS si hay) una especificación de READ MATRIX; como se puede observar en el ejemplo.

Las dos matrices que se muestran podrían haber sido producidas por el PEARSON CORR siguiente:

```
1                16
PEARSON CORR  A,B,C,D,E,F / Z1,Z2,Z3,X1,X2,X3
OPTIONS      4
```

La declaración N OF CASES desarrolla la función de informar al subprograma que los coeficientes que se leen fueron calculados en base a 3157 observaciones.

Se debe tener cuidado con el uso de la convención "TO" para la declaración de las variables en VARIABLE LIST -si se usa el TO, se asume que el orden de las variables en la matriz es el mismo que el orden que se declara en la variable list. Para nuestro ejemplo será válido el uso del "TO" como se muestra.

```
VARIABLE LIST  D,E,F,A,C,B,Z1 TO Z3,X1 TO X3
```

Otro ejemplo similar al anterior, con el propósito de ilustrar la posición de los datos cuando se usa más de una lista de parciales es el siguiente:

```
1                16
PARTIAL CORR  A,B WITH C,D BY E,F (1,2) /
              Z1 WITH Z2,Z3,X1 BY X2,X3 (2)
OPTIONS      4
READ MATRIX
... matriz 1 ....
... matriz 2 ....
FINISH
```

En este ejemplo las declaraciones VARIABLE LIST, INPUT MEDIUM y N OF CASES son exactamente las mismas del primer ejemplo, así como los datos.

B. El método VARIABLE LIST: opción 6

El uso de esta opción requiere que el orden de las variables en la matriz sea el mismo que el orden en que están declarados en la VARIABLE LIST; y que la matriz traiga los coeficientes de todas las variables declaradas y en el orden que están declaradas.

Ejemplo:

```

1                16
RUN NAME         EJEMPLO DE PARTIAL CORR - OPCION 6
VARIABLE LIST   A,B,C,D,E,F
INPUT MEDIUM    TAPE
N OF CASES      3157
PARTIAL CORR    F WITH A,B,D BY C,E (1,2) /
                A WITH F BY B,C,D (1,2,3)
OPTIONS         6
READ MATRIX
PARTIAL CORR    A,B WITH E,F BY C,D (1,2) /
                A WITH D BY B,C,E,F (2,3,4)
OPTIONS         6
READ MATRIX
FINISH

```

Para este caso, el archivo en cinta, en imagen de tarjetas debe contener dos veces la matriz 1 que se describió en la parte A del método estándar, pues hay dos solicitudes de PARTIAL CORR y cada una debe llevar la declaración READ MATRIX.

Resumiendo, la forma del archivo en cinta debería ser para el ejemplo en cuestión:

AA	AB	AF	}	utilizada por la primera solicitud
BA				
				
				
FA	FB		FF	}	utilizada por la segunda solicitud
AA	AB	AF		
BA	BB	BF		
				
FA	FB	FF		

Se puede observar que este método sólo nos permite ingresar una única matriz de coeficientes pero con la ventaja de poder especificar varias listas de parciales en una solicitud, haciendo combinaciones de las variables en juego, sinque la especificación de cada lista requiera disponer de una nueva matriz. En otras palabras, podemos decir que lo que en un método es ventajoso en el otro es desventajoso y vice-versa; y la selección del método a utilizar dependerá de cada aplicación.

El uso de la convención "TO" para variable list es válido en la forma "alfaxxx TO alfayyy".

XV.2.2 Tarjetas del SPSS que son requeridas para el ingreso de matrices.

Cualquiera que sea el método utilizado, el sistema requiere que se especifiquen las siguientes tarjetas para el procesamiento de PARTIAL CORR utilizando matrices:

- 1 VARIABLE LIST: la función es definir los nombres de variables; el orden depende del método utilizado.
2. INPUT MEDIUM: informa al sistema, el medio donde se encuentran almacenadas las matrices a ser utilizadas. Si se especifica TAPE o DISK se debe direccionar el archivo por medio de la unidad lógica FT08.
3. N OF CASES: informa al sistema sobre el número de caso utilizado en el cálculo de los coeficientes.
4. OPTIONS: especificando 4 o 6 según sea el caso.
5. READ MATRIX: se coloca después de OPTIONS (o STATISTICS si hay) y debe ser especificada para cada solicitud PARTIAL CORR.
6. SUBFILE LIST: condicional. Es utilizada en el caso que las matrices fueran generadas en base a un archivo con estructura de subarchivos y hay tantas matrices como subarchivos (o combinaciones). Si se especifica SUBFILE LIST no se debe especificar N OF CASES y además se debe poner la declaración RUN SUBFILES de acuerdo con el proceso que produjo las matrices.

XV.3 Opciones disponibles para PARTIAL CORR

Opción estándar: eliminación de missing por lista. El tratamiento estándar para los casos con missing es la eliminación por lista -esto es, un caso será eliminado del cálculo si alguna de las variables en las listas (de correlación o de control) trae un valor missing. Esta opción es utilizada por defecto, a menos que se utilicen las opciones 1 o 2.

- OPCION 1: inclusión de los missing. Esta opción provoca que se ignore la declaración de MISSING y considere to dos los casos como válidos.
- OPCION 2: eliminación de los missing por pares. Un caso será eliminado del cálculo si el valor de alguna de las dos variables para las cuales se está realizando el cómputo es missing.
- OPCION 3: test de significancia de dos colas. La selección de esta opción provoca que se aplique el test de dos colas sobre cada coeficiente en lugar del test de un cola que es utilizado en forma estándar.
- OPCION 4: Entrada de matriz de coeficientes, por el método estándar.
- OPCION 5: Perfora (graba) matriz de coeficientes. Esta opción provoca la perforación (grabación o impresión) de una matriz de coeficientes de correlación, tomando todas las variables de la lista (lista de correlación y de control). Esta opción es válida tanto para listas de correlación con WITH o en la forma matricial. Así la lista "A,B WITH C,D BY E,F" produciría una matriz cuadrada idéntica a la que produciría "PEARSON CORR A,B,C,D,E,F" e idéntica a la que se produciría con PARTIAL CORR A,B,C,D BY E,F. El formato y forma de grabación es idéntico a lo ya descrito para PEARSON CORR. Se requiere especificar la unidad lógica FT09 o aquella que se especifique en RAW OUTPUT UNIT.
- OPCION 6: entrada de matrices de coeficientes, por el método VARIABLE LIST.
- OPCION 7: elimina la impresión de los grados de libertad y la significancia.
- OPCION 8: impresión de coeficientes no redundantes, en formato de hilera. Si no se usa esta opción el formato estándar en forma matricial es utilizado.

Opciones estándares que asume el sistema si no se especifica
OPTIONS:

- a) eliminación de los missing por lista.
- b) aplica el test de 1 cola
- c) espera lectura de datos crudos y no matrices.
- d) no perfora matrices de correlación
- e) grados de libertad y significancia son impresos
- f) los coeficientes parciales son impresos en formato matricial.

XV.4 Estadísticos disponibles

Estadístico 1: informa al sistema para que imprima los coeficientes de correlación simples (parcial de orden cero) utilizados en el cálculo de las correlaciones parciales.

Estadístico 2: imprimirá la media y la desviación estándar de todas las variables. El cálculo de estos dos estadísticos se realiza en base al mismo número de casos utilizados para el cálculo de los coeficientes parciales si se utilizó la opción estándar de eliminación de missing por lista. Para la opción de eliminación por pares, el cálculo se realiza sobre el número de casos no-missing de cada variable.

Estadístico 3: imprime los coeficientes de correlación simples (igual a estadístico 1) sólo si existe algún coeficiente simple que no fue posible calcular. Si se pide estadístico 1 y 3, prevalecerá el estadístico 1.

XV.5 Limitaciones para PARTIAL CORR

- 1) Máximo de listas = 25
- 2) Máximo de variables = 400
Si una misma variable es nombrada en dos listas debe ser contada dos veces. Así, cada ocurrencia de una variable se contabiliza UNO para esta limitación.
- 3) Un máximo de 5 diferentes órdenes (excluyendo orden cero) de parciales puede ser solicitado. Coeficientes parciales de orden superior a la 5° pueden ser solicitados, pero el número de valores de orden no debe ser mayor que cinco valores.
- 4) Requerimientos de memoria.
 - a) Para la opción estándar de tratamiento de los missing u opción 1:

$$\text{WORKSPACE} = 24 * NV + 8 * \sum N_i^2$$

- b) Si se usa opción 4:

$$\text{WORKSPACE} = 8 * \sum N_i^2$$

c) si se usa opción 2 sin estadístico 2

$$\text{WORKSPACE} = 24 * \sum N_i^2$$

d) si se usa opción 2 con estadístico 2

$$\text{WORKSPACE} = 24 * \text{NV} + 24 * \sum N_i^2$$

donde

NV = número total de variables de todas las listas. Cada ocurrencia contabiliza UNO al valor de NV.

N_i = total de variables en la lista i. La sumatoria es para las n listas.

XV.6. Ejemplos de salida de PARTIAL CORR

Ejemplo a. En esta pagina y en la 204 se muestra la salida con las opciones estandares y STATISTICS ALL. En la solicitud de PARTIAL CORR se pidió parciales de primer y 2º orden. En PAGE 2, se muestra la solicitud de PARTIAL CORR; en "page 3", salen las medias y desviaciones; en "page 4", los parciales de orden cero; en "page 6,7,8" los parciales de primer y 2º orden.

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR - 09R05R76 PAGE 2
 PARTIAL CORR VAR068,VAR053 WITH VAR004,VAR007,VAR008,VAR010,VAR014,VAR015,
 STATISTICS VAR022,VAR025,VAR027 BY VAR083,VAR084(1,2)
 ALL

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR - 09R05R76 PAGE 3
 USO DE OPCIONES ESTANDARES Y TODOS LOS ESTADISTICOS
 FILE FESALEMB (CREATION DATE = 08R24R76)

VARIABLE	MEAN	STANDARD DEV	CASES
VAR068	4.5137	2.9013	1754
VAR053	0.3843	1.4906	1754
VAR004	0.6129	0.7755	1754
VAR007	0.5644	0.6459	1754
VAR008	0.3687	0.6631	1754
VAR010	0.6482	0.7739	1754
VAR014	0.3775	0.7137	1754
VAR015	0.4116	0.9140	1754
VAR022	0.3387	0.6149	1754
VAR025	0.4418	0.6858	1754
VAR027	0.6596	0.7068	1754
VAR083	8.9453	8.2165	1754
VAR084	18.4253	3.7431	1754

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR - 09R05R76 PAGE 4
 USO DE OPCIONES ESTANDARES Y TODOS LOS ESTADISTICOS
 FILE FESALEMB (CREATION DATE = 08R24R76)

----- PARTIAL CORRELATION COEFFICIENTS -----

ZERO ORDER PARTIALS

	VAR068	VAR053	VAR004	VAR007	VAR008	VAR010	VAR014	VAR015	VAR022	VAR025
VAR068	1.0000 (\$=0.001)	-0.0388 (\$=0.052)	-0.1920 (\$=0.001)	-0.1225 (\$=0.001)	-0.1393 (\$=0.001)	-0.1870 (\$=0.001)	-0.2775 (\$=0.001)	-0.0497 (\$=0.019)	0.1282 (\$=0.001)	-0.2927 (\$=0.001)
VAR053		1.0000 (\$=0.001)	0.2304 (\$=0.001)	0.1876 (\$=0.001)	0.1997 (\$=0.001)	0.2072 (\$=0.001)	0.2460 (\$=0.001)	0.1531 (\$=0.001)	-0.0288 (\$=0.114)	0.2283 (\$=0.001)
VAR004			1.0000 (\$=0.001)	0.6005 (\$=0.001)	0.6621 (\$=0.001)	0.8756 (\$=0.001)	0.4449 (\$=0.031)	0.3053 (\$=0.001)	0.0239 (\$=0.159)	0.5460 (\$=0.001)
VAR007				1.0000 (\$=0.001)	0.8711 (\$=0.001)	0.6462 (\$=0.001)	0.4143 (\$=0.001)	0.2043 (\$=0.001)	0.0054 (\$=0.411)	0.4283 (\$=0.001)
VAR008					1.0000 (\$=0.001)	0.7055 (\$=0.001)	0.4535 (\$=0.001)	0.2292 (\$=0.001)	-0.0107 (\$=0.328)	0.4702 (\$=0.001)
VAR010						1.0000 (\$=0.001)	0.4315 (\$=0.001)	0.3044 (\$=0.001)	0.0239 (\$=0.158)	0.5308 (\$=0.001)
VAR014							1.0000 (\$=0.001)	0.2264 (\$=0.001)	-0.1474 (\$=0.001)	0.6376 (\$=0.001)
VAR015								1.0000 (\$=0.001)	0.0609 (\$=0.005)	0.2435 (\$=0.001)
VAR022									1.0000 (\$=0.001)	-0.0723 (\$=0.001)
VAR025										1.0000 (\$=0.001)
VAR027										
VAR083										
VAR084										

(COEFFICIENT A (C.F.) IN SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR -
USO DE OPCIONES ESTANDARES Y TODOS LOS ESTADISTICOS
FILE PESALEMB (CREATION DATE = 08R24R76)

09R05R76

PAGE 6

----- P A R T I A L C O R R E L A T I O N C O E F F I C I E N T S -----

CONTROLLING FOR.. VAR083

	VAR004	VAR007	VAR008	VAR010	VAR014	VAR015	VAR022	VAR025	VAR027
VAR068	-0.2145 (.1751) \$=0.001	-0.1353 (.1751) \$=0.001	-0.1563 (.1751) \$=0.001	-0.2146 (.1751) \$=0.001	-0.1947 (.1751) \$=0.001	-0.1186 (.1751) \$=0.001	-0.0922 (.1751) \$=0.001	-0.2213 (.1751) \$=0.001	-0.1108 (.1751) \$=0.001
VAR053	0.2314 (.1751) \$=0.001	0.1881 (.1751) \$=0.001	0.2003 (.1751) \$=0.001	0.2080 (.1751) \$=0.001	0.2225 (.1751) \$=0.001	0.1529 (.1751) \$=0.001	-0.0313 (.1751) \$=0.095	0.2343 (.1751) \$=0.001	0.2213 (.1751) \$=0.001

(COEFFICIENT R (D.F.) R SIGNIFICANCE)

(A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR -
USO DE OPCIONES ESTANDARES Y TODOS LOS ESTADISTICOS
FILE PESALEMB (CREATION DATE = 08R24R76)

09R05R76

PAGE 7

----- P A R T I A L C O R R E L A T I O N C O E F F I C I E N T S -----

CONTROLLING FOR.. VAR084

	VAR004	VAR007	VAR008	VAR010	VAR014	VAR015	VAR022	VAR025	VAR027
VAR068	-0.1764 (.1751) \$=0.001	-0.1112 (.1751) \$=0.001	-0.1246 (.1751) \$=0.001	-0.1797 (.1751) \$=0.001	-0.2231 (.1751) \$=0.001	-0.0327 (.1751) \$=0.085	0.1122 (.1751) \$=0.001	-0.2759 (.1751) \$=0.001	-0.0976 (.1751) \$=0.001
VAR053	0.2297 (.1751) \$=0.001	0.1866 (.1751) \$=0.001	0.1987 (.1751) \$=0.001	0.2063 (.1751) \$=0.001	0.2467 (.1751) \$=0.001	0.1518 (.1751) \$=0.001	-0.0268 (.1751) \$=0.131	0.2278 (.1751) \$=0.001	0.2198 (.1751) \$=0.001

(COEFFICIENT R (D.F.) R SIGNIFICANCE)

(A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR -
USO DE OPCIONES ESTANDARES Y TODOS LOS ESTADISTICOS
FILE PESALEMB (CREATION DATE = 08R24R76)

09R05R76

PAGE 8

----- P A R T I A L C O R R E L A T I O N C O E F F I C I E N T S -----

CONTROLLING FOR.. VAR083 VAR084

	VAR004	VAR007	VAR008	VAR010	VAR014	VAR015	VAR022	VAR025	VAR027
VAR068	-0.2039 (.1750) \$=0.001	-0.1273 (.1750) \$=0.001	-0.1461 (.1750) \$=0.001	-0.2034 (.1750) \$=0.001	-0.1769 (.1750) \$=0.001	-0.1064 (.1750) \$=0.001	-0.1037 (.1750) \$=0.001	-0.2095 (.1750) \$=0.001	-0.0978 (.1750) \$=0.001
VAR053	0.2309 (.1750) \$=0.001	0.1871 (.1750) \$=0.001	0.1992 (.1750) \$=0.001	0.2070 (.1750) \$=0.001	0.2227 (.1750) \$=0.001	0.1516 (.1750) \$=0.001	-0.0297 (.1750) \$=0.107	0.2335 (.1750) \$=0.001	0.2204 (.1750) \$=0.001

(COEFFICIENT R (D.F.) R SIGNIFICANCE)

(A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

Ejemplo b. este ejemplo ilustra la salida producida bajo el efecto de la opcion 7, que suprime la impresion de grados de libertad y de la significancia. En forma similar a todos los ejemplos se muestra la solicitud del procedimiento y la salida impresa.

TASK NAME USD DE OPCION 7 - NO IMPRIME NI GRADOS DE LIBERTAD NI SIGNIF.
PARTIAL CORR VAR008,VAR009 WITH VAR004,VAR007,VAR008,VAR010,VAR014,VAR015,
OPTIONS VAR022,VAR025,VAR027 BY VAR003,VAR008411,20

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR -
USD DE OPCION 7 - NO IMPRIME NI GRADOS DE LIBERTAD NI SIGNIF.
FILE PESALEMB (CREATION DATE = 08R24R76)

09R05R76

PAGE 10

----- PARTIAL CORRELATION COEFFICIENTS -----

CONTROLLING FOR.. VAR003

	VAR004	VAR007	VAR008	VAR010	VAR014	VAR015	VAR022	VAR025	VAR027
VAR008	-0.2145**	-0.1253**	-0.1363**	-0.2146**	-0.1947**	-0.1186**	-0.0922**	-0.2213**	-0.1198**
VAR009	0.2314**	0.1881**	0.2003**	0.2080**	0.2525**	0.1529**	-0.0313	0.2343**	0.2213**

* - SIGNIF. LE .01

** - SIGNIF. LE .001

(99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR -
USD DE OPCION 7 - NO IMPRIME NI GRADOS DE LIBERTAD NI SIGNIF.
FILE PESALEMB (CREATION DATE = 08R24R76)

09R05R76

PAGE 11

----- PARTIAL CORRELATION COEFFICIENTS -----

CONTROLLING FOR.. VAR004

	VAR004	VAR007	VAR008	VAR010	VAR014	VAR015	VAR022	VAR025	VAR027
VAR008	-0.1764**	-0.1112**	-0.1246**	-0.1707**	-0.2531**	-0.0327	0.1122**	-0.2759**	-0.0976**
VAR009	0.2297**	0.1866**	0.1987**	0.2063**	0.2467**	0.1516**	-0.0268	0.2278**	0.2198**

* - SIGNIF. LE .01

** - SIGNIF. LE .001

(99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR -
USD DE OPCION 7 - NO IMPRIME NI GRADOS DE LIBERTAD NI SIGNIF.
FILE PESALEMB (CREATION DATE = 08R24R76)

09R05R76

PAGE 12

----- PARTIAL CORRELATION COEFFICIENTS -----

CONTROLLING FOR.. VAR003 VAR004

	VAR004	VAR007	VAR008	VAR010	VAR014	VAR015	VAR022	VAR025	VAR027
VAR008	-0.2039**	-0.1273**	-0.1461**	-0.2034**	-0.1769**	-0.1064**	-0.1037**	-0.2095**	-0.0978**
VAR009	0.2305**	0.1871**	0.1992**	0.2070**	0.2527**	0.1516**	-0.0297	0.2335**	0.2204**

* - SIGNIF. LE .01

** - SIGNIF. LE .001

(99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

Ejemplo c. este ejemplo ilustra la salida que produce el uso de la opcion 8 ; impresion en formato de hilera:

TASK NAME USD DE OPCION 8 - COEFICIENTES NO REDUNDANTES EN FORMATO HILERA
 PARTIAL CORR VAR068,VAR003 WITH VAR004,VAR007,VAR008,VAR010,VAR014,VAR015,
 OPTIONS VAR022,VAR025,VAR027 BY VAR003,VAR004(1,2)

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR - 09R05R76 PAGE 14
 USD DE OPCION 8 - COEFICIENTES NO REDUNDANTES EN FORMATO HILERA
 FILE PESALEM8 (CREATION DATE = 08R24R76)

----- PARTIAL CORRELATION COEFFICIENTS -----

CONTROLLING FOR.. VAR083

VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR
VAR068 WITH VAR004	VAR068 WITH VAR007	VAR068 WITH VAR008	VAR068 WITH VAR010	VAR068 WITH VAR014	VAR068 WITH VAR015
-0.2145 DF = 1751 SIG .001	-0.1353 DF = 1751 SIG .001	-0.1563 DF = 1751 SIG .001	-0.2146 DF = 1751 SIG .001	-0.1947 DF = 1751 SIG .001	-0.1186 DF = 1751 SIG .001
VAR068 WITH VAR022	VAR068 WITH VAR025	VAR068 WITH VAR027	VAR053 WITH VAR004	VAR053 WITH VAR007	VAR053 WITH VAR008
-0.0922 DF = 1751 SIG .001	-0.2213 DF = 1751 SIG .001	-0.1108 DF = 1751 SIG .001	0.2314 DF = 1751 SIG .001	0.1881 DF = 1751 SIG .001	0.2003 DF = 1751 SIG .001
VAR053 WITH VAR010	VAR053 WITH VAR014	VAR053 WITH VAR015	VAR053 WITH VAR022	VAR053 WITH VAR025	VAR053 WITH VAR027
0.2080 DF = 1751 SIG .001	0.2525 DF = 1751 SIG .001	0.1529 DF = 1751 SIG .001	-0.0313 DF = 1751 SIG .005	0.2343 DF = 1751 SIG .001	0.2213 DF = 1751 SIG .001

A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR - 09R05R76 PAGE 15
 USD DE OPCION 8 - COEFICIENTES NO REDUNDANTES EN FORMATO HILERA
 FILE PESALEM8 (CREATION DATE = 08R24R76)

----- PARTIAL CORRELATION COEFFICIENTS -----

CONTROLLING FOR.. VAR084

VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR
VAR068 WITH VAR004	VAR068 WITH VAR007	VAR068 WITH VAR008	VAR068 WITH VAR010	VAR068 WITH VAR014	VAR068 WITH VAR015
-0.1764 DF = 1751 SIG .001	-0.1112 DF = 1751 SIG .001	-0.1246 DF = 1751 SIG .001	-0.1707 DF = 1751 SIG .001	-0.2531 DF = 1751 SIG .001	-0.0327 DF = 1751 SIG .085
VAR068 WITH VAR022	VAR068 WITH VAR025	VAR068 WITH VAR027	VAR053 WITH VAR004	VAR053 WITH VAR007	VAR053 WITH VAR008
0.1122 DF = 1751 SIG .001	-0.2759 DF = 1751 SIG .001	-0.0976 DF = 1751 SIG .001	0.2297 DF = 1751 SIG .001	0.1866 DF = 1751 SIG .001	0.1987 DF = 1751 SIG .001
VAR053 WITH VAR010	VAR053 WITH VAR014	VAR053 WITH VAR015	VAR053 WITH VAR022	VAR053 WITH VAR025	VAR053 WITH VAR027
0.2063 DF = 1751 SIG .001	0.2467 DF = 1751 SIG .001	0.1518 DF = 1751 SIG .001	-0.0268 DF = 1751 SIG .131	0.2278 DF = 1751 SIG .001	0.2198 DF = 1751 SIG .001

A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

EJEMPLO DEL PROCEDIMIENTO - PARTIAL CORR -
USO DE OPCIÓN 8 - COEFICIENTES NO REDUNDANTES EN FORMATO MILERA
FILE PESALEMB (CREATION DATE = 08R24R76)

09R05R76

PAGE 16

----- PARTIAL CORRELATION COEFFICIENTS -----

CONTROLLING FOR.. VAR003 VAR004

VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR		VARIABLE PAIR	
VAR068 WITH VAR004	-0.2039 DF = 1750 SIG .001	VAR068 WITH VAR007	-0.1273 DF = 1750 SIG .001	VAR068 WITH VAR008	-0.1461 DF = 1750 SIG .001	VAR068 WITH VAR010	-0.2034 DF = 1750 SIG .001	VAR068 WITH VAR014	-0.1769 DF = 1750 SIG .001	VAR068 WITH VAR015	-0.1064 DF = 1750 SIG .001
VAR068 WITH VAR022	-0.1937 DF = 1750 SIG .001	VAR068 WITH VAR025	-0.2095 DF = 1750 SIG .001	VAR068 WITH VAR027	-0.0978 DF = 1750 SIG .001	VAR053 WITH VAR004	0.2305 DF = 1750 SIG .001	VAR053 WITH VAR007	0.1871 DF = 1750 SIG .001	VAR053 WITH VAR008	0.1992 DF = 1750 SIG .001
VAR053 WITH VAR010	0.2070 DF = 1750 SIG .001	VAR053 WITH VAR014	0.2927 DF = 1750 SIG .001	VAR053 WITH VAR015	0.1516 DF = 1750 SIG .001	VAR053 WITH VAR022	-0.0297 DF = 1750 SIG .107	VAR053 WITH VAR025	0.2335 DF = 1750 SIG .001	VAR053 WITH VAR027	0.2204 DF = 1750 SIG .001

A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

XVI. REGRESSION: ANALISIS DE REGRESION MULTIPLE

XVI.1. La tarjeta de procedimiento

Formato General:

<u>1</u> REGRESSION	<u>16</u> VARIABLES= lista de vars / REGRESSION= lista de regresion .../ REGRESSION= lista de regresion / VARIABLES= lista de vars / REGRESSION= lista de regresion/ REGRESSION= lista de regresion
------------------------	--

En todas las listas de regresión después de la palabra clave "REGRESSION" sólo se puede hacer referencia a las variables declaradas en la parte "VARIABLES =" que antecede.

Ignorando las repeticiones que permite una solicitud de procedimiento y especificando en mayor detalle el formato sería:

<u>1</u> REGRESSION	<u>16</u> VARIABLES= lista de variables / REGRESSION= variable dependiente { (parametros) } WITH lista ₁ vars independientes (nivel de inclusion) , lista ₂ vars indep (nivel inclusion) lista _n vars independientes (nivel inclusion) , { RESID= nump }
------------------------	--

La parte encerrada entre llaves, es opcional.

Para una ilustración sobre el formato de especificación veamos dos ejemplos iniciales:

```
1          16
REGRESSION  VARIABLES = A,B,VAR001 TO VAR010, X1,Z1 /
            REGRESSION = A WITH B, VAR001 (4),
            VAR002 TO VAR006 (3), VAR007 TO Z1 (1) /
            REGRESSION = VAR001 WITH VAR002 TO VAR007 (2)
```

Este primer ejemplo se omitió todo lo que es opcional. Cabe observar que la especificación de un nivel de inclusión tiene efecto sobre todas las variables independientes que lo preceden.

Un ejemplo con las especificaciones opcionales y por lo tanto modificando los valores estándares que asume el sistema en caso de omisión sería:

```
1          16
REGRESSION  VARIABLES = A,B,C,D, X1,Z1,X2,Z2 /
            REGRESSION = A (10,1.5,0.1) WITH B TO X1 (3),
            Z1,X2,Z2 (5),
            RESID = 0.4
```

XVI.2 El nivel de inclusión: definición de la modalidad de cálculo

El subprograma REGRESSION permite que el usuario tenga un control total sobre la forma en que desea que el cálculo de la regresión se realice. Este control se ejerce por medio de la manipulación de los niveles de inclusión que a su vez de finen el modo de cálculo de la ecuación.

- a) Modalidad simple: esta modalidad de cálculo es seleccionada si se especifica como nivel de inclusión un número par.
- b) Modalidad paso a paso: esta modalidad es seleccionada por niveles de inclusión impares.

Rango de los niveles de inclusión - desde 1 a 99 y la mayor jerarquía corresponde al nivel 99, la menor al nivel 1.

Como se puede observar en el formato de la tarjeta de procedimiento, el subprograma nos permite especificar niveles de inclusión distintos para cada variable o sublistas de las variables dependientes. Este aspecto nos permite tener un control total sobre la forma en que deseamos que las variables independientes sean incluidas en la ecuación de regresión; o sencillamente dejar que el subprograma seleccione automáticamente las variables que contribuyen en forma significativa a la ecuación.

Definición de las modalidades:

SIMPLES: al seleccionar esta modalidad, estamos informando al sistema que aquellas variables que traen un número par como nivel de inclusión deben ser incluidas a la ecuación de regresión sin importar si es significativo o no el aporte de tales variables a la ecuación.

PASO A PASO: esta modalidad provoca que de una variable o una sublista de variables solamente deben ser incluidas aquellas variables que tienen un aporte significativo sobre la ecuación. Así, de una lista de variables con un mismo nivel de inclusión impar el sistema podrá no incluir ninguna, incluir sólo algunas o incluir todas a la ecuación. Esta modalidad además, provoca que el sistema imprima una ecuación de regresión para cada variable que es incluida en la ecuación, o sea cada paso significa la inclusión de una variable.

Ejemplos:

(a)

1	16
REGRESSION	VARIABLES = VAR001 TO VAR010 / REGRESSION = VAR001 WITH VAR002 (2), VAR003 (6), VAR004, VAR005 (4)

Este ejemplo, a pesar de que todos los niveles de inclusión son pares y por lo tanto todas las variables serán incluidas a la ecuación, generará tres pasos pues los valores de los niveles de inclusión son distintos. Los pasos generados, produciendo cada uno una ecuación serían:

- PASO 1: VAR001 = f (VAR003) pues nivel = 6 (más alto)
- PASO 2: VAR001 = f (VAR003,VAR004,VAR005)
- PASO 3: VAR001 = f (VAR003,VAR004,VAR005,VAR002)

Así tendríamos tres ecuaciones distintas y el orden en que son producidas está determinado por los niveles de inclusión y no por el orden en que están puestas las variables en la lista.

(b)

1	16
REGRESSION	VARIABLES = VAR001 TO VAR010 / REGRESSION = VAR001 WITH VAR002 TO VAR005 (7), VAR008, VAR009 (4), VAR010 (6)

En este ejemplo tenemos las dos modalidades y como el nivel = 7 es el más alto, el sistema inicia el cálculo paso a paso con variables VAR002, VAR003, VAR004, VAR005. El criterio que utiliza el subprograma para la inclusión de las variables es el siguiente: primero incluye aquella variable que tiene una mayor correlación con la variable dependiente y que su aporte a la ecuación es significativo; segundo incluye (de las variables restantes) aquella que conjuntamente con la variable que ya está en la ecuación aporta una contribución significativa a la ecuación, y así sucesivamente.

Para la solicitud de este ejemplo y en forma de supuesto podríamos tener los siguientes pasos: (se denotará V1 para VAR001, V2 para VAR002, etc. para simplificar la notación).

PASO 1: V1 = f (V4)

PASO 2: V1 = f (V4, V3)

PASO 3: V1 = f (V4, V3, V5) V2 se excluye

PASO 4: V1 = f (V4, V3, V5, V10) pues nivel = 6 > 4

PASO 5: V1 = f (V4, V3, V5, V10, V8, V9)

El nivel de inclusión cero

La especificación para una sublista de variables de un nivel de inclusión con el valor 0 (cero), provocará que dichas variables nunca serán incluidas en la ecuación pero permite al usuario obtener para dichas variables una serie de coeficientes que aparecen en la parte de la salida impresa bajo el título de "--VARIABLES NOT IN THE EQUATION--"

XVI.3 Los parámetros para el criterio de inclusión

El criterio de inclusión para la modalidad paso a paso se realiza en base a los parámetros que se especifican entre paréntesis a la par de la variable dependiente:

1	16
REGRESSION	VARIABLES = A,B,C,D,E,F / REGRESSION = A (n,F,T)

Donde

n = número de pasos máximo
F = F estadístico
T = tolerancia

Los valores que asume el sistema si no son especificados son:

n = 80
F = 0.01
T = 0.001

Si el usuario desea modificar alguno de estos parámetros deberá especificar necesariamente aquellos que lo preceden.
Ejemplos:

- Cambio sólo de n

1	16
REGRESSION	VARIABLES = ----- / REGRESSION = Z1 (30) WITH -----

- Cambio sólo de F (manteniendo n = 80)

1	16
REGRESSION	VARIABLES = ---- / REGRESSION = Z1 (80,0.5) WITH

- Cambio sólo de T

1	16
REGRESSION	VARIABLES = ---- / REGRESSION = Z1 (80,0.01,0.30) WITH

XVI.4 RESID: Tratamiento de los missing

Por medio de RESID = mdrp (missing-data replacement proportion) el sistema nos permite cambiar la proporción para la inclusión de casos con valores missing. Si se usa la opción estándar de tratamiento de los missing -eliminación por lista, el RESID debe ser cero o no debe ser especificado con el cual el sistema asume RESID = 0

La opción 2 -eliminación de los missing por pares-, provoca que un caso sea eliminado del cálculo si alguna de las variables independientes trae un valor missing. Con el uso de esta opción, el sistema nos permite por medio de RESID una proporción que indicará que un caso sea incluido del cálculo si el número de variables independientes es superior a la proporción especificada.

Un ejemplo sería:

1	16
	REGRESSION = Y WITH VAR001 TO VAR010 (4),
	RESID = 0.30

NOTA: El valor de mdrp debe variar entre 0 y 1

En este ejemplo, RESID = 0.30 indica al sistema que todos los casos que tengan 3 o menos variables independientes (tenemos 10 variables -VAR001 TO VAR010 y 3 = 30%) con valores declarados como missing, que sean incluidos en el cálculo de los predictores y residuos. En este ejemplo un caso será excluido del cálculo si simultáneamente 4 o más variables independientes (mayor que 30%) traen valores missing.

El SPSS permite dos métodos de estimar los valores en los casos que son missing.

- a) Método estándar: reemplaza los valores missing por la media.
- b) Método ponderado: este método es activado por el uso de la opción 13. Así llamando:
 - Y' = predictor estandarizado ponderado.
 - N = número total de variables independientes en la ecuación.
 - NNM = número de variables con valores no missing (varía para cada caso).

La fórmula queda:

$$Y' = \frac{N}{NNM} (\sum \beta_i Z_i)$$

La sumatoria se realiza sobre todas las variables con valores no missing.

El residuo es calculado a partir de v'

XVI.5 Convenciones para ingreso de matrices

En forma similar al subprograma PARTIAL CORR, el procedimiento REGRESSION permite la entrada tanto de datos crudos o matrices de coeficientes de correlación simples y las medias y desviaciones estándares. Las matrices pueden ser producidas por PEARSON CORR, PARTIAL CORR y por el mismo subprograma REGRESSION.

Las opciones para leer matrices son:

OPCION 4: sólo leerá las matrices de coeficientes de correlación y asume media = 0 y desviación estándar = 1.0 para todas las variables.

OPCION 4,5: requiere la especificación de:

- fila con las medias
- fila con desviaciones estándares
- matriz de coeficientes.

Formato de las matrices, medias y desviaciones

El formato de las matrices de coeficientes es exactamente igual a lo descrito en PARTIAL CORR.

Para las medias y desviaciones estándares, el formato es 8F10.0. Esto implica que es recomendable la perforación de los puntos decimales y se puede especificar tantos decimales como se requiera dentro del campo de 10 columnas ya que el punto perforado prevalece sobre el formato de lectura. Si no se perforan puntos decimales los valores deben ser ajustados a la derecha del campo.

Métodos para el orden de las variables

Método estándar: el orden de las variables, tanto de medias, desviaciones y coeficientes debe ser el mismo que el orden en que las variables son declaradas en la parte "VARIABLES =" de la tarjeta. Se debe especificar medias, desviaciones y matrices para cada ocurrencia de "VARIABLES =".

Método VARIABLE LIST: opción 9, (junto con 4 o 4,5). En este caso el orden de las variables debe ser el mismo en que son declaradas en la tarjeta VARIABLE LIST.

Tarjetas SPSS requeridas

1. VARIABLE LIST
2. INPUT MEDIUM
3. N OF CASES (SUBFILE LIST)
4. READ MATRIX
5. OPTIONS con (4); (4,5); (4,9); (4,5,9)

A excepción de la especificación de las medias y desviaciones estándares y que el orden se refiere al orden en "VARIABLES =" (método estándar) todo lo demás, referente a la entrada con matrices es idéntico a lo ya descrito para PARTIAL CORR por lo cual se recomienda que el lector se reporte a dicho subprograma para un mayor detalle.

XVI.6 Opciones disponibles para REGRESSION

OPCION ESTANDAR: Eliminación de missing por lista. Esta opción es utilizada a menos que se especifique opción 1 ó 2. Un caso es excluido de todos los cálculos si alguna variable de la lista "VARIABLES =" trae un valor missing. Esto garantiza que todos los cálculos: media, desviación estándar y coeficientes de correlación son elaborados en base al mismo número de casos.

- OPCION 1: inclusión de los missing. Informa al sistema que considere todos los casos como válidos.
- OPCION 2: eliminación de missing por pares. Esta opción provoca que un caso en que el valor de una variable es missing, sea excluido de los cálculos que involucran dicha variable. El uso de esta opción puede provocar ciertas anomalías tales como coeficientes de correlación múltiple mayor que 1.0; F y suma de los cuadrados con valores negativos; motivo por el cual su uso debe ser cuidadoso.
- OPCION 3: elimina impresión de VAR LABELS.
- OPCION 4: entrada con matrices. El uso de esta opción puede combinarse con la opción 5 y 9. Si se usa sólo opción 4, el sistema sólo espera la especificación de matrices, asumiendo medias = 0 y desviación estándar = 1.0.
- OPCION 5: entrada de medias y desviaciones estándares. Debe ser usada junto con opción 4 e informa al sistema que antes de la matriz se encuentran especificadas:
- 1) las medias
 - 2) las desviaciones estándares
- OPCION 6: elimina impresión de los pasos. Solamente imprimirá la tabla sumaria.
- OPCION 7: elimina la impresión de la tabla sumaria.
- OPCION 8: grabación de matrices de coeficientes. Esta opción provoca la grabación (perforación) de las matrices de coeficientes de correlación simples. Grabará una matriz por cada declaración "VARIABLES =". El orden y las variables para las cuales graba la matriz son aquellas que son declaradas en "VARIABLES =". Esta opción utilizada conjuntamente con opción 15 (graba medias y desviaciones estándares) genera un archivo en que el orden es: medias, desviaciones y matriz; o sea en el orden que el subprograma requiere para la entrada de datos con el uso de opción 4,5. El uso de esta opción requiere direccionamiento de la unidad lógica FT09 (estándar) o la unidad que se especifique por RAW OUTPUT UNIT.
- OPCION 9: entrada de matrices por método VARIABLE LIST. Esta opción debe ser utilizada conjuntamente con opción 4 o 4 y 5.

.... opciones para analisis de residuos - 10 a 14

OPCION 10: provoca la grabación de información de secuencia para el uso de opciones 11, 12.

La siguiente información es grabada en columnas 1-20 de cada registro grabado:

col. 1 - 8 = SEQNUM

col. 9 - 10 = Clase de tarjeta (RECORD N)

col. 12 - 15 = 4 caracteres alfanuméricos de identificación del subarchivo (si hay).

col. 21 - 80 = 6 valores en formato 6F10.6.

Si no se usa opción 10, el formato de salida con el uso de opción 11 o 12 es 8F10.6.

OPCION 11: graba residuos estandarizados (normalizados).

Se grabará por FT09 (o la deseada) los residuos para cada caso y para cada especificación de "REGRESSION ="

OPCION 12: graba valores de Y' normalizados (estandarizado).

Si se usa opción 11 y 12, grabará residuo y Y' para cada caso. Tanto para 11 como 12 el formato de grabación es el 8F10.6 a no ser que se use con juntamente la opción 10.

OPCION 13: método ponderado para tratamiento de los missing.

Esta opción debe ser usada junto con opción 2 y cuando "RESID = mdrp" con mdrp > 0

OPCION 14: suprime impresión de ejes en los gráficos de los

valores de predicción normalizados versus residuos normalizados. Los gráficos son obtenidos por medio del estadístico 6.

OPCION 15: graba medias y desviaciones estándares. El forma

to que utiliza es 8F10.4. Puede ser utilizada so la o junto con opción 8.

XVI.7 Estadísticos disponibles para REGRESSION

Estadístico 1: provoca la impresión de las matrices de coefi
cientes de correlación simples. Se imprime
una matriz por cada ocurrencia de "variables =".

Estadístico 2: impresión de las medias, desviaciones estándares y número de casos válidos, para todas las variables especificadas en "VARIABLES =". Si está en efecto la opción 2 -eliminación por pares- el número de casos válidos de una variable dada, serán todos aquellos en que dicha variable es no missing. Para la opción estándar -eliminación por lista- el número de casos será igual para todas las variables ya que el caso se excluye del cálculo si alguna de las variables declarada en "VARIABLES =" trae un valor missing.

Estadístico 3: impresión de las matrices de coeficiente en casos anormales, o sea en el caso en que uno o más coeficientes simples no pudo ser calculado. Si se pide estadístico 1 y 3, el estadístico 1 prevalece y se imprimirán todas las matrices.

Estadístico 4: gráfico de residuos normalizados por caso, el gráfico es producido sólo para la última especificación de "REGRESSION =" con el uso de "RESID = 0". Sólo 500 casos serán graficados.

Estadístico 5: calcula el estadístico de DURBIN-WATSON para los residuos.

$$\text{Est. DURBIN - W} = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

e_i = residuo para el caso i .

n = número de casos.

Estadístico 6: produce un gráfico de residuos normalizados por Y' normalizado; con los residuos en el eje vertical.

Estadístico 7: Imprime las matrices de correlación y número de casos. Este estadístico debe ser utilizado en el caso que esté en efecto la opción 2. Si es solicitado con opción 1 o la opción estándar para el tratamiento de missing el sistema producirá la salida del estadístico 1. Si se solicita estadístico 1 y 7 sólo el 7 será considerado (en caso de uso de opción 2).

XVI.8 Limitaciones para REGRESSION

- 1) Máximo de 10 especificaciones de "VARIABLES ="
- 2) Máximo de 50 declaraciones "REGRESSION =", independiente del número de "VARIABLES =".
- 3) Máximo de 200 variables en todas "VARIABLES =". Variables que aparecen en más de una lista deben ser contabilizadas una por cada lista.
- 4) Máximo de 400 variables pueden aparecer en el total de "REGRESSION =". Cada ocurrencia de una misma variable se debe contabilizar como UNO para esta limitación.
- 5) Requerimientos de memoria

$$\boxed{\text{WORKSPACE} = 8 * (8 * \text{NVARs} + \text{MATRIX})}$$

Donde

NVARs = número de variables de todas "VARIABLES ="

Convencionando:

L = número de listas (VARIABLES =)

R = número de listas de regresión (REGRESSION =)

N_i = número de variables en la lista i-ésima de "VARIABLES =".

M_{ij} = número de variables de la j-ésima "REGRESSION =", referente a la i-ésima VARIABLES =

i = 1, 2, 3 ... L

j = 1, 2, 3 ... R para cada i

Tenemos que:

- a) con el uso de opción 2

$$\text{MATRIX} = 3 * \sum N_i^2 \quad i = 1, 2, \dots L$$

- b) opción estándar (por defecto) u opción 1

$$\text{MATRIX} = \sum N_i^2 + \max (M_{ij} * N_i)$$

Ejemplo de cálculo

i	N _i	N _i ²	Mij			Mij * Ni		
			j = 1	j = 2	j = 3	j = 1	j = 2	j = 3
1	6	36	6	0	0	36	0	0
2	10	100	8	10	5	80	100	50
3	20	400	10	12	15	200	240	<u>300</u>
4	12	144	5	6	10	60	72	<u>120</u>
5	4	16	4	0	0	16	0	0
	<u>52</u>	<u>696</u>						

$\max (Mij * Ni) = \underline{300}$, $NVARS = \underline{52}$

a) $MATRIX = 3 * N_i^2 = 3 * 696 = 2088$

$WORKSPACE = 8 * (8 * 52 + 2088) = \underline{20032 \text{ bytes.}}$

b) $MATRIX = N_i^2 + \max (Mij * Ni) = 696 + 300 = 996$

$WORKSPACE = 8 * (8 * 52 + 996) = \underline{11296 \text{ bytes}}$

6) Requerimientos de memoria si se usa "RESID ="

$WORKSP1 = 20 * V + 12 * I + 94 * N + (Q * 2072 * N)$

Donde:

V = número total de variables independientes, considerando todas "REGRESSION = " que se especifique "RESID ="

N = número de especificaciones "RESID = 0"

Q = 1 si se pide el estadístico 6.

Q = 0 si no se pide estadístico 6.

NOTA: Entre el valor calculado de WORKSPACE de la limitación 5 y el valor de WORKSP1 de la limitación 6 se debe considerar el que resulte mayor para efectos del ALLOCATE.

Si el espacio de memoria es menor que lo requerido para los gráficos del estadístico 6, el sistema sólo calcula los gráficos que permite el espacio dado.

- 7) Residuos no pueden ser calculados con la entrada de matrices.
- 8) Residuos y predictores normalizados por las opciones 11 y/o 12 se limitan al rango de -99.0 a + 99.0; todos los valores que exceden dicho rango son ajustados a los valores máximos del rango (-99, 99).

XVI.9 Descripción de la salida impresa de REGRESSION

En la parte de ejemplos de salida se puede ver el producto del subprograma REGRESSION, que básicamente se divide en dos partes:

- Parte A: resultados paso a paso (STEP)
- Parte B: tabla sumaria.

Parte A:

- 1) Indica la variable dependiente (nombre y var labels)
- 2) Indica las variables independientes que entran en el paso i (STEP i)
- 3) Para cada paso , antes de la parte
 - VARIABLES IN THE EQUATION - ; - VARIABLES NOT IN THE EQUATION -, publica un resumen estadístico para la ecuación de regresión producida en ese paso: R múltiplo, R cuadrado, R cuadrado ajustado y error estándar. (MULTIPLE R; R SQUARE; ADJUSTED A SQUARE y ERROR STANDARD) en la parte izquierda de la página; y en la parte derecha una tabla de análisis de varianza para la regresión y residuos.
- 4) - VARIABLES IN THE EQUATION - bajo este encabezado indica las variables ya incorporadas a la ecuación y sus coeficientes: B, BETA, y además el error estándar de B y el estadístico F.
- 5) - VARIABLES NOT IN THE EQUATION - bajo este encabezado salen las variables que todavía no han ingresado a la ecuación de regresión y para cada variable produce:

- BETA IN: el valor de BETA si la variable fuera ingresada a la ecuación en el próximo paso.
- PARTIAL: coeficiente de correlación parcial entre cada variable y la variable dependiente, teniendo como variables de control aquellas vars independientes que ya están en la ecuación.

Los otros valores, TOLERANCIA y F, son utilizados por el subprograma para la decisión de qué variable ingresará a la ecuación en el próximo paso para el modo paso a paso.

PARTE B: Tabla sumaria

La tabla sumaria trae los nombres de las variables y sus var labels en el orden que fueron incorporadas a la ecuación; el R y R^2 (MULTIPLE R y R SQUARE) producido en cada paso del análisis. Bajo el título RSQ CHANGE trae el cambio sufrido por R^2 entre un paso y el anterior. En SIMPLE R se produce el valor de los coeficientes de correlación de orden cero entre la variable dependiente y cada una de las variables independientes.

Bajo el título B y BETA salen los valores de los coeficientes de cada variable que nos da la ecuación final. BETA son los coeficientes normalizados (estandarizados).

XVI.10 Ejemplos de salida de REGRESSION

Ejemplo a . este ejemplo ilustra la salida producida por la modalidad paso a paso, opciones estandares y con la solicitud de STATISTICS ALL sin la especificacion de RESID, motivo por el cual solo los estadisticos 1 y 2 son producidos:
 1= medias, desviaciones y nº de casos; 2= coeficientes de correlacion simples. En la pagina 224 se puede observar los 4 pasos producidos - todas las variables fueron ingresadas a la ecuacion pero una en cada paso. La tabla sumaria es impresa despues del ultimo paso ("page 7") :

EJEMPLO DEL PROCEDIMIENTO - REGRESSION - 09R09R76 PAGE 2
 TASK NAME OPCIONES ESTANDARES Y ESTADISTICOS - PASO A PASO ..SIN RESID..
 REGRESSION VARIABLES=VAR068,VAR004,VAR007,VAR008,VAR010,VAR047,
 VAR025,VAR047,VAR083
 STATISTICS REGRESSION= VAR068 WITH VAR010,VAR025,VAR047,VAR083(1) R
 ALL

EJEMPLO DEL PROCEDIMIENTO - REGRESSION - 09R09R76 PAGE 3
 OPCIONES ESTANDARES Y ESTADISTICOS - PASO A PASO ..SIN RESID..
 FILE FESALEMB (CREATION DATE = 08R24R76)

VARIABLE	MEAN	STANDARD DEV	CASES
VAR068	4.3230	2.8994	1260
VAR004	1.5921	0.7720	1260
VAR007	1.5571	0.6555	1260
VAR008	1.5651	0.6797	1260
VAR010	1.6270	0.7753	1260
VAR047	9.3484	22.0807	1260
VAR025	1.4746	0.7086	1260
VAR047	9.3484	22.0807	1260
VAR083	11.7548	8.1123	1260

EJEMPLO DEL PROCEDIMIENTO - REGRESSION - 09R10R76 PAGE 4
 OPCIONES ESTANDARES Y ESTADISTICOS - PASO A PASO ..SIN RESID..
 FILE FESALEMB (CREATION DATE = 08R24R76)

CORRELATION COEFFICIENTS
 A VALUE OF 99.00000 IS PRINTED
 IF A COEFFICIENT CANNOT BE COMPUTED.

	VAR068	VAR004	VAR007	VAR008	VAR010	VAR047	VAR025	VAR047	VAR083
VAR068	1.00000	-0.21184	-0.15203	-0.16564	-0.20148	-0.00125	-0.30741	-0.00125	0.80861
VAR004	-0.21184	1.00000	0.63946	0.70001	0.89486	0.27008	0.57344	0.27008	-0.06177
VAR007	-0.15203	0.63946	1.00000	0.86167	0.67660	0.22410	0.45973	0.22410	-0.04823
VAR008	-0.16564	0.70001	0.86167	1.00000	0.73948	0.21724	0.50475	0.21724	-0.04683
VAR010	-0.20148	0.89486	0.67660	0.73948	1.00000	0.26071	0.55385	0.26071	-0.05005
VAR047	-0.00125	0.27008	0.22410	0.21724	0.26071	1.00000	0.22715	1.00000	0.21289
VAR025	-0.30741	0.57344	0.45973	0.50475	0.55385	0.22715	1.00000	0.22715	-0.18686
VAR047	-0.00125	0.27008	0.22410	0.21724	0.26071	1.00000	0.22715	1.00000	0.21289
VAR083	0.80861	-0.06177	-0.04823	-0.04683	-0.05005	0.21289	-0.18686	0.21289	1.00000

EJEMPLO DEL PROCEDIMIENTO - REGRESSION - OPCIONES ESTANDARES Y ESTADISTICOS - PASO A PASO ..SIN RESID.. FILE PESALEMB (CREATION DATE = 08R24R76)

..... MULTIPLE REGRESSION VARIABLE LIST REGRESSION LIST

DEPENDENT VARIABLE.. VAR068 TOTAL DE NACIDOS VIVOS
VARIABLE(S) ENTERED ON STEP NUMBER 1.. VAR083 EXPOSICION AL RIESGO DE EMBARAZARSE

Table with 6 columns: MULTIPLE R, R SQUARE, ADJUSTED R SQUARE, STANDARD ERROR, ANALYSIS OF VARIANCE (REGRESSION, RESIDUAL), DF, SUM OF SQUARES, MEAN SQUARE, F. Values: 0.80861, 0.65386, 0.65386, 1.76280, 1258., 6920.11737, 6920.11737, 2.91209, 2376.337

Table with 10 columns: VARIABLES IN THE EQUATION (VARIABLE, B, BETA, STD ERROR B, F) and VARIABLES NOT IN THE EQUATION (VARIABLE, BETA IN, PARTIAL, TOLERANCE, F). Includes VAR083 (CONSTANT), VAR010, VAR047, VAR025.

VARIABLE(S) ENTERED ON STEP NUMBER 2.. VAR047 MESES USO ULTIMO METODO

Table with 6 columns: MULTIPLE R, R SQUARE, ADJUSTED R SQUARE, STANDARD ERROR, ANALYSIS OF VARIANCE, DF, SUM OF SQUARES, MEAN SQUARE, F. Values: 0.82786, 0.68535, 0.68510, 1.62765, 1257., 7253.43580, 3626.71790, 2.64924, 1368.964

Table with 10 columns: VARIABLES IN THE EQUATION and VARIABLES NOT IN THE EQUATION. Includes VAR083, VAR047 (CONSTANT), VAR010, VAR025.

EJEMPLO DEL PROCEDIMIENTO - REGRESSION - OPCIONES ESTANDARES Y ESTADISTICOS - PASO A PASO ..SIN RESID.. FILE PESALEMB (CREATION DATE = 08R24R76)

..... MULTIPLE REGRESSION VARIABLE LIST REGRESSION LIST

DEPENDENT VARIABLE.. VAR068 TOTAL DE NACIDOS VIVOS
VARIABLE(S) ENTERED ON STEP NUMBER 3.. VAR010 DONDE VIVIO HACE 5 ANOS

Table with 6 columns: MULTIPLE R, R SQUARE, ADJUSTED R SQUARE, STANDARD ERROR, ANALYSIS OF VARIANCE, DF, SUM OF SQUARES, MEAN SQUARE, F. Values: 0.83601, 0.69891, 0.69844, 1.59282, 1256., 7396.98558, 3698.49279, 2.93706, 971.8580

Table with 10 columns: VARIABLES IN THE EQUATION and VARIABLES NOT IN THE EQUATION. Includes VAR083, VAR047, VAR010, VAR025 (CONSTANT).

VARIABLE(S) ENTERED ON STEP NUMBER 4.. VAR025 EDUCACION ESPOSO

Table with 6 columns: MULTIPLE R, R SQUARE, ADJUSTED R SQUARE, STANDARD ERROR, ANALYSIS OF VARIANCE, DF, SUM OF SQUARES, MEAN SQUARE, F. Values: 0.83839, 0.70290, 0.70219, 1.58288, 1255., 7439.12204, 3719.56120, 2.90551, 742.277

Table with 10 columns: VARIABLES IN THE EQUATION and VARIABLES NOT IN THE EQUATION. Includes VAR083, VAR047, VAR010, VAR025 (CONSTANT).

MAXIMUM STEP REACHED

EJEMPLO DEL PROCEDIMIENTO - REGRESSION - OPCIONES ESTANDARES Y ESTADISTICOS - PASO A PASO ..SIN RESID.. FILE PESALEMB (CREATION DATE = 08R24R76)

..... MULTIPLE REGRESSION VARIABLE LIST REGRESSION LIST

DEPENDENT VARIABLE.. VAR068 TOTAL DE NACIDOS VIVOS

SUMMARY TABLE with 6 columns: VARIABLE, MULTIPLE R, R SQUARE, RSQ CHANGE, SIMPLE R, B, BETA. Lists variables VAR083, VAR047, VAR010, VAR025, and (CONSTANT) with their respective regression statistics.

Ejemplo b. este ejemplo ilustra la salida de la modalidad mixta. El nivel de inclusion para VARO10 y VARO25 es 4, por lo tanto de mayor jerarquia que el nivel 3 de VARO47 y VARO83. De esta forma las 2 variables con nivel 4 (par) son ingresadas a la ecuacion en el primer paso. La modalidad paso a paso definida para VARO47 y VARO83 con un nivel impar, provoca el ingreso de VARO83 en el paso 2 y despues de la VARO47 en el paso 3.

TASK NAME OPCIONES ESTANDARES SIMPLES Y PASO A PASO
 REGRESSION VARIABLES=VARO68,VARO10,VARO25,VARO47,VARO83
 REGRESSION= VARO68 WITH VARO10,VARO25 (4), VARO47,VARO83 (3)

EJEMPLO DEL PROCEDIMIENTO - REGRESSION -
 OPCIONES ESTANDARES SIMPLES Y PASO A PASO
 FILE FESALEMB (CREATION DATE = 08R24R76)

09R10R76 PAGE 9

..... MULTIPLE REGRESSION VARIABLE LIST 1
 REGRESSION LIST 1

DEPENDENT VARIABLE.. VARO68 TOTAL DE NACIDOS VIVOS
 VARIABLE(S) ENTERED ON STEP NUMBER 1.. VARO10 DCNDE VIVIO HACE 5 ANOS
 VARO25 EDUCACION ESPOSO

		ANALYSIS OF VARIANCE			DF	SUM OF SQUARES	MEAN SQUARE	F
MULTIPLE R	0.30969	REGRESSION		2.	1015.04019	507.52009	66.6722	
R SQUARE	0.09591	RESIDUAL		1257.	9568.49235	7.61217		
ADJUSTED R SQUARE	0.07519							
STANDARD ERROR	2.75902							

VARIABLES IN THE EQUATION					VARIABLES NOT IN THE EQUATION				
VARIABLE	B	BETA	STD ERROR B	F	VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
VARO10	-0.16845	-0.04504	0.12046	1.956	VARO47	0.08096	0.08176	0.92215	8.48
VARO25	-1.15573	-0.28246	0.13179	76.902	VARO83	0.78419	0.80848	0.96096	2370.28
(CONSTANT)	6.30132								

VARIABLE(S) ENTERED ON STEP NUMBER 2.. VARO83 EXPOSICION AL RIESGO DE EMBARAZARSE

		ANALYSIS OF VARIANCE			DF	SUM OF SQUARES	MEAN SQUARE	F
MULTIPLE R	0.82877	REGRESSION		3.	7269.38503	2423.12834	918.3203	
R SQUARE	0.68486	RESIDUAL		1256.	3314.14751	2.63865		
ADJUSTED R SQUARE	0.68636							
STANDARD ERROR	1.62439							

VARIABLES IN THE EQUATION					VARIABLES NOT IN THE EQUATION				
VARIABLE	B	BETA	STD ERROR B	F	VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
VARO10	-0.39455	-0.10550	0.07108	30.816	VARO47	-0.13658	-0.22631	0.85972	67.24
VARO25	-0.41918	-0.10245	0.07305	28.115					
VARO83	0.28027	0.78419	0.00576	2370.280					
(CONSTANT)	2.26851								

EJEMPLO DEL PROCEDIMIENTO - REGRESSION -
 OPCIONES ESTANDARES SIMPLES Y PASO A PASO
 FILE FESALEMB (CREATION DATE = 08R24R76)

09R10R76 PAGE 10

..... MULTIPLE REGRESSION VARIABLE LIST 1
 REGRESSION LIST 1

DEPENDENT VARIABLE.. VARO68 TOTAL DE NACIDOS VIVOS
 VARIABLE(S) ENTERED ON STEP NUMBER 3.. VARO47 MESES USO ULTIMO METODO

		ANALYSIS OF VARIANCE			DF	SUM OF SQUARES	MEAN SQUARE	F
MULTIPLE R	0.83839	REGRESSION		4.	7439.12264	1859.78161	742.2773	
R SQUARE	0.70290	RESIDUAL		1255.	3144.41050	2.50551		
ADJUSTED R SQUARE	0.70219							
STANDARD ERROR	1.58288							

VARIABLES IN THE EQUATION					VARIABLES NOT IN THE EQUATION				
VARIABLE	B	BETA	STD ERROR B	F	VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
VARO10	-0.30519	-0.08160	0.07010	18.951					
VARO25	-0.31977	-0.07815	0.07797	16.818					
VARO83	0.29272	0.81900	0.00581	2538.456					
VARO47	-0.01793	-0.13658	0.00218	67.746					
(CONSTANT)	2.01793								

ALL VARIABLES ARE IN THE EQUATION

Ejemplo c. este ejemplo ilustra el uso de la modalidad simples - con la definicion de un nivel de inclusion par (2), para todas las variables. Con la especificacion de RESID=0 y STATISTICS 4,5, se produce el grafico de residuos para los primeros 500 casos. El grafico se muestra en la pagina 227.

TASK NAME	OPCIONES ESTANDARES Y ESTADISTICOS 4,5 - CON RESID=0
REGRESSION	VARIABLES=VAR068,VAR004,VAR007,VAR006,VAR010,VAR047,VAR025,VAR047,VAR083
STATISTICS	REGRESSION=VAR06 WITH VAR010,VAR025,VAR047,VAR083(2),RESID=0 4,5

EJEMPLO DEL PROCEDIMIENTO - REGRESSION - 09R10R76 PAGE 13

OPCIONES ESTANDARES Y ESTADISTICOS 4,5 - CON RESID=0
FILE FESALEMB (CREATION DATE = 08R24R76)

..... MULTIPLE REGRESSION VARIABLE LIST
REGRESSION LIST

DEPENDENT VARIABLE.. VAR068 TOTAL DE NACIDOS VIVOS

VARIABLE(S) ENTERED ON STEP NUMBER 1.. VAR083 EXPOSICION AL RIESGO DE EMBARAZARSE
VAR010 DONDE VIVIO HACE 5 ANOS
VAR047 MESES USO ULTIMO METODO
VAR025 EDUCACION ESPOSO

MULTIPLE R	0.83839	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.70290	REGRESSION	4	7439.12204	1859.78051	742.277
ADJUSTED R SQUARE	0.70219	RESIDUAL	1255	3144.41050	2.50551	
STANDARD ERROR	1.58288					

VARIABLES IN THE EQUATION					VARIABLES NOT IN THE EQUATION				
VARIABLE	B	BETA	STD ERROR B	F	VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
VAR083	0.29272	0.81900	0.00581	2538.456					
VAR010	-0.30519	-0.08160	0.07010	18.951					
VAR047	-0.31793	-0.13658	0.00218	67.746					
VAR025	-0.31977	-0.07815	0.07797	16.818					
(CONSTANT)	2.01793								

ALL VARIABLES ARE IN THE EQUATION

EJEMPLO DEL PROCEDIMIENTO - REGRESSION - 09R10R76 PAGE 14

OPCIONES ESTANDARES Y ESTADISTICOS 4,5 - CON RESID=0
FILE FESALEMB (CREATION DATE = 08R24R76)

..... MULTIPLE REGRESSION VARIABLE LIST
REGRESSION LIST

DEPENDENT VARIABLE.. VAR068 TOTAL DE NACIDOS VIVOS

SUMMARY TABLE									
VARIABLE		MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA		
VAR083	EXPOSICION AL RIESGO DE EMBARAZARSE	0.80861	0.65386	0.65386	0.80861	0.29272	0.81900		0.81900
VAR010	DONDE VIVIO HACE 5 ANOS	0.82453	0.67985	0.02599	-0.20148	-0.30519	-0.08160		-0.08160
VAR047	MESES USO ULTIMO METODO	0.83601	0.69891	0.01907	-0.00125	-0.31793	-0.13658		-0.13658
VAR025	EDUCACION ESPOSO	0.83839	0.70290	0.00398	-0.30741	-0.31977	-0.07815		-0.07815
(CONSTANT)						2.01793			

Grafico de residuos producido con estadisticos 4,5 y la especificacion de RESID=0.

EJEMPLO DEL PROCEDIMIENTO - REGRESSION - OPCIONES ESTANDARES Y ESTADISTICOS 4,5 - CON RESID=0 FILE FESALEMB (CREATION DATE = 08R24N76)

09N10R76

PAGE 16

..... MULTIPLE REGRESSION

DEPENDENT VARIABLE VAR068 FROM VARIABLE LIST 1 REGRESSION LIST 1

SEGNUM	OBSERVED VAR068	PREDICTED VAR068	RESIDUAL	PLOT OF STANDARDIZED RESIDUAL				
				-2.0	-1.0	0.0	1.0	2.0
1	4.000000	6.044463	-2.044463					
2	2.000000	1.129124	-.870878					
3	3.000000	2.817032	-.1829479					
4	4.000000	10.45470	-6.454704	X				
5	5.000000	4.370136	-.6798631					
6	10.00000	7.786760	2.233238					
7	5.000000	7.842728	-2.842730					
8	1.000000	1.837045	-.8370456					
9	4.000000	4.940175	-.9401751					
10	7.000000	7.807785	-.8077872					
11	2.000000	2.271124	-.2711238					
12	2.000000	2.856556	-.8565561					
13	13.00000	7.832728	5.167269					
14	7.000000	4.083962	-.083962					
15	5.000000	4.452341	-.452341					
16	7.000000	9.003593	-2.003594					
17	11.00000	9.296309	1.703688					
18	11.00000	MISSING**	MISSING**					
19	.0	MISSING**	MISSING**					
20	.0	MISSING**	MISSING**					
21	9.000000	6.976716	2.023283					
22	.0	MISSING**	MISSING**					
23	1.000000	1.033685	-.3368288E-01					
24	4.000000	2.416710	1.583289					
25	2.000000	1.033685	-.9663171					
26	3.000000	1.619115	1.380884					
27	1.000000	1.757827	-.2021733					
28	5.000000	4.018925	-.018925					
29	1.000000	2.013339	-1.013339					
30	3.000000	3.665079	-.6650796					
31	1.000000	1.126074	-.1260733					
32	4.000000	MISSING**	MISSING**					

EJEMPLO DEL PROCEDIMIENTO - REGRESSION - OPCIONES ESTANDARES Y ESTADISTICOS 4,5 - CON RESID=0

09N10R76

PAGE 22

472	7.000000	6.076433	-.9235662					
473	8.000000	8.405690	-.4056912					
474	3.000000	3.955381	-.9553818					
475	1.000000	1.673220	-.6732137					
476	4.000000	8.671353	-4.671354					
477	4.000000	3.115516	-.8844841					
478	2.000000	6.954581	-4.954582					
479	5.000000	7.247296	-2.247298					
480	1.000000	2.055912	-1.055912					
481	3.000000	3.734704	-.7347046					
482	5.000000	5.198285	-.1982853					
483	.0	MISSING**	MISSING**					
484	4.000000	5.158761	-1.158761					
485	8.000000	5.230850	2.769149					
486	3.000000	2.271124	-.7288762					
487	4.000000	4.070368	-.3684293E-03					
488	7.000000	7.820258	-.8202589					
489	3.000000	3.149272	-.1492722					
490	5.000000	4.405569	-.9443080E-01					
491	4.000000	3.710371	-.2896287					
492	6.000000	5.171232	-.8287669					
493	4.000000	4.427420	-.2742071E-01					
494	3.000000	4.320136	-1.320136					
495	10.00000	11.63804	-1.638041					
496	2.000000	2.271124	-.2711238					
497	5.000000	4.452743	-.452743					
498	5.000000	4.470907	-.2909247E-01					
499	4.000000	4.302202	-.3022025					
500	1.000000	4.320136	-3.320136					

Ejemplo d. este ejemplo ilustra el uso de las opciones 10,11,12, que provocan la grabacion de archivo con predictores y residuos, para las listas de regresion con la especificacion RESID. En esta pagina, "page 28", el SPSS publica un reporte con la informacion relativa al formato de grabacion del archivo. En la pagina 229 se muestra el listado del archivo grabado - esta listado no es producido por el SPSS. El listado del archivo grabado, realizado con un programa de utileria, se muestra para efectos de visualizacion.

```
TASK NAME      USO DE OPCION 10,11,12- SALIDA DE ARCHIVO CON PRED.Y RESIDUOS
REGRESSION     VARIABLES=VAR068,VAR004,VAR007,VAR008,VAR010,VAR047,
              VAR025,VAR047,VAR083
OPTIONS        REGRESSION= VAR068 WITH VAR010,VAR025,VAR047,VAR08312a,RESID=0
              10,11,12
```

EJEMPLO DEL PROCEDIMIENTO - REGRESSION -
 USO DE OPCION 10,11,12- SALIDA DE ARCHIVO CON PRED.Y RESIDUOS
 FILE PESALEMB (CREATION DATE = 08R24R76)

09A10R76

PAGE 28

1 PAIRS OF RESIDUAL AND PREDICTOR WERE WRITTEN ON LOGICAL UNIT 9 FOR 1471 UNWEIGHTED CASES. 1 RECORDS OUTPUT PER CASE.
 OUTPUT FORMAT IS (F8.0,F2.0,1X,A4,5X,6F10.6). RECORD NUMBER APPEARS LEFT-ZERO-FILLED.
 MISSING INFORMATION IS OUTPUT AS 999.0. NON-MISSING BUT EXTREME INFORMATION IS TRUNCATED TO +99.0 OR -99.0

RESIDUAL AND PREDICTOR FROM VARIABLE LIST	REGRESSION LIST	OUTPUT INFORMATION	OUTPUT RECORD NUMBER PER CASE	RECORD COLUMNS	UNWEIGHTED NUMBER OF MISSING CASES
		SEQNUM		1-8	
		RECORD NUMBER		9-10	
		SUBFILL		12-15	
		RESIDUAL		21-30	211
		PREDICTOR		31-40	211

Listado del archivo generado con el uso de opciones 10,11,12. Este listado se produjo con el uso de un programa utilitario y no con SPSS.

00001.	101	FESA	-0.705142	0.593733	
00002.	201	FESA	0.300369	-1.101585	
00003.	301	FESA	0.063106	-0.519419	
00004.	401	FESA	-2.226250	2.114641	
00005.	501	FESA	0.234487	-0.000993	
00006.	601	FESA	0.770252	1.107760	
00007.	701	FESA	-0.977019	1.210513	
00008.	801	FESA	-0.288700	-0.537420	
00009.	901	FESA	-0.254270	0.212860	
00010.	1001	FESA	-0.278239	1.701910	
00011.	1101	FESA	-0.094212	-0.797705	
00012.	1201	FESA	-0.294229	-0.505787	
00013.	1301	FESA	1.742709	1.210513	
00014.	1401	FESA	0.322843	0.600458	
00015.	1501	FESA	0.016437	0.217057	
00016.	1601	FESA	-0.621647	1.614348	
00017.	1701	FESA	0.257800	1.715307	
00018.	1801	FESA	999.000000999	0.000000	
00019.	1901	FESA	999.000000999	0.000000	
00020.	2001	FESA	999.000000999	0.000000	
00021.	2101	FESA	0.697637	0.915270	
00022.	2201	FESA	999.000000999	0.000000	
00023.	2301	FESA	-0.011617	-1.134502	
00024.	2401	FESA	0.346582	-0.637402	
00025.	2501	FESA	0.333286	-1.144502	
00026.	2601	FESA	0.476272	-0.932585	
00027.	2701	FESA	0.069730	-0.870946	
00028.	2801	FESA	-0.351448	0.584942	
<hr/>					
01426.	142601	FESA	0.866391	1.091620	
01427.	142701	FESA	0.412211	-0.523721	
01428.	142801	FESA	999.000000999	0.000000	
01429.	142901	FESA	0.646525	-1.413431	
01430.	143001	FESA	0.857821	1.110951	
01431.	143101	FESA	0.153068	-0.609381	
01432.	143201	FESA	-0.114671	-1.031449	
01433.	143301	FESA	-0.126340	-0.927556	
01434.	143401	FESA	0.227298	-1.026515	
01435.	143501	FESA	999.000000999	0.000000	
01436.	143601	FESA	-0.286926	-0.514390	
01437.	143701	FESA	-0.421594	0.999992	
01438.	143801	FESA	-0.016646	-1.129473	
01439.	143901	FESA	999.000000999	0.000000	
01440.	144001	FESA	0.221504	-0.332913	
01441.	144101	FESA	0.006719	-0.837935	
01442.	144201	FESA	-0.110416	-0.000993	
01443.	144301	FESA	999.000000999	0.000000	
01444.	144401	FESA	999.000000999	0.000000	
01445.	144501	FESA	0.259609	1.008595	
01446.	144601	FESA	-0.089210	-0.712006	
01447.	144701	FESA	-0.506414	1.084812	
01448.	144801	FESA	-0.337456	-0.808663	
01449.	144901	FESA	0.147276	-0.258685	
01450.	145001	FESA	0.318541	0.654760	
01451.	145101	FESA	0.259693	-0.712006	
01452.	145201	FESA	-0.089210	-0.712006	
01453.	145301	FESA	999.000000999	0.000000	
01454.	145401	FESA	-0.112708	-0.913924	
01455.	145501	FESA	-0.126292	1.394497	
01456.	145601	FESA	-0.223492	-0.334901	
01457.	145701	FESA	-1.466425	1.210513	
01458.	145801	FESA	1.411729	1.236089	
01459.	145901	FESA	0.288885	0.289512	
01460.	146001	FESA	-0.051485	-0.404828	
01461.	146101	FESA	0.133528	-0.079966	
01462.	146201	FESA	-0.337456	-0.808663	
01463.	146301	FESA	999.000000999	0.000000	
01464.	146401	FESA	-0.232196	-0.913924	
01465.	146501	FESA	-1.784689	2.018183	
01466.	146601	FESA	-0.110416	-0.000993	
01467.	146701	FESA	0.631524	-0.742933	
01468.	146801	FESA	0.601636	1.311472	
01469.	146901	FESA	0.041520	-0.842736	
01470.	147001	FESA	-0.143840	-0.312473	
01471.	147101	FESA	-0.606157	-0.195059	

Ejemplo e. este ejemplo ilustra la salida producida con el uso de STATISTICS 6, y RESID=0. El grafico producido se encuentra en pagina 231. La modalidad definida como paso a paso, genera 9 pasos, de los cuales se muestra solamente el primero y el ultimo.

EJEMPLO DEL PROCEDIMIENTO - REGRESSION -

09R10R76

PAGE 2

TASK NAME OPCIONES ESTANDARES Y ESTADISTICO 6
REGRESSION VARIABLES= VAR001 TO VAR012
STATISTICS REGRESSION= VAR001 WITH VAR003 TO VAR011(1),RESID=0

EJEMPLO DEL PROCEDIMIENTO - REGRESSION -
OPCIONES ESTANDARES Y ESTADISTICO 6
FILE NAME (CREATION DATE = 09R10R76)
SUBFILE SUBF1 SUBF2

09R10R76

PAGE 3

..... MULTIPLE REGRESSION VARIABLE LIST
REGRESSION LIST

DEPENDENT VARIABLE.. VAR001 TASA DE MORTALIDAD INFANTIL
VARIABLE(S) ENTERED ON STEP NUMBER 1.. VAR011 PCT MAC.VIVOS ATENDIDOS POR COMADRLNA

MULTIPLE R		ANALYSIS OF VARIANCE		DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.71081	REGRESSION	1.	1340624.47543	1340624.47543		69.443
ADJUSTED R SQUARE	0.50525	RESIDUAL	68.	1312764.11028	19305.35456		
STANDARD ERROR	138.94371						

----- VARIABLES IN THE EQUATION -----

VARIABLE	B	BETA	STD ERROR B	F
VAR011	6.03151	0.71081	0.72379	69.443
(CONSTANT)	424.85192			

----- VARIABLES NOT IN THE EQUATION -----

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
VAR003	-0.15756	0.16045	0.51304	1.7
VAR004	-0.11025	-0.05553	0.40323	0.6
VAR005	-0.18630	0.20133	0.57755	0.6
VAR006	0.03080	0.02904	0.43591	0.6
VAR007	0.32541	0.23913	0.46162	0.6
VAR008	0.17755	0.19681	0.73988	0.6
VAR009	0.16084	-0.19681	0.73988	0.6
VAR010	-0.38064	-0.43784	0.65461	15.8

EJEMPLO DEL PROCEDIMIENTO - REGRESSION -
OPCIONES ESTANDARES Y ESTADISTICO 6
FILE NAME (CREATION DATE = 09R10R76)
SUBFILE SUBF1 SUBF2

09R10R76

PAGE 9

..... MULTIPLE REGRESSION VARIABLE LIST
REGRESSION LIST

DEPENDENT VARIABLE.. VAR001 TASA DE MORTALIDAD INFANTIL
VARIABLE(S) ENTERED ON STEP NUMBER 9.. VAR006 PCT DE VIVIENDAS BUENAS

MULTIPLE R		ANALYSIS OF VARIANCE		DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.82103	REGRESSION	9.	178832.24169	198736.91574		13.768
ADJUSTED R SQUARE	0.63135	RESIDUAL	60.	864756.34402	14412.60573		
STANDARD ERROR	120.05251						

----- VARIABLES IN THE EQUATION -----

VARIABLE	B	BETA	STD ERROR B	F
VAR011	6.08884	0.71757	1.48732	16.759
VAR010	-20.71652	-0.61749	1.80462	8.371
VAR008	-5.64336	-0.68219	1.83494	8.459
VAR009	5.99101	0.19428	3.62672	2.462
VAR007	1.57801	0.14072	1.55234	0.678
VAR004	4.13651	0.12923	5.06892	0.666
VAR007	1.62978	0.14674	1.70679	0.912
VAR005	1.02685	0.10586	1.80735	0.323
VAR006	-0.86419	-0.05213	2.30834	0.140
(CONSTANT)	975.86414			

----- VARIABLES NOT IN THE EQUATION -----

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
VAR003	-0.15756	0.16045	0.51304	1.7
VAR004	-0.11025	-0.05553	0.40323	0.6
VAR005	-0.18630	0.20133	0.57755	0.6
VAR006	0.03080	0.02904	0.43591	0.6
VAR007	0.32541	0.23913	0.46162	0.6
VAR008	0.17755	0.19681	0.73988	0.6
VAR009	0.16084	-0.19681	0.73988	0.6
VAR010	-0.38064	-0.43784	0.65461	15.8

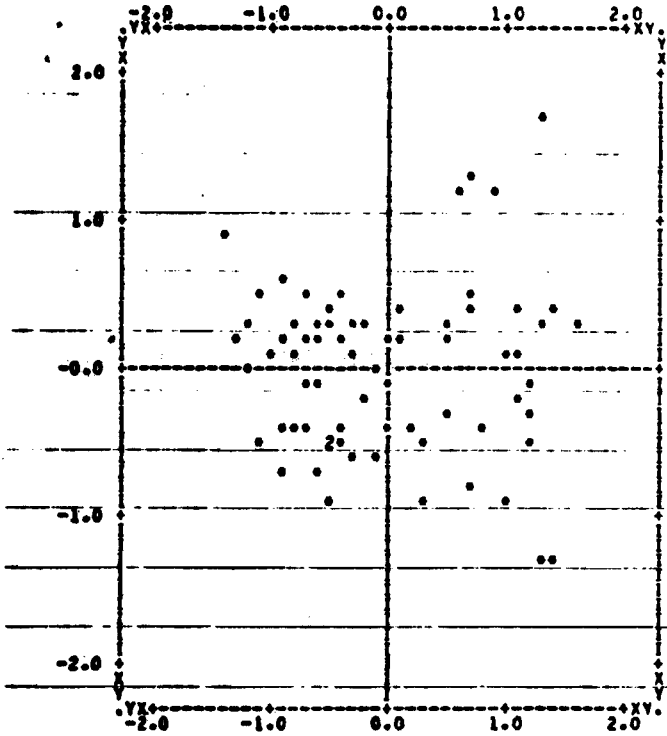
Grafico de residuos producido con el uso del estadistico 6 y RESID=0.

EJEMPLO DEL PROCEDIMIENTO - REGRESSION -
OPCIONES ESTADISTICAS Y ESTADISTICO 6
FILE NONAME (CREATION DATE = 09R10K76)
SUBFILE SUBF1 SUBF2

09R10K76

PAGE 12

..... PLOT0 STANDARDIZED RESIDUAL (DOWN) -- PREDICTED STANDARDIZED DEPENDENT VARIABLE (ACROSS)
DEPENDENT VARIABLE VAR001 VARIABLE LIST 1
 REGRESSION LIST 1



ROWS,COLUMNS YO VALUES OUTSIDE (-3.0,3.0) ROWS,COLUMNS XO VALUES IN (-3.0,-2.05) OR (2.05,3.0)

APENDICE 1: TARJETAS JCL PARA O.S. IBM 360/370

La versión SPSS para computadores IBM opera normalmente bajo el sistema operativo O.S. (Operating System).

El propósito de este apéndice es hacer un resumen sobre las tarjetas de JCL (Job Control Language) que requiere el uso de SPSS.

A. La declaración EXEC

```
//PASO EXEC PGM=SPSS
```

Si en la tarjeta EXEC no se especifica el valor de SPACE, el sistema asume SPACE = 80000 bytes.

Para modificar el valor de SPACE:

```
// EXEC PGM=SPSS,PARM=23K
```

con lo cual SPACE = 23K

El valor de SPACE depende de cada equipo considerando que el SPSS ocupa para las múltiples funciones un total de 140K

B. La declaración STEPLIB

Esta declaración debe direccionar la biblioteca particionada donde reside el miembro SPSS llamado con el EXEC PGM = SPSS

```
//STEPLIB DD dirección de la biblioteca.
```

C. FT01F001

Esta unidad lógica es utilizada para almacenar información relacionada a títulos - VAR LABELS y value labels. Requiere un bloque de 800 caracteres por variable. Así:

```
//FT01F001 DD UNIT=SYSDA,SPACE=(800,(600,100))
```

tendría un espacio inicial para 600 variables y se incrementa en 100 bloques de 800 caracteres si se requiere más espacio.

D. FT02F001

Esta unidad lógica direcciona el archivo (temporal) que utiliza el SPSS para el archivo borrador generalmente se usa disco pero se permite la utilización de cintas.

El cálculo del espacio se realiza en términos de bloque de 2012 caracteres.

Para 100 variables y 2000 casos se requiere

$$NB = \frac{4 * NVAR * NCASE}{2012} = \frac{4 * 100 * 2000}{2012}$$

NB = 398 bloques de 2012 caracteres

Ejemplos:

```
//FT02F001 DD UNIT=SYSDA,SPACE=(2012,(1000,100))
```

```
//FT02F001 DD UNIT=SYSDA,SPACE=(CYL,(30,5)),DCB=BLKSIZE=2012
```

Dispositivo	Bytes por pista	Pistas por cilindro	Cilindros por volumen	Capacidad total en bytes
2311	3625	10	200	7,250,000
2314	7294	20	200	29,176,000
3330 (Mode- lo 1)	13030	19	404	100,018,280

Tabla de capacidad de almacenamiento en discos.

E. FT06F001

Esta unidad define la salida impresa:

```
//FT06F001 DD SYSOUT=A
```

F. FT05F001

Define el medio donde reside el programa:


```
//FT05F001 DD *  
RUN NAME  
-----  
FINISH  
/*
```

G. FT08F001: Archivo BCD

Si en INPUT MEDIUM se especificó TAPE o DISK se debe especificar la dirección del archivo por medio de FT08:

```
//FT08F001 DD UNIT=2400,VOL=SER=CCO47,DISP=OLD,  
// LABEL=1,DSN=FESAL.BCD
```

H. FT04F001: Archivo SPSS - SAVE FILE

El uso de SAVE FILE exige la especificación de FT04. Si no se especifica DGB con el parámetro BLSIZE, el sistema utiliza BLKSIZE = 2012.

```
//FT04F001 DD UNIT=SYSDA,VOL=SER=CELA01,DISP=(NEW,KEEP),  
// SPACE=(2012,(1000,100),RLSE),  
// DSN=SAVE.FESALEMB
```

Es recomendable que el último parámetro del DSN lleve el mismo nombre utilizado en FILE NAME que para el ejemplo sería:

```
"FILE NAME FESALEMB"
```

Se permite para FT04 tanto disco como cintas. Si se usa disco el cálculo del espacio requerido según el número de casos es igual a FT02.

I. FT03F001 - Entrada de Archivo SPSS - GET FILE

La especificación de FT03 es requerida si se usa GET FILE.

```
//FT03F001 DD UNIT=SYSDA,VOL=SER=CELA01,  
// DSN=SAVE.FESALEMB,DISP=OLD
```

J. FT09F001 - salida con opciones (FT15, ... FT20)

Si el usuario no especificó RAW OUTPUT UNIT el sistema asume que la grabación o perforación de datos solicitada por las opciones de los procedimientos se realizará a la unidad lógica FT09F001.

A excepción de WRITE CASES en que el LRECL depende del formato, todos las salidas solicitadas con OPTIONS tendrán LRECL = 80.

```
//FT09F001 DD UNIT=OOD,DCB=BLKSIZE=80          (perforadora)
//FT09F001 DD UNIT=2400,VOL=SER=GCO91,
//          LABEL=1,DSN=OUT.MATRIZ.FESAL,
//          DCB=(RECFM=FB,LRECL=80,BLKSIZE=4000)
```

Para WRITE CASES que cada caso requiere un total de 100 caracteres tendríamos:

```
//FT09F001 DD UNIT=2400,VOL=SER=GCO91,LABEL=2,
//          DSN=WRTCASE.FESAL,DCB=(RECFM=FB,LRECL=100,
//          BLKSIZE=5000)
```

K. El procedimiento catalogado SPSSH6

En el computador de la UCR se encuentra catalogado el procedimiento SPSSH6 que ahorra esfuerzo al usuario, evitando que tenga que especificar FT01, FT02, FT06, FT05; requiriendo según sea el caso que especifique FT08, FT04, FT03, FT09.

Ejemplos:

```
// EXEC SPSSH6
//FT08F001 DD archivo del usuario-bcd
//FT04F001 DD archivo del sistema para SAVE FILE
//SYSIN DD *
RUN NAME
.....
FINISH
/*
//
```

Para cambiar el valor del parámetro SPACE, la especificación sería la siguiente:

```
// EXEC SPSSH6,SPACE=20K
```

En el caso de que no se especifique el parámetro SPACE, el procedimiento catalogado asume SPACE=23K que es el máximo permitido en el equipo IBM/360 del Centro de Informática de la Universidad de Costa Rica.

Para modificar los valores de los espacios para FT01 y/o FT02, el procedimiento catalogado permite la especificación de los parámetros SP1, SP2, como se ilustra a continuación:

```
// EXEC SPSSH6,SP1=200
```

```
// EXEC SPSSH6,SP2=40
```

```
// EXEC SPSSH6,SP1=200,SP2=40
```

Si los parámetros no son especificados, son asumidos los siguientes valores:

SP1= 300 tracks para la unidad lógica FT01FOO1

SP2= 30 cilindros para la unidad lógica FT02FOO1

BIBLIOGRAFIA

1. SPSS - Statistical Package for the Social Sciences. Second Edition.
 - Norman Nie, C. Hadlai Hull, Jean Jenkins, Karin Steinbrenner, Dale Bent.
 - Mac Graw Hill, 1975.

2. Manual de SPSS
 - Juan Jorge Méndez
 - ECOM, Empresa Nacional de Computación e Informática, Santiago de Chile, 1973.

3. SPSS, First Edition
 - Norman Nie, Dale Bent, C. Hadlai Hull
 - Mac Graw Hill, 1970

4. Job Control Language. OS/360; OS/VS1 370.
 - IBM

5. Data Management Service Guide OS/VS1 370.
 - IBM