

201  
1A



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**FACULTAD DE CIENCIAS**

**USO Y APLICACIONES  
DEL PAQUETE SPSS.**

**T E S I S**

QUE PARA OBTENER EL TÍTULO DE :

**A C T U A R I O**

P R E S E N T A :

**BEATRIZ DEL SOCORRO ALCALA VELAZQUEZ**

México, D. F.

1987



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**CONTENIDO.**

Introducción. 1

I. Comunicación con la computadora. 5

1.1 Archivos utilizados por SPSS. 7

1.2 La ejecución del paquete. 8

1.3 Ejemplo de resultados de SPSS. 9

II. Definición de caso y variable. 13

III. Formato de instrucciones. 15

IV. Procedimientos estadísticos de SPSS. 17

V. Ejemplos de las instrucciones de SPSS. 20



acento

APH) → V<sup>3</sup>a

APH~ → V<sup>2</sup>a

% \_\_\_\_\_

**UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO**

6.12 OPTIONS 21

6.13 RAW OUTPUT UNIT 21

6.14 READ INPUT DATA 23

6.15 FINISH 23

24

25

27

29

30

31

31

32

33

33

34

VII. Instrucciones que controlan la impresión de resultados.

7.1 PRINT FORMAT	35
7.2 PAGESIZE	36
7.3 PRINT BACK	36
7.4 LIST CASES	37

VIII. Instrucciones de etiquetado para variables y valores.

8.1 VAR LABELS	38
8.2 VALUE LABELS	39

IX. Instrucción para control de valores missing.

9.1 MISSING VALUES	40
--------------------	----

X. Generación de archivos en disco por SPSS.

10.1 FILE NAME	42
10.2 SAVE FILE	43
10.3 GET FILE	44
10.4 DOCUMENT	44
10.5 LIST FILEINFO	45
10.5.1 Ejemplo de generación de archivos.	46
10.6 WRITE CASES	48
10.7 WRITE FILEINFO	49

XI. Creación de subarchivos.

11.1 SUBFILE LIST	50
11.2 RUN SUBFILES	51



# CONTENIDO.

Introducción.	1
I. Comunicación con la computadora.	5
1.1 Archivos utilizados por SPSS.	7
1.2 La ejecución del paquete.	8
1.3 Ejemplo de resultados de SPSS.	9
II. Definición de caso y variable.	13
III. Formato de instrucciones.	15
IV. Procedimientos estadísticos de SPSS.	17
V. Tabla de precedencia de las instrucciones de SPSS.	20
VI. Instrucciones básicas.	
6.1 COMMENT	21
6.2 EDIT	21
6.3 NUMBERED	23
6.4 RUN NAME	23
6.5 VARIABLE LIST	24
6.6 INPUT FORMAT	25
6.7 DATA LIST	27
6.8 INPUT MEDIUM	29
6.9 N OF CASES	30
6.10 TASK NAME	31
6.11 STATISTICS	31
6.12 OPTIONS	32
6.13 RAW OUTPUT UNIT	33
6.14 READ INPUT DATA	33
6.15 FINISH	34

VII. Instrucciones que controlan la impresión de resultados.

7.1 PRINT FORMAT	35
7.2 PAGESIZE	36
7.3 PRINT BACK	36
7.4 LIST CASES	37

VIII. Instrucciones de etiquetado para variables y valores.

8.1 VAR LABELS	38
8.2 VALUE LABELS	39

IX. Instrucción para control de valores missing.

9.1 MISSING VALUES	40
--------------------	----

X. Generación de archivos en disco por SPSS.

10.1 FILE NAME	42
10.2 SAVE FILE	43
10.3 GET FILE	44
10.4 DOCUMENT	44
10.5 LIST FILEINFO	45
10.5.1 Ejemplo de generación de archivos.	46
10.6 WRITE CASES	48
10.7 WRITE FILEINFO	49

XI. Creación de subarchivos.

11.1 SUBFILE LIST	50
11.2 RUN SUBFILES	51

XII. Transformación y creación de variables.		52
12.1 RECODE	y *RECODE	53
12.2 COMPUTE	y *COMPUTE	55
12.3 IF	e *IF	57
12.4 COUNT	y *COUNT	60
12.5 SELECT IF	y *SELECT IF	62
12.6 WEIGHT	y *WEIGHT	63
12.7 SAMPLE	y *SAMPLE	66
12.8 DO REPEAT	y END REPEAT	66

XIII. Descripción de subrutinas.		
13.1 CONDESCRIPTIVE		69
13.2 FREQUENCIES		72
13.3 AGGREGATE		77
13.3.1 SORT CASES		82
13.4 CROSSTABS		83
13.5 BREAKDOWN		90
13.6 T-TEST		98
13.7 PEARSON CORR		102
13.8 NONPAR CORR		106
13.9 SCATTERGRAM		109
13.10 PARTIAL CORR		113
13.11 REGRESSION		121
13.12 ANOVA		132
13.13 ONEWAY		138
13.14 DISCRIMINANT		144
13.15 FACTOR		159
13.16 CANCORR		173
13.17 GUTTMAN SCALE		180

XIV. Versiones 7 a 9 de SPSS.	186
14.1 REPORT	187
14.2 SURVIVAL	187
14.3 NPAR TESTS	188
14.4 MULT RESPONSE	189
14.5 RELIABILITY	190
14.6 MANOVA	190
14.7 BOX JENKINS	191
14.8 NEW REGRESSION	191
14.9 Gráficas: PIECHART, BARCHART y LINECHART	192
14.10 Selección y modificación de datos.	
13.10.1 SAMPLE	193
13.10.2 SEED	194
13.10.3 Funciones: NORMAL, UNIFORM e YRMODE	195
14.11 Lectura de datos.	
13.11.1 N OF CASES	196
13.11.2 END INPUT DATA	196
13.11.3 READ INPUT DATA y READ MATRIX	197
14.12 SORT CASES	197
XV. Versión X de SPSS.	200
15.1 NUEVAS RUTINAS EN PROX. VERSIONES DE SPSS	203
XVI. Versión micro de SPSS.	205
Conclusión	210
APENDICE I. Palabras reservadas de SPSS.	215
APENDICE II. Otros paquetes estadísticos:	
SAS, BMDP y OSIRIS.	216
BIBLIOGRAFIA	222
INDICE ALFABETICO.	228

## INTRODUCCION

El SPSS es un sistema de análisis de información y una magnífica herramienta para manejar, analizar y desplegar información. Puede tomar los datos desde cualquier archivo secuencial (o combinación de archivos) y convertirlos en valiosa información: reportes tabulados, gráficas de distribuciones y tendencias, y resultados de una amplia variedad de procedimientos estadísticos. SPSS combina el manejo de datos, la impresión de reportes y el análisis estadístico en un sólo sistema con un sólo lenguaje.

Las siglas de SPSS significan "Statistical Package for the Social Sciences" (Paquete estadístico para las ciencias sociales) y su mayor logro, desde sus inicios en 1966, es el proporcionar a sus múltiples usuarios un sistema que contiene todos los requerimientos computacionales, permitiéndoles así, concentrarse en los resultados más que en la mecánica para obtenerlos.

El SPSS se puede ejecutar en una gran variedad de computadoras y sistemas operativos, cada uno de los cuales requiere información adicional para esa versión de SPSS.

Además de las estadísticas descriptivas usuales, distribución simple de frecuencias y tabulaciones cruzadas, SPSS contiene procedimientos para: correlaciones simples (ordinal e intervalar), correlaciones parciales, medias y varianzas para submuestras estratificadas, análisis de varianza de una vía (involucra sólo una variable independiente) ó de n-vías (involucra 'n' variables independientes) que incluyen tablas de análisis de clasificación múltiple, regresión múltiple, análisis discriminante, análisis factorial, correlaciones canónicas y escalas de Guttman.

Para las versiones 7 y 8 se cuenta, además, con los siguientes procedimientos: generador de reportes, tablas de vida, 14 pruebas no paramétricas, distribución de frecuencias y tabulaciones cruzadas para variables de respuesta múltiple.

En la versión 9 se agregaron los siguientes procedimientos: análisis de varianza multivariado, análisis de series de tiempo que utiliza la técnica Box-Jenkins, análisis de regresión múltiple con extensión para análisis de residuales y procedimientos para generar gráficas de barra, de línea y de pastel o "pie".

En la versión 10, más conocida como SPSSx, la mayoría de los cambios fueron realizados en el área de definición de archivos y transformación de datos, es decir, comandos no de procedimiento. Sin embargo, algunos procedimientos tienen cambios mínimos, otros cambios importantes y algunos fueron totalmente reescritos. Los procedimientos GUTTMAN SCALE y CANCORR ya no fueron incluidos (el nuevo procedimiento MANOVA es una alternativa para CANCORR).

En 1984 salió al mercado el programa SPSS/PC, versión de SPSS para micro. El SPSS/PC se ejecuta en una IBM PC/XT o sistema compatible, con un disco duro de 10 megabytes, con versión 2.0 ó 2.1 de sistema operativo MS DOS o PC DOS y 320 k de memoria. El SPSS/PC se distribuye en 10 diskettes (5 1/4 pulg. DSDD), consistentes en un Tutorial (diskette de demostración), un disco llave de seguridad y 8 diskettes de programas. El SPSS/PC cuenta con una gran cantidad de pantallas de ayuda, que pueden accederse desde cualquier punto del programa. Aunque limitado en memoria, procedimientos y variables permitidas (200 activas durante una sesión), el SPSS/PC cuenta con las mismas características y ventajas que SPSSx.

Actualmente al SPSS lo usan en todo el mundo estudiantes, estadísticos, investigadores de mercado, biólogos, auditores, investigadores sociales, ejecutivos de negocios, investigadores médicos, analistas de cómputo y muchos otros que confían en sus algoritmos estadísticos.

El paquete SPSS fue desarrollado en la Universidad de Chicago con la colaboración de otras universidades de Estados Unidos y es revisado, actualizado y comercializado por la compañía SPSS Inc.

El SPSS está en constante revisión de sus manuales y procedimientos que ofrece, por lo que periódicamente publica y distribuye sus nuevas versiones que incluyen nuevos programas, con mejores y más robustos algoritmos estadísticos.

Todos los programas están revisados y verificados en respuesta a las sugerencias de los usuarios para mejorar su eficiencia, corregir errores o solicitar nuevas funciones.

Debido a que el SPSS fue pensado como una herramienta de análisis de datos para todo propósito, muchas características del SPSS sólo son necesarias para problemas complejos. Así, no se necesita saber todo este manual para poder usar SPSS. Este libro contiene toda la información que se necesita para usar SPSS, está formulado para usarse como un manual de referencia. Cuando ud. ve una instrucción en particular encontrará toda la información sobre ella. Esta cantidad de información puede ser excesiva, siéntase libre de saltar partes que no se apliquen a ud. Utilice el índice, cuyo propósito es ahorrarle tiempo.

Otra manera de aprender a usar el paquete SPSS es asistiendo a cursos. SPSS Inc. ofrece periódicamente cursos en varias ciu-

dades de Estados Unidos. Muchas organizaciones en México también dan cursos rápidos de SPSS.

La intención de este trabajo es proporcionar a los usuarios de SPSS, principiantes o con experiencia, un libro de consulta rápida, accesible, completa y en español. Consta de 16 capítulos. Los primeros 13 explican ampliamente las características de la versión 6 de SPSS, la cual sirve como base para trabajar con las mejoras que ofrecen las versiones 7, 8 y 9 presentadas en el capítulo 14. Las principales características de las versiones SPSSx y SPSS/PC se presentan en los capítulos 15 y 16 respectivamente.

El enfoque que se le ha dado al manual es fundamentalmente de Uso del Paquete, y aunque se presentan los objetivos y características de las técnicas tratadas por cada subrutina del SPSS, y se buscó que los ejemplos cumplieran los supuestos y presentaran resultados coherentes, no se profundiza en ellas, ya que cada técnica podría ser un tema de libro por sí mismo.

Por lo tanto, se requiere que los lectores de este manual y usuarios del SPSS, cuenten con conocimientos estadísticos básicos, al menos, para poder apreciar la capacidad y resultados producidos por cada subrutina, ya que, una vez que se aprende a usar el paquete SPSS, lo más importante será la interpretación de sus resultados.

Para profundizar en alguna o todas las técnicas estadísticas como análisis de regresión, análisis de varianza, análisis discriminante, análisis factorial, etc., se recomienda consultar la amplia bibliografía existente.



## I. COMUNICACION CON LA COMPUTADORA

Las personas que deseen utilizar el paquete SPSS deberán, en primer lugar, obtener la autorización y una clave de usuario ante el Centro de Cómputo donde se va a trabajar.

El sistema necesita instrucciones que identifiquen al trabajo y que indiquen los dispositivos que se van a utilizar: discos, cintas magnéticas, impresoras, así como paquetes, programas y archivos de datos.

Esta información se proporciona a la computadora a través de tarjetas perforadas o de una terminal remota. Este capítulo tratará el procedimiento general, ya que las instrucciones y el lenguaje dependerá de la computadora utilizada, de su editor y su sistema operativo. Las instrucciones y comandos de SPSS, permanecen inalterables de computadora a computadora, salvo quizá, cambios mínimos debidos a la instalación.

Actualmente el uso de tarjetas perforadas no es común ya que se prefiere la terminal remota, pero como algunos centros de cómputo todavía las utilizan como alternativa, se explicará la secuencia correcta de las tarjetas en el uso del paquete SPSS.

- A) Tarjetas de control que identifican al trabajo, al usuario y que indican la utilización de SPSS.
- B) Tarjetas de instrucciones SPSS que describen los datos.
- C) Instrucción READ INPUT DATA de SPSS que precede a los datos y acciona su lectura.
- D) Datos.
- E) Tarjetas que describen procedimientos estadísticos, los cuales

pueden ejecutarse en cualquier secuencia y tantas veces como se desee en una sola corrida.

- F) FINISH, instrucción de SPSS que indica el fin de la utilización del paquete.
- G) Tarjetas de control que indican el fin del trabajo.

Para trabajar con una terminal de teleproceso, además de conocer las instrucciones del paquete SPSS, se deberán conocer los comandos del editor que interactúa entre el usuario y la computadora. Se recomienda consultar los manuales existentes para aprovechar al máximo los recursos del sistema.

Los pasos que se deben llevar a cabo para poder hacer uso del paquete, por medio de una terminal, son los siguientes:

- A) Iniciar el proceso identificándose como usuarios autorizados por medio de la clave y contraseña. Si la identificación es correcta, el sistema dará el inicio de sesión.
- B) Crear un archivo que contenga únicamente instrucciones de SPSS y dejarlo grabado en disco.
- C) Crear un archivo con los datos en el orden, formato y número especificado en el programa SPSS. Dejarlo grabado en disco.
- D) Ejecutar el paquete por medio de las instrucciones con las que cuenta cada computador.
- E) Verificar en pantalla la correcta ejecución del programa, en caso de error, corregir ya sea los datos o el archivo de instrucciones, grabar en disco los archivos corregidos y volver a ejecutar. Repetir hasta obtener los resultados deseados.
- F) Imprimir los resultados.
- G) Indicar el fin del trabajo y el fin de sesión.

## 1.1 ARCHIVOS UTILIZADOS POR SPSS

- FILE1 Archivo que almacena la información asociada al programa SPSS, etiquetas, valores perdidos, formatos, subarchivos en código BCD lo que agiliza notablemente el procesamiento interno. Su permanencia es en disco. Es transparente al usuario (no interviene en su formación y tampoco puede transformarlo) y se destruye al terminar la corrida.
- FILE2 Almacena en código BCD, los datos del usuario (más tres variables del sistema). Archivo transparente al usuario. Al terminar la corrida se destruye.
- FILE3 Archivo de entrada a SPSS, contiene lo almacenado en FILE1 y FILE2. Por medio de la instrucción GET FILE se puede acceder (Ver capítulo X).
- FILE4 Archivo SPSS de salida, contiene a FILE1 y FILE2. Se graba con la instrucción SAVE FILE de SPSS (Ver capítulo X).
- FILE5 Archivo de entrada que contiene las instrucciones del programa SPSS.
- FILE6 Archivo de impresión de resultados a cualquier dispositivo de salida como son: cinta, disco, impresora o terminal.
- FILE8 Archivo de entrada que contiene los datos. Se usa cuando los datos vienen en un dispositivo diferente a la lectora de tarjetas. Se debe dar a SPSS las características con las que fue creado.
- FILE9 y FILE15 al FILE20 Archivos asociados a las opciones de grabado específicas de cada subrutina (ver la instrucción RAW OUTPUT UNIT).

En los archivos (excepto FILE1 y FILE2) el usuario puede controlar los atributos de archivo como son: el dispositivo físico en el que se encuentran, nombre físico, tiempo de permanencia en el sistema, formato de los registros, estructura del archivo, etc. Consultar los manuales existentes para cada sistema.

## 1.2 LA EJECUCION DEL PAQUETE.

- A) Con la instrucción de ejecución de SPSS, se activa la impresión de encabezados, inicialización de variables, tablas, etc.
- B) Comienza a leer el programa SPSS, verificando la sintaxis de cada tarjeta.
- C) SPSS va almacenando la información en el FILE1, asociándola con cada variable, etiquetas, valores faltantes, ponderación, subarchivo al que pertenece, etc.
- D) Acciona la lectura de los datos de acuerdo a las instrucciones VARIABLE LIST, INPUT FORMAT y N OF CASES, que indican cuáles, cómo y cuántos deberán leerse respectivamente.
- E) Si una vez leídos el programa SPSS y los datos no se encuentran errores, se llevan a cabo los análisis deseados, de otra manera producirá un listado de errores y cancelará la ejecución de los procedimientos.
- F) Si encuentra la tarjeta SAVE FILE, realiza la unión de los archivos FILE1 y FILE2 en el archivo FILE4, el cual se puede acceder posteriormente como FILE3.
- G) SPSS encuentra la instrucción FINISH y la ejecución termina, dejando los resultados en el FILE6 (generalmente impresora).

### 1.3 EJEMPLO DE RESULTADOS DE SPSS.

A continuación se presenta una salida común de SPSS. En ella se combinan 3 tipos de mensajes: los que manda SPSS, la impresión de las instrucciones proporcionadas por el usuario y los resultados de procedimientos.

SPSS BATCH SYSTEM

03/31/87 PAGE 1

SPSS WITH GRAPHICS OPTION FOR OS/360, VERSION N, RELEASE 9.1, FEBRUARY 1, 1982

CURRENT DOCUMENTATION FOR THE SPSS BATCH SYSTEM

ORDER FROM McGRAW-HILL: SPSS, 2ND ED. (PRINCIPAL TEXT) ORDER FROM SPSS INC.: SPSS STATISTICAL ALGORITHMS  
SPSS UPDATE 7-9 (USE W/SPSS, 2ND FOR REL. 7, 8, 9) KEYWORDS: THE SPSS INC. NEWSLETTER  
SPSS POCKET GUIDE, RELEASE 9  
SPSS INTRODUCTORY GUIDE: BASIC STATISTICS AND OPERATIONS  
SPSS PRIMER (BRIEF INTRO TO SPSS)

DEFAULT SPACE ALLOCATION.. ALLOWS FOR.. 230 TRANSFORMATIONS  
WORKSPACE 161280 BYTES 921 RECODE VALUES + LAB VARIABLES  
TRANSSPACE 23040 BYTES 3489 IF/COMPUTE OPERATIONS

1 RUN NAME EJEMPLO DE RESULTADOS DE SPSS.  
2 VARIABLE LIST REGION SEXO OCUP ESCO EMAD EXOCIVIL HIJOS  
3 EDREAL SUELDO  
4 INPUT FORMAT FIXED (F1.0,F2.0,F2.0,F2.0,F7.0)

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

VARIABLE	FORMAT	RECORD	COLUMNS
REGION	F 1. 0	1	1- 1
SEXO	F 2. 0	1	2- 3
OCUP	F 2. 0	1	4- 5
ESCO	F 2. 0	1	6- 7
EMAD	F 2. 0	1	8- 9
EXOCIVIL	F 2. 0	1	10- 11
HIJOS	F 2. 0	1	12- 13
EDREAL	F 2. 0	1	68- 69
SUELDO	F 7. 0	1	70- 76

THE INPUT FORMAT PROVIDES FOR 9 VARIABLES. 9 WILL BE READ  
IT PROVIDES FOR 1 RECORDS ('CARDS') PER CASE. A MAXIMUM OF 76 'COLUMNS' ARE USED ON A RECORD.

5 INPUT MEDIUM DISK  
6 N OF CASES UNKNOWN  
7  
8 RECODE EDREAL (15 THRU 19=1) (20 THRU 24=2) (25 THRU 29=3)  
9 (30 THRU 34=4) (35 THRU 39=5) (40 THRU 44=6)  
10 (45 THRU 50=7) (ELSE=99)

11 MISSING VALUES EDREAL (99)  
 12  
 13 VAR LABELS EDREAL EDAD REAL DEL ENTREVISTADO/  
 HIJOS TIENE HIJOS?  
 14  
 15 VALUE LABELS REGION (1)VALLE DE MEXICO (2)RURAL (3)URBANO/  
 16 SEXO (1)HOMBRES (2)MUJERES/  
 17 OCUP (1)EMPLEADO (2)HOGAR (3)ESTUDIANTE/  
 18 ESCO (1)SIN INSTRUCCION (2)PRIM INCOMP  
 19 (3)PRIM COMP (4)SEGUND O MAS/  
 20 EDREAL EDAD (1)15-19 (2)20-24 (3)25-29 (4)30-34

EJEMPLO DE RESULTADOS DE SPSS.

03/31/87 PAGE 2

21 (5)35-39 (6)40-44 (7)45 O MAS/  
 22 EDUCIVIL (1)ACT UNIDO (0)NO UNIDO/  
 23 HIJOS (1) SI (0)NO/  
 24 PAGESIZE 86  
 25  
 26 FRECUENCIES GENERAL= EDREAL  
 27 STATISTICS ALL

GIVEN WORKSPACE ALLOWS FOR 11520 VALUES AND 3456 LABELS PER VARIABLE FOR 'FRECUENCIES'

28 READ INPUT DATA

AFTER READING 2813 CASES FROM SUBFILE NONAME , END OF DATA WAS ENCOUNTERED ON LOGICAL UNIT N 8  
 EJEMPLO DE RESULTADOS DE SPSS.

03/31/87 PAGE 3

FILE NONAME (CREATION DATE = 03/31/87)

EDREAL EDAD REAL DEL ENTREVISTADO

CATEGORY LABEL	CODE	ABSOLUTE		RELATIVE ADJUSTED		CUM
		FREQ	FREQ	FREQ	FREQ	FREQ
15-19	1.	524	18.4	18.4	18.8	18.8
20-24	2.	517	18.4	18.6	37.4	
25-29	3.	472	16.8	16.9	54.3	
30-34	4.	428	15.2	15.4	69.7	
35-39	5.	346	12.3	12.4	82.1	
40-44	6.	228	8.1	8.2	90.3	
45 O MAS	7.	271	9.6	9.7	100.0	
	99.	27	1.0	MISSING	100.0	
	TOTAL	2813	100.0	100.0		

MEAN	3.475	STD ERR	0.036	MEDIAN	3.246
MODE	1.000	STD DEV	1.912	VARIANCE	3.657
KURTOSIS	-0.981	SKEWNESS	0.349	RANGE	6.000
MINIMUM	1.000	MAXIMUM	7.000		

VALID CASES 2786 MISSING CASES 27

EJEMPLO DE RESULTADOS DE SPSS.

03/31/87 PAGE 4

TRANSPIECE REQUIRED.. 500 BYTES  
 5 TRANSFORMATIONS  
 19 RECORD VALUES + LAG VARIABLES  
 9 IF/COMPUTE OPERATIONS  
 CPU TIME REQUIRED.. 1.76 SECONDS



Todo archivo de instrucciones del SPSS se puede dividir en tres partes:

- a) Definición de variables.
- b) Modificación y creación de variables.
- c) Ejecución de procedimientos.

Las instrucciones VARIABLE LIST, INPUT FORMAT, INPUT MEDIUM y N OF CASES entran en la primera parte ya que definen: cuáles son y en qué orden se leerán las variables, qué tipo son y con qué formato se leerán, en dónde se encuentra el archivo de datos y de cuántos casos consta, respectivamente.

En la segunda clasificación caen las instrucciones RECODE, MISSING VALUES, VAR LABELS y VALUE LABELS que se utilizan en el ejemplo.

Por último se dan los nombres de los procedimientos del SPSS, y sus parámetros asociados, que se desean ejecutar, colocando únicamente después del primero la instrucción READ INPUT DATA que indica la lectura del archivo de datos.

La última instrucción en un archivo de instrucciones en el SPSS es FINISH, que indica que ya no siguen más instrucciones y SPSS termina.



## II. DEFINICION DE CASO Y VARIABLE

El paquete tiene como entrada datos, observaciones o casos. El CASO es la unidad básica de análisis y usualmente es un individuo, tal vez una gran unidad como una ciudad, nación, institución, o más abstracto, como una condición experimental o un tiempo de prueba.

Cada unidad o caso se compone de valores para una o más medidas, que se denominan VARIABLES, cada caso dentro del estudio tendrá un valor para cada variable. El orden de éstas será el mismo para todos los casos y los datos podrán constar de uno o más registros.

Como ejemplo se puede mencionar a un ESTUDIANTE como caso y a su SEXO como variable que puede tomar los valores de FEMENINO o MASCULINO, EDAD con el rango de valores enteros comúnmente usado para años cumplidos de 0 a 99, ESCOLARIDAD y como valores posibles todos los niveles existentes: NINGUNA, PRIMARIA, SECUNDARIA, PREPARATORIA, LICENCIATURA, POSGRADO o aquéllos a los que se restrinja la población en estudio.

Las variables deben estar codificadas de tal manera que puedan leerse por computadora. Los códigos pueden tomar valores numéricos o alfanuméricos y es importante distinguirlos ya que las variables numéricas pueden tratarse matemáticamente y las rutinas que realizan procedimientos matemáticos no pueden analizar variables alfanuméricas.

Debido a los pocos procedimientos disponibles para variables alfanuméricas y el tiempo extra que requieren, se recomienda manejar sólo códigos numéricos o recodificarlos a numéricos.

Existen ocasiones que se requiere hacer referencia a la fuente original de los datos, una medida muy útil es colocar al inicio y/o al final de cada tarjeta una identificación del caso así como un número que identifique cada tarjeta dentro del caso, tratarlo como una variable y en caso de duda, las consultas serán más rápidas y eficientes.

### III. FORMATO DE INSTRUCCIONES.

Como en muchos sistemas y paquetes, el usuario controla el procesamiento de los datos por medio de instrucciones. En el paquete SPSS cada instrucción es específica, algunas definen y describen los datos, otras controlan las subrutinas estadísticas y otras permiten transformar los datos del archivo.

Aunque cada instrucción tiene una función única, comparten un formato similar que consta de dos partes:

- A) AREA DE CONTROL Se encuentra de las columnas 1 a la 15 y contiene las palabras de control que identifican la instrucción.
- B) AREA DE ESPECIFICACION Ocupa de las columnas 16 a 80 y todas las tarjetas necesarias para complementar las especificaciones, conteniendo los parámetros y argumentos requeridos por cada instrucción (ver la instrucción NUMBERED).

Las especificaciones son combinaciones de nombres, valores, llaves, etiquetas, delimitadores y operadores.

**NOMBRES** Son a elección del usuario y referencias: archivos, variables o subarchivos. Están limitados a 8 caracteres alfanuméricos debiendo ser el primero alfabético (subarchivos máximo 4). No se pueden utilizar como nombres ninguna de las palabras de control del paquete.

**VALORES** Son numéricos o alfanuméricos, éstos últimos siempre deben ir entre apóstrofes.

**LLAVES** No deben usarse como nombres las palabras especiales de SPSS como: TO, THRU, BY, WITH, LO, HI, ELSE, EQ, etc., ya que tienen una función predeterminada según el contexto. (Consultar el Apéndice I)

**ETIQUETAS** Se pueden etiquetar archivos, variables, valores de variables. Su longitud varía según el tipo de etiqueta.

**DELIMITADORES** Nombres, valores, llaves y etiquetas deben estar separados entre sí por delimitadores. Los comunes son el BLANCO y la COMA que son equivalentes e intercambiables. Hay delimitadores especiales como "()/=" (paréntesis redondos, diagonal, signo igual) que son utilizados según el formato específico de la instrucción. Los delimitadores especiales no necesitan delimitadores comunes, pero pueden insertarse para claridad.

**OPERADORES** Las instrucciones de transformación los requieren y son: "+", "-", "\*", "/" (más, menos, por y entre). No necesitan delimitadores comunes pero pueden insertarse.

## **IV. PROCEDIMIENTOS ESTADISTICOS DE SPSS.**

La finalidad de este capítulo, es hacer una presentación breve de las subrutinas que contiene SPSS para que el usuario pueda tener una visión de conjunto. La descripción detallada de cada una se encuentra en el capítulo XIII.

### **AGGREGATE**

Describe y resume la distribución de los valores de una variable, mediante estadísticas básicas, entre los miembros de un grupo. Permite hasta 4 variables de clasificación.

### **CONDESCRIPTIVE**

Calcula estadísticas descriptivas básicas que ayudan a conocer las características distribucionales de cada variable.

### **FREQUENCIES**

Obtiene reportes de frecuencias y estadísticas descriptivas.

### **CROSSTABS**

Genera tablas cruzadas, calcula estadísticas no paramétricas y el coeficiente de Pearson, basadas en ellas.

### **BREAKDOWN**

Obtiene medias estadísticas de subgrupos generados por una variable de criterio, estudiando así relaciones entre conjuntos.

### **T-TEST**

Realiza la prueba-t, es decir, una prueba de comparación de medias.

## **PEARSON CORR**

Produce coeficientes de correlación de Pearson que miden la relación lineal entre dos variables.

## **NONPAR CORR**

Obtiene los coeficientes no paramétricos de correlación de Spearman y/o Kendall.

## **SCATTERGRAM**

Grafica la relación entre dos variables y calcula estadísticas asociadas con regresión simple.

## **PARTIAL CORR**

Obtiene coeficientes de correlación parcial, que son una medida de asociación que describen la relación lineal entre dos variables bajo el efecto de otras.

## **REGRESSION**

Permite estudiar la relación lineal entre una variable dependiente y un conjunto de variables independientes, mediante una combinación lineal que hace posible predecir los valores de la variable dependiente.

## **ANOVA**

Efectúa análisis de varianza que mide los efectos de una o más variables o factores. También realiza análisis de covarianza que incluye variables continuas o covariables.

## **ONEWAY**

Realiza análisis de varianza de una vía, es decir, que considera una sola variable independiente o factor, cuenta además con pruebas de contrastes, de tendencia y diferentes pruebas estadísticas.

## **DISCRIMINANT**

En el Análisis Discriminante los casos están divididos en grupos y el análisis es usado para encontrar funciones de clasificación que mejor caracterizan las diferencias entre grupos. Se calculan los efectos de un conjunto de variables independientes en una dependiente nominal.

## **FACTOR**

Ayuda a determinar si una o más variables forman parte de un dimensión. Representa en un número menor de dimensiones un número mayor, en función al grado de correlación existente entre ellas.

## **CANCORR**

Ayuda a ver la relación entre un conjunto de variables independientes contra otro de dependientes, considerados así, como conjuntos.

## **GUTTMAN SCALE**

Analiza las características operacionales de una escala, para ver si cumple con los supuestos de unidimensionalidad y acumulatividad.

## V. TABLA DE PRECEDENCIA DE LAS INSTRUCCIONES DE SPSS.

Las Tarjetas de Control del SPSS deben seguir un orden llamado precedencia. La siguiente tabla muestra ese orden de mayor a menor. Ninguna instrucción de orden menor puede preceder a una de orden mayor, excepto para tarjetas de definición de procedimiento (precedencia 16 a la 23), que pueden repetirse. Con las instrucciones de la misma precedencia el orden no importa.

1	EDIT	12	COMPUTE	18	*WEIGHT
2	NUMBERED		IF	19	MISSING VALUES
3	RUN NAME		SELECT IF		PRINT FORMAT
	PRINT BACK		COUNT	20	RUN SUBFILES
	PAGESIZE		RECODE		RAW OUTPUT UNIT
4	GET FILE	13	END REPEAT	21	LIST CASES
5	FILE NAME	14	WEIGHT		TASK NAME
6	DATA LIST	15	VAR LABELS	22	Nombre de la Rutina
	VARIABLE LIST		VALUE LABELS	23	STATISTICS
7	SUBFILE LIST		DOCUMENT		OPTIONS
8	INPUT FORMAT	16	*SAMPLE	24	READ INPUT DATA
	INPUT MEDIUM	17	*COMPUTE		READ MATRIX
9	N OF CASES		*IF	25	SORT CASES
10	SAMPLE		*SELECT IF	26	SAVE FILE
11	DO REPEAT		*COUNT	27	FINISH
			*RECODE		



## VI. INSTRUCCIONES BASICAS.

### 6.1) COMMENT.

Sintaxis:       1                   16  
                 COMMENT        <texto de comentario>

Instrucción opcional que permite insertar comentarios en cualquier lugar de la impresión para dar claridad al trabajo. "COMMENT" puede usarse en cualquier parte del programa SPSS, excepto entre los datos o entre las rutinas de ejecución (FRECUENCIES, CROSSTABS, ANOVA, etc) y sus opciones asociadas (STATISTICS y OPTIONS). Se pueden usar cuantas instrucciones "COMMENT" se deseen y serán impresas en el orden y lugar en el que fueron colocadas.

El comentario se coloca a partir de la columna 16; puede contener cualquier caracter y extenderse por más de un renglón, dejando en blanco las columnas 1 a la 15.

Ejemplo:

COMMENT DESARROLLO DE UN PROBLEMA DE ESTADISTICA

### 6.2) EDIT.

Sintaxis:       1                   16  
                 EDIT

Al mandar un proceso, SPSS revisa la sintaxis de las instrucciones, si está correcta, las ejecuta y si encuentra error lo señala, deja de revisar esa instrucción y continúa con la siguiente. Esto puede ser un grave desperdicio de tiempo y de má-

quina si se trata de un programa muy largo o con archivos extensos de datos o varios archivos, contenidos, tal vez, en cintas magnéticas que hay que montar cada vez o tarjetas perforadas que hay que leer, y que el programa, finalmente, resulte con errores.

SPSS por medio de la instrucción EDIT permite ejecutar el programa, únicamente para revisión de sintaxis sin necesidad de leer los datos, por consiguiente sin impresión de resultados estadísticos y requiriendo un mínimo de tiempo de máquina. Al utilizar esta instrucción se deben omitir todas aquellas tarjetas de control que hagan referencia a los archivos de datos, en su caso, omitir los datos que le siguen a las instrucciones READ MATRIX o READ INPUT DATA, y las órdenes existentes para montado de cintas magnéticas o lectura de tarjetas perforadas.

La impresión de una corrida SPSS con EDIT incluirá una lista de las tarjetas de control y de todos los errores encontrados. Como SPSS deja de revisar una instrucción al encontrar un error, pudiera ser que tuviera más errores que sólo se detectarán hasta haber corregido el primero. EDIT no detecta el hecho de que falte alguna instrucción importante ni el orden de ellas, sólo el usuario podrá detectar errores tales como etiquetas o cálculos equivocados, o mal asignados, listas de valores o variables incompletas, o mal proporcionadas, a las diferentes instrucciones del paquete que las requieran.

Si se usa esta instrucción, deberá ser la primera.

### 6.3) NUMBERED.

Sintaxis:       1                   16  
                 NUMBERED        YES

SPSS por default supone que el usuario ha utilizado hasta la columna 80 de las tarjetas para dar información. Algunos sistemas utilizan de la columna 73 a la 80 para secuencia interna y si no se indica lo contrario, SPSS tratará de leerla como información que interferirá con el sistema. Para evitarlo se da NUMBERED y el parámetro YES a partir de la columna 16 y así solamente trabajará con las primeras 72 columnas.

Esta instrucción únicamente hace referencia al formato de las tarjetas de control y no a los archivos de datos. En caso de incluirla deberá ser la primera, a menos que vaya la instrucción EDIT, en cuyo caso, será la segunda.

### 6.4) RUN NAME.

Sintaxis:       1                   16  
                 RUN NAME        <etiqueta>

Instrucción opcional que identifica la corrida, imprimiendo al principio de cada hoja de salida una etiqueta como título. La etiqueta se coloca a partir de la columna 16, puede contener cualquier caracter y extenderse por más de un renglón, dejando en blanco las columnas 1 a la 15. Si RUN NAME se omite, aparecerá una línea en blanco.

Ejemplo:

RUN NAME   EJEMPLO DE UN TITULO

## 6.5) VARIABLE LIST.

Sintaxis: <sup>1</sup> VARIABLE LIST <sup>16</sup> <lista de variables>

VARIABLE LIST indica a SPSS las variables que serán leídas.

El orden de las variables señala el orden de entrada de los datos, el cual debe ser el mismo para todos los casos.

Los nombres de las variables son dados por el usuario los que deberán ser referenciados en todos los procesos, pueden estar compuestos de letras y números (sin caracteres especiales o blancos), limitados a 8 caracteres de longitud debiendo ser el primero alfabético. Si el número de caracteres fuera mayor, SPSS los trunca a 8. Cada variable debe tener un nombre único, el cual se convierte en palabra reservada. Acepta una lista máxima de 500 variables. Para nombrarlas se pueden utilizar varios registros a partir de la columna 16.

En la lista no se deben incluir los nombres de las variables que se vayan a generar durante el programa y deberán separarse por comas y/o blancos.

Cuando se han de nombrar muchas variables el trabajo se complica, SPSS da la facilidad del nombramiento IMPLICITO que consta de un prefijo alfabético y un número, la palabra TO, y el mismo prefijo alfabético con otro número: VARXXX TO VARYYY, donde "VAR" es un prefijo dado por el usuario, "XXX" y "YYY" números donde "YYY" debe ser mayor que "XXX". El prefijo y su número asociado no deben de exceder de 8 caracteres. El prefijo a la derecha e izquierda de "TO" debe ser idéntico. El número de variables declaradas serán los números entre "XXX" y "YYY" incluidos éstos.

Ejemplo:

VARIABLE LIST EDAD ABC01 TO ABC30 ALTURA NA1 TO NA25

Se indica la lectura de 57 variables: EDAD, ABC01, ABC02, ABC03, ABC04, ..., ABC29, ABC30, ALTURA, NA1, NA2, NA3, ..., NA24, NA25. Es importante hacer notar la diferencia entre la notación "01" y "1", el primero será "VAR01" mientras que el segundo "VAR1" y esos serán los nombres mediante los cuales serán referenciados.

## 6.6) INPUT FORMAT.

	1	16
Sintaxis:	INPUT FORMAT	FIXED (formato) FREEFIELD

Instrucción que indica a SPSS el tipo de las variables (numéricas o alfanuméricas), su localización en el archivo, su organización y su formato. Se encuentra asociada a VARIABLE LIST ya que la primera variable nombrada corresponderá al primer formato definido y así hasta que coincidan ambas listas.

Si se desea utilizar el FORMATO LIBRE, se da en la columna 16 la opción FREEFIELD. Sólo se requiere que la secuencia de las variables permanezca constante en todos los casos, sin correspondencia necesaria entre las columnas y las variables. Varios casos pueden ocupar la misma tarjeta. Los valores de cada variable deben estar separados por una coma o uno o más blancos. Las variables alfanuméricas deben, además, ir entre apóstrofes. El uso de este tipo de formato se recomienda para archivos de datos pequeños ya que en caso de error, mientras más grande sea el archivo, mayor dificultad implicará encontrarlo.

El **FORMATO FIJO** indica que los valores de cada variable se encuentran en las mismas columnas de las mismas tarjetas en todos los casos. Un caso puede constar de todas las tarjetas que se requieran. Se coloca en la columna 16 la palabra **FIXED** y a continuación, entre paréntesis, un formato cuyas especificaciones serán reconocidas por aquellos familiarizados con **FORTRAN IV**. Según este formato será cómo el sistema leerá los datos y cómo el usuario deberá colocarlos. Los elementos principales del formato son:

**nfw.d**      Variables numéricas reales o enteras donde

- "n" = número de variables adyacentes con el mismo formato.
- "w" = número de campos que ocupa la variable incluyendo el punto, los decimales y el signo (máximo 9).
- "d" = número de decimales (en caso de ser variable entera "d=0").

**nAw**      Variables alfanuméricas donde

- "n" = número de variables adyacentes con el mismo formato.
- "w" = número de campos que ocupa la variable alfanumérica (máximo 4).

**nX**      Indica saltar n columnas.

**n(/)**      Salto de "n" registros o tarjetas.

**n(formato)**      Es una manera de escribir en forma abreviada formatos repetitivos.

Ejemplo:

```
VARIABLE LIST PESO ALTURA EDAD SEXO OCUPACIO ESCOLARI
X01 TO X04
INPUT FORMAT  FIXED (2F5.2,F2.0,3A2,/,4(F8.2,Z1))
```

Formato que describe 10 variables por caso que serán leídas de la siguiente manera: el primer formato se asociará a la variable numérica real PESO que ocupará las columnas 1 a 5 incluyendo el punto y 2 decimales; a continuación la variable numérica real ALTURA en las columnas 6 a 10 con el mismo formato que PESO, EDAD variable numérica entera en las columnas 11 y 12 sin punto decimal; las variables SEXO, OCUPACIO Y ESCOLARI son alfanuméricas y ocupan 2 campos cada una; en el segundo registro de datos de cada caso se encuentran, en las columnas 1 a la 30, cuatro variables numéricas reales: X01, X02, X03 y X04 que tienen 8 campos cada una incluyendo punto decimal y 2 decimales, separadas entre sí por dos blancos. Este formato debe ser el mismo para todos los casos que se vayan a leer.

## 6.7) DATA LIST.

```
Sintaxis:  1          16
           DATA LIST  FIXED ((número de registros por caso))/
           <registro> <variables> <col.inicio>
           - <col.final> (<tipo variable>)
           <variable> ... / <registro> ...
```

Esta instrucción presenta otra opción de especificar la entrada de los datos, reemplaza a VARIABLE LIST e INPUT FORMAT por lo que es recomendable tener claro la utilización de estas tarjetas antes de usar DATA LIST. La única diferencia funcional que presenta es que no está definida para lectura en formato libre.

La manera de describir las variables y su formato ayuda a evitar errores pero requiere de mayor tiempo de preparación. Se coloca la opción FIXED en la columna 16 y a continuación, entre paréntesis, el número de tarjetas que se requerirán por caso. Después el delimitador "/" (slash) y el número del registro del cual se empezarán a leer los datos. Si sólo es una tarjeta por caso, el número de tarjetas, el slash y el número de registro se pueden omitir.

Se procederá a describir la primer variable dando su nombre, su columna de inicio, un "-" (guión) y la columna final (si sólo ocupa una columna se puede omitir la indicación de la columna final y el guión) y el tipo de la variable (SPSS toma por default que es numérica entera por lo que para alfanuméricas se indicará con "(A)" y para reales con "(n)" donde "n" es el número de decimales, máximo 5). Así se seguirán especificando las variables del mismo registro y las variables de los siguientes registros.

Ejemplo:

```
DATA LIST      FIXED (2)/1 SEXO 1(A) EDAD 3-4 PESO 5-9
              (2)/2 X001 TO X018 1-36 (A)
```

Se describen 2 tarjetas y 21 variables por caso. La variable SEXO es alfanumérica y ocupa la columna 1 del primer registro, EDAD es numérica ocupando las sig. 2 columnas, la variable PESO es numérica con 2 decimales, ocupa 5 columnas: de la 5 a la 9.

En el segundo registro se encuentran adyacentes 18 variables del mismo tipo (alfanumérico), ocupan de las columnas 1 a la 36, el número total de columnas debe ser un número entero, múltiplo del número de variables en la lista; cada una ocupa igual número de columnas y se calcula dividiendo el número total de campos



entre el número de variables, en este ejemplo: 36 campos entre 18 variables = 2 campos por variable.

Las diferencias que presenta el usar la tarjeta DATA LIST o VARIABLE LIST e INPUT FORMAT son:

- \* DATA LIST ayuda a evitar errores al dar las variables una por una junto con su tipo y localización.
- \* Con las tarjetas VARIABLE LIST e INPUT FORMAT se debe verificar el orden de las variables al mismo tiempo que sus especificaciones correspondientes, lo que es más lento.
- \* DATA LIST es recomendable cuando se van a leer pocas variables que se encuentren en varios registros.
- \* VARIABLE LIST e INPUT FORMAT es recomendable al usar la declaración "VARxxx TO VARYyy" y un formato repetitivo.

La elección de cualquiera de los 2 métodos depende del usuario y el tipo de datos con los que vaya a trabajar.

## 6.8) INPUT MEDIUM.

1	16
Sintaxis: INPUT MEDIUM	CARD
	DISK
	TAPE
	OTHER

Informa a SPSS el dispositivo periférico de entrada por medio del cual entrarán los datos al programa. La instrucción INPUT MEDIUM debe llevar una de cuatro opciones.

Con CARD los datos vienen junto con las tarjetas de control después de READ INPUT DATA, aunque el propósito es leer por medio de tarjetas perforadas, esta opción la puede utilizar una termi-

nal remota, indicando así que los datos acompañan a las tarjetas de control.

DISK indica que los datos se encuentran dentro de un archivo en disco. TAPE es usada cuando los datos se encuentran en cinta magnética y OTHER cuando los datos están en algún otro dispositivo disponible en las instalaciones al servicio del usuario.

Además del programa SPSS, el usuario siempre debe preparar algunas tarjetas orientadas al sistema operativo, informándole la localización exacta de los archivos utilizados, su factor de bloqueo, su nombre físico y otros atributos necesarios.

Ejemplo:

INPUT MEDIUM CARD

Se indica que los datos vienen por tarjetas y SPSS los buscará en un archivo de tales características.

## 6.9) N OF CASES.

Sintaxis: 1 N OF CASES 16 <número de casos>  
UNKNOWN

Informa al sistema el número de casos contenidos en el archivo de datos a procesar. Esta instrucción se omite si los datos tienen estructura de subarchivo y en su lugar va la instrucción SUBFILE LIST. Se coloca en la columna 16 el número de casos o la palabra UNKNOWN que sólo es permitida si la entrada de los datos es diferente a CARD, ya que por tarjetas al sistema no le es posible detectar el fin de archivo.

Ejemplo:

N OF CASES 500

Indica que son 500 los casos a procesar.

#### 6. 10) TASK NAME.

Sintaxis:       1                   16  
          TASK NAME           <etiqueta>

Permite al usuario agregar un título que identifique a cada una de las tareas de la corrida. El título se coloca a partir de la columna 16 y sólo por el resto del registro, puede contener cualquier caracter. Se imprimirá en la primera página de salida de la subrutina señalada y después del RUN NAME y el FILE NAME, si se omite, sólo aparecerá una línea en blanco. Debe colocarse inmediatamente antes de la instrucción de ejecución. Permanecerá vigente hasta el final de la corrida u otro TASK NAME.

#### 6. 11) STATISTICS.

Sintaxis:       1                   16  
          STATISTICS       <lista de estadísticas>  
                              ALL

Cada subrutina tiene asociadas ciertas estadísticas adicionales a sus cálculos e impresión normal; para obtener su ejecución, la instrucción STATISTICS debe colocarse después de la instrucción de ejecución.

Cada una de las estadísticas tiene asignado un número entero único y tanto las estadísticas cuanto los números varían de subrutina a subrutina. Las estadísticas deseadas se colocan a

partir de la columna 16 separadas por comas o blancos. Si no se desea su cálculo, simplemente se omite la instrucción STATISTICS, mientras que con la palabra ALL, se obtendrán todas.

Ejemplo:

```
STATISTICS 1 4 2 7
```

Esta instrucción acciona el cálculo de las estadísticas 1, 4, 2, 7 que en particular proporciona la subrutina a la que esta instrucción va asociada .

## 6. 12) OPTIONS.

Sintaxis:       1                   16  
          OPTIONS                <lista de opciones>

OPTIONS permite controlar las opciones de presentación y cálculo de resultados de la subrutina a la que está asociada. Cada subrutina cuenta con métodos de procesamiento que se pueden elegir, se identifican con números enteros únicos y tanto las opciones cuanto los números varían de subrutina a subrutina.

Las opciones deseadas se colocan a partir de la columna 16 y se separan por medio de comas o blancos. Todas las subrutinas dan algunas opciones por default, si éstas son las que se desean la instrucción OPTIONS se puede omitir. Se recomienda verificar cada opción y, en su caso, incluir las instrucciones asociadas que se requieran. No acepta la palabra ALL.

Ejemplo:

```
OPTIONS 5
```

Así se indica que se desea la ejecución de la opción 5 de la subrutina a la que esta instrucción está asociada .

### 6.13) RAW OUTPUT UNIT.

Sintaxis: RAW OUTPUT UNIT <sup>1</sup> <número del 15 al 20 ó 9> <sup>16</sup>

Algunas subrutinas cuentan con opciones que imprimen datos adicionales a la salida común de SPSS y esta instrucción indica cuál será el archivo lógico en el que estarán impresos. A partir de la columna 16 se da el número 9 ó del 15 al 20 que denota al FILE9; FILE15; FILE16, ..., FILE20. Si se pide alguna opción que requiera de esta tarjeta, y se omite, SPSS toma el FILE9 por default. La instrucción permanece hasta el final de la corrida u otro RAW OUTPUT UNIT. Debe preceder a la instrucción de ejecución que genera dicha salida (CONDESCRIPTIVE, NONPAR CORR, etc).

Como la impresión es a un archivo, el usuario deberá preparar una tarjeta de control, aparte del programa SPSS, orientada al sistema operativo del computador en particular utilizado que describa los archivos utilizados por SPSS.

### 6.14) READ INPUT DATA.

Sintaxis: READ INPUT DATA <sup>1</sup> <sup>16</sup>

Comando que informa al sistema que el usuario ha terminado de definir los datos y el primer trabajo estadístico y que está listo para empezar a leer los datos. Si la opción de INPUT MEDIUM es CARD, los datos deben ir a continuación. Sólo debe haber una instrucción READ INPUT DATA por programa y se coloca después de la primera instrucción de ejecución y sus tarjetas asociadas.

Los datos se deben proporcionar según el orden, formato y cantidad indicados por las instrucciones: VARIABLE LIST, INPUT FORMAT y N OF CASES, respectivamente.

6.15)

FINISH.

Sintaxis: FINISH

1

16

Instrucción obligatoria que indica que el trabajo de SPSS ha terminado y regresa el control al sistema. FINISH debe ser la última instrucción del programa y cuando SPSS la identifica, manda la lista de errores encontrados durante la corrida o en caso de no haberlos, el mensaje "NORMAL END OF JOB".

## VII. INSTRUCCIONES QUE CONTROLAN LA IMPRESION DE RESULTADOS.

Para la impresión de resultados, SPSS maneja ciertas opciones preestablecidas: supone los valores de las variables numéricos sin decimales, un largo de hoja de 55 líneas, impresión de las tarjetas de control e impresión de una tabla entre las variables y su formato de lectura. Si se desea modificar alguna o todas las opciones anteriores, se tienen las sig. instrucciones:

### 7.1) PRINT FORMAT.

Sintaxis: <sup>1</sup> PRINT FORMAT <sup>16</sup> <lista de variables> ( valor o A )

Esta instrucción informa al sistema el tipo de las variables y cómo serán impresos sus valores. El SPSS por default supone que todas las variables son numéricas enteras por lo que si son alfanuméricas o numéricas reales se debe indicar de la siguiente manera: colocar el nombre de la variable o lista de variables y "A" entre paréntesis si es alfanumérica o un valor entre paréntesis que indicará el número de decimales a ser impresos (máximo 5). Un slash ("/") separará una lista de nombres y su formato de impresión de otra. Se pueden utilizar cuantas tarjetas se requieran, dejando en blanco de las columnas 1 a la 15.

Ejemplo:

```
PRINT FORMAT SEXO TO OCUPACION (A) / INGRESO (2)
```

PRINT FORMAT indica que las variables que se encuentran en VARIABLE LIST entre "SEXO" y "OCUPACION", incluyendo éstas, son alfanuméricas y que la variable "INGRESO" es numérica y se desea a impresión con 2 decimales.

### 7.2)

### PAGESIZE.

Sintaxis:    1                    16  
          PAGESIZE            <número de líneas>  
                                  NOEJECT

Por medio de esta instrucción se puede controlar el número de líneas a imprimir por página. SPSS por default imprime 55 líneas antes de saltar de página, ya que éste es el tamaño de las formas comúnmente usadas, pero al trabajar con formas más pequeñas o grandes se debe indicar con PAGESIZE, colocando el número deseado de líneas por página a partir de la columna 16.

PAGESIZE cuenta con la opción NOEJECT con la cual no habrá un salto físico de hoja, es decir, imprimirá en forma continua, indicando el final lógico de cada página con una línea punteada. Tal vez esto no es muy bueno para la visualización de resultados, pero representa considerable ahorro de papel. El efecto de PAGESIZE dura hasta el final de la corrida o hasta otro PAGESIZE.

### 7.3)

### PRINT BACK.

Sintaxis:    1                    16  
          PRINT BACK            NO  
                                  CONTROL  
                                  FORMAT

Normalmente SPSS imprime las tarjetas de control y una tabla de correspondencia entre las variables y su formato, además de los resultados estadísticos en la salida de cada corrida. Mediante PRINT BACK el usuario puede controlar la información que se imprime dando una de las siguientes tres opciones:



Con NO, sólo se imprimirán los resultados de las subrutinas omitiendo las tarjetas de control y la tabla de correspondencia, CONTROL omitirá la impresión de la tabla de correspondencia e imprimirá tarjetas de control y resultados; FORMAT imprimirá resultados y tabla de correspondencia sin tarjetas de control. Si PRINT BACK se omite, la impresión será la normal de SPSS.

#### 7.4) LIST CASES.

Sintaxis: <sup>1</sup> LIST CASES <sup>16</sup> CASES= <n>/ VARIABLES= <lista de var>  
ALL

Imprime el contenido de los casos y variables que se especifiquen. Esto es muy útil cuando se desea verificar la correcta lectura de los datos, si las transformaciones y cálculos proporcionan los resultados esperados o simplemente los datos.

Con la opción CASES se indica el número de casos a imprimir, si se omite imprimirá los 10 primeros casos del archivo de datos, la lista de variables se da mediante VARIABLES, acepta la palabra ALL y si se omite imprimirá todas las variables.

Para que LIST CASES sea ejecutado, debe acompañarse de una tarjeta de procedimiento. Recodificaciones y transformaciones realizadas con anterioridad afectarán los reportes de LIST CASES.

Ejemplo:

```
LIST CASES CASES= 20/ VARIABLES= EDAD SEXO
```

Acciona la impresión de los valores de las variables EDAD y SEXO, pero solamente para los primeros 20 casos.

## VIII. INSTRUCCIONES DE ETIQUETADO PARA VARIABLES Y VALORES.

El SPSS permite "etiquetar" a las variables y sus valores, lo que es excelente para documentar tanto el archivo de instrucciones cuanto la salida de resultados.

### 8.1) VAR LABELS.

Sintaxis:           1                           16  
                  VAR LABELS           <variable> <etiqueta>/ ...

Instrucción que permite asociar una etiqueta a una o todas las variables del archivo, lo que permite documentar la salida y los cálculos obtenidos por SPSS. Siempre que una variable sea utilizada, su etiqueta se imprimirá junto a ella.

Se coloca el nombre de la variable a partir de la columna 16 y a continuación una etiqueta de hasta 40 caracteres. Un slash ("/") separa una variable y su etiqueta, de otra. Se pueden utilizar cuantas tarjetas como se requieran dejando en blanco las columnas 1 a la 15. Cada variable sólo puede tener una etiqueta y ésta puede contener cualquier caracter, con excepción de "/" y "," y formar cualquier palabra, incluso palabra reservada. La separación entre variables y etiquetas se da por coma o blancos.

Ejemplo:

```
VAR LABELS    11 EDAD DE LA PERSONA/ 12 PESO/ 13 ALTURA
```

En el momento que aparezca la variable "X1" se imprimirá junto a ella la etiqueta "EDAD DE LA PERSONA" y lo mismo sucederá para "X2" y "X3" con la impresión de sus respectivas etiquetas, a menos que se indique lo contrario por medio de las opciones de impresión de cada subrutina.

## 8.2) VALUE LABELS.

```
1          16
Syntax:  VALUE LABELS <lista de variables> (<valor 1>
          <etiqueta> (<valor 2>) <etiqueta> ...
          (<valor n>) <etiqueta>/ .../
          <lista de variables> ...
```

SPSS permite asignar etiquetas a los valores de todas las variables del archivo, las que son importantes tanto en la presentación cuanto en la documentación. Se coloca el nombre de la variable o variables que comparten las mismas etiquetas para sus valores a partir de la columna 16, el primer valor a etiquetar entre paréntesis y una etiqueta de hasta 20 caracteres (la subrutina CROSSTABS sólo toma los primeros 16), a continuación el siguiente valor entre paréntesis y su etiqueta y de igual manera hasta que se hayan etiquetado todos los valores deseados. Un slash ("/") separa cada conjunto de etiquetas de valores de otro y pueden estar compuesta por cualquier caracter excepto "/" (").

Siempre que estos valores vayan a impresión, sus etiquetas se imprimirán junto a ellos. Se puede utilizar la palabra ALL para indicar que las etiquetas de los valores son comunes para todas las variables. Los valores de variables alfanuméricas deben ir, además, entre apóstrofes.

Ejemplo:

```
VALUE LABELS SEXO ('M') MASCULINO ('F') FEMENINO /
RELIGION (1) PROTESTANTE (2) CATOLICO
(3) JUDIO (4) OTRO
```

El valor M de la variable SEXO se etiqueta con MASCULINO y F con FEMENINO, estos valores, por ser alfanuméricos, van entre apóstrofes. Para el caso de las etiquetas de los valores de la variable numérica RELIGION, el procedimiento es similar.

## IX. INSTRUCCION PARA CONTROL DE VALORES MISSING.

Es muy usual en la investigación social no obtener en todos los casos el valor de una variable debido a: NO QUISO CONTESTAR, NO SABE, NO APLICABLE o simplemente mala codificación. Por esto, el SPSS incluye esta opción de procesamiento que señala códigos incompletos, incorrectos o inútiles.

### 9.1) MISSING VALUES.

```
1          16
Sintaxis: MISSING VALUES <lista de variables> (<val missing>)/
          .../ <variable> (<valores missing>)/
```

Cada subrutina da un trato especial a estos valores, además de contar con opciones para su manejo. Se coloca a partir de la columna 16 el nombre de la variable o la lista de las variables y sus valores MISSING entre paréntesis y separados por comas. Un slash ("/") separa una lista de variables y valores de otra. Los valores alfanuméricos deben ir, además, entre apóstrofes. Se pueden utilizar todas las tarjetas que se requieran dejando en blanco las columnas 1 a la 15. Se puede utilizar la palabra ALL para hacer referencia a todas las variables y hacer comunes los valores MISSING definidos.

El usuario puede seleccionar como MISSING hasta 3 valores, o rangos por variable, pero se ahorra tiempo y trabajo si un conjunto predeterminado de valores es usado como MISSING para todas las variables.

**Ejemplo:**

**MISSING VALUES** EDAD (01) / YYY2 TO YYY8 (4,7,15-23)

Con esta instrucción se excluye el caso en que la edad tome el valor "0", lo que no tendría ningún sentido. Para las variables YYY2, YYY3, ..., YYY8 los valores MISSING son "4", "7" y del 15 al 23, esto es, en caso de que alguna de las variables tome cualquiera de estos valores se rechaza la observación y no interviene en el análisis, a menos que se especifique lo contrario por medio de las opciones de las subrutinas.



Sintaxis: SAVE FILE

Instrucción que almacena el Archivo de Protección creado por SPSS en disco bajo el nombre que señala la instrucción FILE NAME, en caso de no haberla, lo guarda bajo NONAME.

SAVE FILE debe colocarse en forma previa a FINISH. Una vez ejecutado el programa que la utilizó, se informará al usuario si fue correctamente salvado así como la lista de variables, subarchivos y número de casos que lo fueron. A la lista de variables SPSS agrega 3 más:

SEQNUM número de secuencia de cada caso.

SUBFILE A qué subarchivo corresponde cada caso.

CASWGT Peso o ponderación de cada caso.

Con la información que contienen estas variables, SPSS puede reorganizar los casos al acceder el archivo salvado por medio de otro programa que contenga la instrucción GET FILE. SEQNUM, SUBFILE y CASWGT no pueden ser modificadas por el usuario, pero sí las puede acceder y utilizar por medio de las instrucciones COMPUTE e IF.

Una tarjeta de control que informe al sistema operativo las características físicas del archivo salvado, debe acompañar al programa SPSS. El archivo estará identificado como FILE4.





## 10.5) LIST FILEINFO.

Sintaxis:	1	16
	LIST FILEINFO	VARLIST
		SORTVARS
		VARINFO
		LABELS
		SUBDIRECTORY
		COMPLETE
		DOCUMENTS

LIST FILEINFO proporciona junto con la salida impresa de resultados, un reporte de la estructura del programa SPSS que la lleva, esto es particularmente útil cuando se accedió algún archivo por medio de GET FILE y se desea conocer sus características. Cuenta con varias opciones que se colocan a partir de la columna 16 separadas por coma o blancos y son:

VARLIST proporciona la lista de variables.

SORTVARS produce la lista de variables ordenada alfabéticamente y un número asociado que indica la posición de la variable.

VARINFO genera información sobre valores missing, formatos de impresión (definido por PRINT FORMAT) y el status que señala variables permanentes (T) o temporales (F), lo que puede ser útil para detectar las transformaciones realizadas con anterioridad.

LABELS imprime etiquetas asociadas a variables y valores.

SUBDIRECTORY genera una lista de nombres y número de casos de cada subarchivo.

COMPLETE imprime la información de las opciones anteriores.

DOCUMENTS reproduce el texto del DOCUMENT original.

# 10.5.1) EJEMPLO DE GENERACION DE ARCHIVOS.

RUN NAME EJEMPLO DE COMO SALVAR UN ARCHIVO EN DISCO.  
 FILE NAME ARCHIVO1  
 VARIABLE LIST REGION SEXO OCUP ESCO EDAD EDOCCIVIL HIJOS SUELDO  
 INPUT FORMAT FIXED (F1.0,GF2.0,S6X,F7.0)  
 INPUT MEDIUM DISK  
 N OF CASES UNKNOWN  
 MISSING VALUES REGION (3)/ SEXO EDAD (99)  
 VALUE LABELS REGION (1)VALLE DE MEXICO (2)RURAL/  
 SEXO (1)HOMBRES (2)MUJERES/  
 OCUP (1)CAMPO (2)JEFES (3)OBREROS (4)OFICIOS  
 (5)EMPLEADO (6)VENDEDOR COMERC (7)SUBEMPL  
 (8)HOGAR (9)ESTUDIANTE/  
 ESCO (1)SIN INSTRUCCION (2)PRIM INCOMP  
 (3)PRIM COMP (4)SECUND O MAS/  
 EDAD (1)15-19 (2)20-24 (3)25-29 (4)30-34  
 (5)35-39 (6)40-44 (7)45 O MAS/  
 EDOCCIVIL (1)ACT UNIDO (0)IND UNIDO/  
 HIJOS (1) SI (0)NO/

DOCUMENT ESTE ARCHIVO QUEDA SALVADO CON 8 VARIABLES  
 QUE SON REGION, SEXO, OCUPACION, ESCOLARIDAD,  
 EDAD, ESTADO CIVIL, SI TIENE HIJOS Y SUELDO MENSUAL.

FRECUENCIAS GENERAL=SEXO

SEXO

CATEGORY LABEL	CODE	ABSOLUTE	RELATIVE		CUM
		FREQ	FREQ	FREQ	FREQ
		(PCT)	(PCT)	(PCT)	(PCT)
HOMBRES	1.	1389	49.4	49.4	49.4
MUJERES	2.	1421	50.5	50.6	100.0
	99.	3	0.1	MISSING	100.0
TOTAL		2813	100.0	100.0	

VALID CASES 2810 MISSING CASES 3

SAME FILE

AFTER READING 2813 CASES FROM SUBFILE ARCHIVO1, END OF DATA WAS ENCOUNTERED ON LOGICAL UNIT N 8  
 FILE ARCHIVO1 HAS BEEN SAVED WITH 11 VARIABLES..

SEQNUM SUBFILE CASNUM REGION SEXO OCUP ESCO EDAD EDOCCIVIL HIJOS  
 SUELDO

THE SUBFILES ARE..

NO OF  
 NAME CASES  
 ARCHIVO1 2813

FINISH

RUN NAME EJEMPLO DE COMO ACCESAR UN ARCHIVO GRABADO EN DISCO.  
 GET FILE ARCHIVO1

FILE ARCHIVO1 HAS 11 VARIABLES

THE SUBFILES ARE..

NO OF  
 NAME CASES

ARCHIVO1 2813

LIST FILE INFO VARINFO DOCUMENTS

VARIABLE	FORMAT	PERM	MISSING VALUES
SERIAL#	0	T	
SUBFILE	A	T	
CASSET	4	T	
REGION	0	T	3
SEXO	0	T	99
Ocup	0	T	
ESCO	0	T	
EDAD	0	T	99
EDOCIVIL	0	T	
HIJOS	0	T	
SUELDO	0	T	

DUMP OF DOCUMENTARY INFORMATION..

04/02/87 ESTE ARCHIVO QUEDA SALVADO CON 8 VARIABLES  
 QUE SON REGION, SEXO, OCUPACION, ESCOLARIDAD,  
 EDAD, ESTADO CIVIL, SI TIENE HIJOS Y SUELDO MENSUAL.

FRECUENCIAS GENERAL=SEXO

SEXO

CATEGORY LABEL	ABSOLUTE		RELATIVE		ADJUSTED		CUM	
	CASE	FREQ	FREQ	(PCT)	FREQ	(PCT)	FREQ	(PCT)
HOMBRES	1.	1389	49.4	49.4	49.4	49.4		
MUJERES	2.	1421	50.5	50.6	100.0	100.0		
	99.	3	0.1	MISSING	100.0	100.0		
TOTAL		2813	100.0	100.0				

VALID CASES 2810 MISSING CASES 3

FINISH

Sintaxis:     1                             16  
 WRITE CASES   (<formato>) <lista de variables>

WRITE CASES proporciona al usuario una impresión de los datos en el dispositivo que se indique mediante RAW OUTPUT UNIT, así el usuario puede aprovechar las facilidades de cálculo, recodificación y selección de SPSS utilizando los datos con otro paquete o programa.

Esta instrucción recibe el mismo tratamiento que un procedimiento y puede ir antes o entre ellos. El reporte que proporciona se verá afectado por las transformaciones realizadas en las variables o su selección previa a la ejecución.

A partir de la columna 16 se da el formato de impresión entre paréntesis y la lista de variables. El formato es de tipo FORTRAN (Ver INPUT FORMAT), con la diferencia que al utilizar el formato F no imprime puntos decimales y el total de campos únicamente incluirá el signo (si hay números negativos), la parte entera y la fraccionaria en base al máximo valor que toma la variable. Una especificación de formato, disponible sólo para WRITE CASES es dar un elemento constante alfanumérico entre apóstrofes, el cual se imprimirá en todos los casos en la columna que se indique.

Las variables en la lista pueden llevar cualquier orden, e incluso repetirse, y pueden anexarse las variables SUBFILE, SEQNUM y CASWGT.

Cuenta con la OPCION 1 e indica la omisión a impresión de los casos con valores missing.

## 10.7) WRITE FILEINFO.

```
Sintaxis      1          16
              WRITE FILEINFO VARIABLES=<lista de variables>/
              <opciones>/ CHAR=EBCD/ NUMBERED
              ALL      BCD
```

Proporciona la descripción del archivo de datos así como de variables y subarchivos, información complementaria a la generada por WRITE CASES.

A partir de la columna 16 se da la lista de variables (acepta la palabra ALL) y el tipo de información que se desea de ellas mediante una o varias de las siguientes opciones:

FILENAME El nombre del archivo y su etiqueta.

VARLIST La lista de las variables solicitadas.

SUBFLIST Nombres de los subarchivos y su número de casos.

NCASES Número de casos del archivo.

MISVALS Los valores missing de las variables solicitadas.

PRNTFMTS El formato de impresión correspondiente.

VARLABS Las etiquetas de las variables.

VALLABS Las etiquetas de los valores de las variables.

DOCUMENTS Reproduce el texto del DOCUMENT del programa.

ALL Obtendrá toda la información anterior.

Después de un slash ("/"), sigue la opción CHAR con la que se indica imprimir en código EBCDIC (Extended Binary Coded Decimal Interchange Code) con "EBCD" o en BCD (Binary Coded Decimal) con "BCD", si se omite, lo hará en el código default de la computadora. Por último, la opción NUMBERED que indica numeración secuencial de las tarjetas de las columnas 73 a 80, por default no numera.

## XI. CREACION DE SUBARCHIVOS.

El SPSS permite estructurar los datos leídos en subarchivos y ejecutar rutinas con datos agrupados en uno o más conjuntos, que son muestras del archivo general, dividido según alguna variable crítica a juicio del investigador.

### 11.1) SUBFILE LIST.

```
1          16
Sintaxis: SUBFILE LIST <nombre subarchivo> (<número casos>) /
          .../ <nombre subarchivo> (<casos>)
```

SUBFILE LIST informa el nombre del subarchivo y el número de casos que contiene. El nombre dado debe ser único y constar de hasta 4 caracteres alfanuméricos, debiendo ser el primero alfabético. Se deben dar en el mismo orden que se encuentran los casos en el archivo de datos.

Al trabajar con estructura de subarchivo N OF CASES se omite ya que indica que se está trabajando con un sólo archivo. Se pueden generar desde 2 hasta 100 subarchivos que se pueden procesar por separado, agrupados, o como uno sólo según las especificaciones de RUN SUBFILES.

Ejemplo:

```
SUBFILE LIST SB1(5) SB2(7) SB3(8) SB4(6)
```

Son 4 subarchivos llamados SB1, SB2, SB3 y SB4 con 5, 7, 8 y 6 casos respectivamente. El archivo de datos general contendrá en total 30 casos.

## 11.2) RUN SUBFILES.

Sintaxis:     1                   16  
          RUN SUBFILES        (< lista de subarchivos agrupados >)  
                              EACH  
                              ALL

Si los datos tienen estructura de subarchivo (esto es si se usó la instrucción SUBFILE LIST), RUN SUBFILES indicará al programa cómo se procesarán en las diferentes subrutinas, los diferentes subarchivos. Si se omite esta instrucción, SPSS procesará como si no hubiera estructura de subarchivo, y si se coloca, la instrucción permanece hasta el final del programa o hasta otro RUN SUBFILES.

El modo de proceso se indica a partir de la columna 16 con la palabra EACH que procesa cada archivo declarado en forma independiente, con la palabra ALL trabajará como si no existiera estructura de subarchivo, o el nombre del subarchivo o subarchivos separados por comas y entre paréntesis, los que procesará como si fueran uno sólo. Se pueden dar varias indicaciones de proceso separadas únicamente por coma o blancos.

Ejemplo:

```
SUBFILE LIST  SB1(5) SB2(7) SB3(8) SB4(6)
RUN SUBFILES  (SB1) (SB1,SB3) (SB2,SB1,SB4) ALL
CONDESCRIPTIVE VAR01 TO VAR15
```

La subrutina CONDESCRIPTIVE procesará siete veces, uno con los datos que pertenecen al subarchivo nombrado SB1, después y en forma conjunta con los que pertenecen al subarchivo SB1 y SB3, a continuación, tratando como un sólo subarchivo los datos de SB2, SB1 y SB4 (en este orden) y por último, la subrutina ejecutará el procedimiento estadístico para cada subarchivo.

## XII. TRANSFORMACION Y CREACION DE VARIABLES.

Durante el análisis de los datos, el investigador puede necesitar modificar las variables del archivo o crear nuevas en base a las ya existentes. SPSS cuenta con esta facilidad, que realiza en forma temporal o permanente, a través de las instrucciones RECODE, COMPUTE, IF, SELECT IF, WEIGHT, SAMPLE y COUNT.

Cuando se realiza una transformación temporal se coloca una de las instrucciones anteriores precedidas por un asterisco (\*IF, \*RECODE, etc) en forma previa a un proceso, afectándolo solamente a él. Si el programa tuviera SAVE FILE, sus condiciones no serán almacenadas al final de la corrida. Pueden ir todas las instrucciones de transformación temporal que se requieran antes de cada proceso y para todos los procesos.

Si un asterisco no precede a la instrucción, se indica una transformación o creación permanente, la cual si es almacenada por SAVE FILE y afecta a la variable para todos los procesos del programa. Debe colocarse antes de la primera instrucción de proceso, si el mismo proceso llevara los dos tipos de transformaciones, las permanentes deben preceder a las temporales.

Cuando se accesa por GET FILE un archivo que sufrió modificaciones permanentes, la opción VARINFO de LIST FILEINFO da a conocer el status de las variables y las modificaciones previas.

Aparte de las diferencias antes mencionadas, las instrucciones con asterisco en su inicio o sin él, realizan la misma función, por esto, su explicación se puede ver de manera conjunta.



## 12.1) RECODE y \*RECODE.

```
Sintaxis: 1 RECODE 16 <variables> (<valores>=<nuevo valor>)  
          ... (<valores>=<nuevo valor>)/ .../  
          <variables> ...
```

RECODE permite modificar el código original de una variable a uno más conveniente al tipo de datos y al tipo de análisis. Se puede cambiar desde uno hasta todos los valores, variables continuas convertirse en discretas o alfanuméricas en numéricas.

Se coloca a partir de la columna 16 el nombre de la variable o variables cuyos valores serán reasignados y a continuación, entre paréntesis, los valores o rangos de ellos y el valor al cual se van a recodificar separados por "=". Se pueden recodificar todos los valores que se deseen.

La forma de indicar la recodificación es igualando el valor original de la variable al nuevo valor. Cuando se desea indicar un rango de valores se puede utilizar la palabra THRU. Se tienen, además, varias opciones que facilitan la tarea, como las palabras LOWEST y HIGHEST, que hacen referencia al valor mínimo y máximo de la variable y se pueden abreviar, respectivamente, como LO y HI. La palabra ELSE, recodifica todos los valores que hasta ese momento no hayan sido afectados por la instrucción RECODE. Como la ejecución se realiza de izquierda a derecha, la opción ELSE debe ser la última de la lista.

Hay que tener precaución de no incluir un valor en varias listas, si se hiciera, será recodificado en la primera lista que se encuentre, así como con recodificar valores que pudieran ser resultado de recodificaciones anteriores.

Es común encontrar en la lista de valores de las variables blancos, los cuales son tratados normalmente por SPSS como ceros pero si el investigador desea diferenciarlos, se puede hacer con la palabra BLANK que recodifica los blancos en forma independiente, pero se debe colocar antes de recodificar al valor cero.

RECODE permite convertir variables alfanuméricas en numéricas, de manera explícita, nombrando los caracteres entre apóstrofes y el valor numérico que les será asignado o implícita, con la palabra CONVERT entre paréntesis que convierte los caracteres numéricos a sus numéricos correspondientes, '-' a 11, '& +' a 12 y cualquier otro a 13, en el caso de variables alfanuméricas de más de una columna, la transformación dependerá del código de la computadora utilizada.

Ejemplo:

```
RECODE X1 (5=1) X2 (4, 7, 9=90)
      X4 (1, 5, 22, 37 THRU 68 = 2)
      X5 (LOWEST THRU 9=1) (10 THRU HI=2)
      X7 TO X11 (BLANK=9) (ELSE=99)
      Y1 TO Y7 ( CONVERT )
      Y8 ('A', 'B', 'C'=1) (ELSE=2)
```

El ejemplo recodifica valores de 17 variables. El valor 5 de X1 a 1, los valores 4, 7, 9 de X2 a 90, X4 haciendo uso de la palabra THRU indica desde el valor 37 a 68 que se recodificará a 2 junto con los valores 1, 5 y 22. Todo el código de la variable X5 cambiará de la siguiente manera, desde su valor mínimo hasta 9 a 1 y desde 10 hasta el máximo a 2. Los blancos de las variables X7 a X11 tomarán el nuevo valor de 9 y todos los demás valores 99. Las variables alfanuméricas Y1 a Y7 se convertirán en numéricas en forma implícita y por último Y8, variable alfanumérica convertirá sus valores 'A', 'B', 'C' a 1 y todos los demás a 2.

## 12.2) COMPUTE y \*COMPUTE.

Sintaxis:       1                               16  
COMPUTE        <variable> = <expresión aritmética>

Esta instrucción crea o modifica una variable en base al resultado que tome la expresión aritmética indicada. La transformación es completamente incondicional y si la variable ya existía sus valores originales son reemplazados. La expresión aritmética es de tipo FORTRAN y puede contener nombres de variables ya existentes, operadores aritméticos, funciones matemáticas y constantes numéricas. El nombre de la variable y la expresión aritmética son separados por el signo "=". Para cada transformación, se utiliza un nuevo COMPUTE, pero si ésta necesita más de una tarjeta, se pueden utilizar dejando de las columnas 1 a la 15 en blanco.

Los operadores y las funciones disponibles son:

Operadores:	+	suma	Z=X+Y
	-	resta	Z=X-Y
	*	multiplicación	Z=X*Y
	/	división	Z=X/Y
	**	exponenciación	Z=X**Y
Funciones:	LN	logaritmo natural	Z=LN(Y)
	LG10	logaritmo en base 10	Z=LG10(Y)
	SQRT	raíz cuadrada	Z=SQRT(Y)
	SIN	seno (argumento en radianes)	Z=SIN(Y)
	COS	coseno (argumento en radianes)	Z=COS(Y)
	ATAM	arco tangente (arg en radianes)	Z=ATAM(Y)
	EXP	exponencial	Z=EXP(Y)
	RND	redondeo	Z=RND(Y)
	TRUNC	truncamiento	Z=TRUNC(Y)
	ABS	valor absoluto	Z=ABS(Y)
	MOD10	residuo de dividir entre 10	Z=MOD10(Y)

Los lugares indicados por X y Y pueden ser ocupados por constantes, variables o, a su vez, expresiones aritméticas; pero Z a la izquierda del signo "=", sólo por el nombre de una variable. Los paréntesis se pueden usar para indicar el orden de ejecución de las operaciones y cuando se omitan, las prioridades serán las siguientes:

1. Funciones
2. Exponenciales
3. Multiplicación y división
4. Suma y resta

Cuando se encuentren varias operaciones de igual prioridad, la ejecución será de izquierda a derecha, a excepción de la exponenciación que al encontrarse varias en forma sucesiva, será de derecha a izquierda. Los operadores matemáticos y los paréntesis no necesitan separadores, pero pueden insertarse blancos.

Ejemplo:

```
COMPUTE VAR3 = SIN (VAR2 + 35)  
*COMPUTE VAR01 = 1/VAR01
```

En el COMPUTE inicial se ejecuta en primer lugar la suma que se encuentra entre los paréntesis, al resultado se le aplica la función SEND, asignándolo a VAR3; esta transformación queda para el resto del trabajo. En el segundo ejemplo se obtiene el recíproco de VAR01, asignando el resultado a la misma variable, transformación que sólo afecta al proceso inmediato.

## 12.3)

## IF e \*IF.

Sintaxis: IF            1            16  
 ((exp lógica) <variable> = <exp aritmética>

Al igual que la instrucción COMPUTE, IF puede transformar y crear variables pero de manera condicional, es decir, en base a una expresión lógica que especifica el investigador, si es verdadera, la transformación es realizada y si es falsa, la variable permanece con su valor original.

La expresión aritmética cumple todas las características definidas para COMPUTE. La expresión lógica debe ir entre paréntesis y está formada por dos expresiones matemáticas a comparar y un operador relacional, combinándolos, con operadores lógicos.

Relacionales:	EQ	igual a	X EQ Y
	NE	diferente a	X NE Y
	LT	menor que	X LT Y
	LE	menor o igual que	X LE Y
	GT	mayor que	X GT Y
	GE	mayor o igual que	X GE Y

Lógicos:	AND	intersección	X AND Y
	OR	unión	X OR Y
	NOT	negación	X NOT Y

Los paréntesis se pueden usar para indicar el orden de ejecución, si se omiten, las prioridades serán como sigue:

1. Expresiones algebraicas
  - A. Funciones
  - B. Exponenciales
  - C. Multiplicación y división
  - D. Suma y resta
2. Operadores de relación
3. Operadores lógicos

Si se encuentran varias operaciones de igual prioridad, la ejecución se llevará a cabo de izquierda a derecha.

Para cada transformación se requiere una nueva instrucción IF, pero si la instrucción no alcanza en una tarjeta, se pueden utilizar todas las necesarias.

Ejemplo:

```
IF      ( V1 NE 0 ) V1 = V2/V1
IF      ( EDAD LE 6 AND NACIONAL EQ 1 ) ESCO=10
*IF     ( A LT B*(C+2) AND D EQ E/B ) V4 = A*V2+D
```

En el primer ejemplo se pregunta si V1 es diferente a cero y si esto es verdadero, evalúa a V1 como V2 entre el valor original de V1. El siguiente ejemplo pregunta si la variable EDAD es menor o igual que 6 y si NACIONAL es igual a uno, si y sólo si el resultado de ambas expresiones es verdadero, asigna el valor 10 a la variable ESCO. Por último, se pregunta si el valor de A es estrictamente menor que el valor que tome la expresión algebraica C+2 y luego por B (si no se hubieran colocado los paréntesis, la multiplicación se ejecutaría primero, por tener más alta prioridad que la suma) o si D es igual al valor de E entre B, basta que una de las dos relaciones sea verdadera para que V4 tome el valor de A\*V2+D, pero sólo será cierta esta asignación para el proceso siguiente por ser un IF temporal.

Cuando en una expresión lógica existen expresiones y operadores comunes, SPSS permite que estén implícitos, abreviando así expresiones largas y complicadas. Redefiniendo, una expresión lógica se compone de relaciones unidas por los operadores lógicos AND y OR y, estas a su vez, por expresiones aritméticas unidas por los operadores relacionales EQ, NE, GE, LE, GT y LT.

Denominaremos como sujeto a la primera expresión aritmética a la izquierda en la relación. Si el sujeto es el mismo en relaciones adyacentes puede ser abreviado, aún más, si el operador

relacional tiene el mismo tipo de relación que los sujetos (también puede ser implícito en vez de repetirse). Los operadores lógicos no pueden abreviarse.

Para abreviar la expresión lógica de un IF, se debe especificar la primera mención del sujeto (o sujeto y operador relacional) de la relación. SPSS utilizará el primer sujeto explícitamente establecido a la izquierda, en aquellos lugares de la relación donde se encuentre implícito. En el caso de operador implícito SPSS supondrá que es el operador de la condición de relación simple, más cercana a la izquierda.

#### Ejemplos:

1) IF (AREA GT 100 AND AREA LT 200) TIPO=2

puede abreviarse como

IF (AREA GT 100 AND LT 200) TIPO=2

2) IF (A LT B OR A LT C OR A LT D OR E LT F) I=5

puede ser abreviado como

IF (A LT B OR LT C OR LT D OR E LT F) I=5

y aún más, por tener LT como operador común

IF (A LT B OR C OR D OR E LT F) I=5

donde hay que hacer notar que E es el primer sujeto a la izquierda de F.

3) IF (V1 GT V2 AND V1 EQ V3 OR V4 LT V5 OR V4 LT V6) CONSUMO = 3.5

puede quedar en forma más breve de la siguiente manera

IF (V1 GT V2 AND EQ V3 OR V4 LT V5 OR V6)  
CONSUMO = 3.5

Finalmente se recomienda a los usuarios ser muy cuidadosos al construir expresiones lógicas y todavía más al abreviarlas.

```

1
16.
Syntax: COUNT <variable>=<lista variables>(<valores>)
/ ... / <variable>= ...

```

Existen ocasiones que el investigador desea conocer el número de veces que se repiten una o varias características dentro de un conjunto de variables, esto se puede calcular utilizando las instrucciones IF, COMPUTE y RECODE pero su desarrollo usualmente es laborioso. SPSS cuenta con la instrucción COUNT diseñada especialmente para la construcción de este tipo de escalas aditivas.

Se coloca la palabra COUNT a partir de la columna 1 y a partir de la 16 la variable a la cual le será asignada el cálculo, el signo igual y a continuación, las variables de las que dependerá el proceso, por último, entre paréntesis y separados por comas, los valores que determinarán el criterio de conteo.

La función de COUNT es contar cada vez que una variable de la lista tome un valor predeterminado, es decir, la variable de la izquierda del signo igual tendrá un valor, para cada caso, igual al número de variables de la lista que, para ese caso, toman valores iguales a los especificados.

La manera de dar la lista de valores es muy semejante a la de la instrucción RECODE, puede contener valores individuales, rangos de valores utilizando la palabra THRU o combinación de ambos separados por comas. Se pueden utilizar, además, las palabras LOWEST y HIGHEST y sus abreviaciones LO y HI, así como la palabra BLANK que distingue los blancos de variables numéricas, de los ceros. Para variables alfanuméricas no está permitido indicar rangos y cada uno de sus valores deberán ir entre apóstrofes.



Una facilidad adicional de COUNT es que permite especificar diferentes valores para cada variable de la lista. Más de una variable puede ser calculada en una instrucción COUNT y simplemente se separan una de otra por un slash ("/") sin requerir de un nuevo COUNT.

Ejemplo:

Supóngase que se ha preguntado a todos los elementos (máximo 10) de varias familias, su edad y se quiere saber cuantos adultos hay en cada una.

```
COUNT      ADULTO = EDAD01 TO EDAD10 (18 THRU HI)
```

Resolviendo el mismo problema con otras instrucciones de SPSS

```
COMPUTE    ADULTO = 0
IF         ( EDAD01 GE 18 ) ADULTO = ADULTO + 1
IF         ( EDAD02 GE 18 ) ADULTO = ADULTO + 1
:
:
IF         ( EDAD10 GE 18 ) ADULTO = ADULTO + 1
```

lo que fácilmente demuestra las ventajas de usar COUNT.

## 12.5) SELECT IF y \*SELECT IF

Sintaxis:       1                   16  
          SELECT IF            (<expresión lógica>)

En ocasiones, la investigación requiere seleccionar un subgrupo específico de la muestra para el análisis. SPSS cuenta con la instrucción SELECT IF que permite procesar solamente aquellos casos que cumplen cierto criterio.

La construcción de la expresión lógica de SELECT IF, es completamente similar a la de la instrucción IF y se coloca a partir de la columna 16 entre paréntesis.

Ejemplo:

```
SELECT IF ( XYZ GE 50 AND XYZ LT 100 )
```

Los casos que entrarán al análisis sólo serán aquellos cuya variable XYZ se encuentre entre 50 y 100, incluyendo el valor 50 y sin incluir el valor 100, los casos que no cumplan con este criterio, serán rechazados.

Si el programa contiene más de una instrucción SELECT IF (ya sean permanentes o temporales), se seleccionarán aquellos casos que satisfagan todos los criterios.

Se recomienda al usuario asegurarse de que existan casos que cumplan con los diferentes criterios a utilizar con la instrucción SELECT IF y no dar lugar a situaciones en que el paquete se quede sin datos.

## 12.6) WEIGHT y \*WEIGHT

Sintaxis:       1                   16  
                  WEIGHT            < variable >

En la investigación social es muy común llevar a cabo estudios con la finalidad de aplicar los resultados obtenidos a toda una población y también es común no trabajar con toda la población debido, principalmente, a limitaciones económicas, de tiempo o de personal. Así que el estudio se lleva a cabo en una subpoblación o muestra, elegida de forma que sea lo más representativa posible de la población y así poder inferir los resultados finales al total de la población.

La forma en que la muestra es elegida depende de un diseño conceptual (definir la población que se quiere estudiar), de un diseño estadístico (muestreo probabilístico\*) y de un diseño administrativo (recursos con que se cuentan).

Para utilizar la instrucción WEIGHT de SPSS, se proporciona el nombre de la variable que contiene los factores a partir de la columna 16. Esa variable se puede generar dando los factores a leer a SPSS junto con los datos o, mediante instrucciones IF y COMPUTE dentro del programa SPSS donde, finalmente, WEIGHT efectuará el cálculo. Cuando los casos sean procesados, el valor de esa variable determinará con qué peso cada caso será considerado en los procedimientos estadísticos.

\* Si se desea profundizar en las técnicas de muestreo se recomienda consultar los siguientes libros:  
Técnicas de Muestreo de William Cochran. Edit. CEECSA  
Muestreo de Encuestas de Leslie Kish. Edit. Trillas  
Teoría del Muestreo de Des Raj. Edit. FCE  
Teoría de Encuestas por Muestreo con Aplicaciones de Pandurang V. Sukhatme. Edit. FCE

La aplicación de WEIGHT puede ser para ponderar y/o para proyectar. Ponderar es darle a cada caso de la muestra el peso que proporcionalmente tiene con respecto a la población, bien sea respetando el número de casos de la muestra o bien, si el tamaño de la muestra es suficientemente grande, proyectar los casos incluidos en la muestra al tamaño de la población, dándole a cada caso el factor de ponderación que le corresponde. Los factores de ponderación van a depender de criterios que tome el investigador así como de las características importantes, para el estudio, de la población como podrían ser el Nivel Socioeconómico, la Edad, el Sexo, la Escolaridad, la Ocupación, etcétera.

Cuando el muestreo es probabilístico, cualquiera que sea el diseño empleado, WEIGHT sirve para aplicar a cada caso el peso que, por la inversa de la probabilidad que tuvo de ser seleccionado, represente de la población o estrato de que procede.

Es posible tener diferentes ponderaciones en un mismo archivo y aquella que se aplique será la asociada a la última instrucción WEIGHT (ya sea permanente o temporal) encontrada por SPSS.

#### Ejemplo:

Supongamos que se quiere realizar un estudio en el Área Metropolitana de la Cd. de México. Sabemos que en 1980 la población alcanzaba un total de 12'457'709 habitantes\* (sin contar la población rural colindante incluidos en los censos de municipios).

\* Los datos aquí presentados de la población de 1980, de 1986 y la distribución por Nivel Socioeconómico han sido obtenidos por el Instituto de Investigación de la Comunicación A.C. en base a la información proporcionada en los censos y métodos de proyección muy particulares. Estos resultados se pueden comparar con publicaciones de la ONU, del Colegio de México y del Buró de Investigación de Mercados en documento anexo al mapa mercadológico BIREA.

Se calcula que para 1986 se ha tenido un crecimiento del 42.57% lo que hace un total de 17'760'823 habitantes.

Para el estudio en cuestión, la característica importante de la población es el Nivel Socioeconómico que se reparte 8% en el Nivel Alto, 38% en el Nivel Medio y 54% en el Nivel Bajo.

Se han levantado 450 encuestas en una muestra previamente elegida por algún método, de las cuales 150 pertenecen al Nivel Alto, 150 al Medio y 150 al Bajo. Se podría preguntar porqué no se levantaron de esas 450 el 8%, 38% y 54% del Nivel Alto, Medio y Bajo respectivamente, porque si se hubiera hecho así, las encuestas del Nivel Alto serían pocas y no representativas mientras que las del Nivel Bajo demasiadas. Naturalmente, los criterios para determinar una submuestra demasiado reducida o bien, excesiva, dependen de los objetivos de la encuesta y los recursos disponibles.

Creando ahora los factores de ponderación y proyección dentro del programa SPSS.

```
IF      ONIVEL = 1) PESO = 450 * .08 / 150
IF      ONIVEL = 2) PESO = 450 * .38 / 150
IF      ONIVEL = 3) PESO = 450 * .54 / 150
COMPUTE PROY = 17760823 / 450
COMPUTE PESOPROY = PESO * PROY
WEIGHT  PESOPROY
```

## 12.7) SAMPLE y \*SAMPLE

Sintaxis:       1                               16  
                  SAMPLE                        < número >

Hay ocasiones en que el investigador se encuentra ante un archivo de datos tan grande que su procesamiento resulta demasiado caro. SPSS cuenta con la instrucción SAMPLE que permite trabajar con una submuestra del archivo original, obtenida mediante procedimientos aleatorios.

El número que se coloca a partir de la columna 16 indica el porcentaje del archivo que se va a analizar, debe ser mayor que cero y menor que uno (no es válido obtener cero% que representaría ningún caso ni 100% que sería la muestra original).

## 12.8) DO REPEAT y END REPEAT

Sintaxis:       1                               16  
                  DO REPEAT                    <parámetro1>=<lista de variables>/  
  <parámetro2>=<lista de variables>/  
  ... /  
                  <instrucción1>            <especificaciones>  
                  <instrucción2>            <especificaciones>  
                  !  
                  !  
                  END REPEAT

Al preparar el conjunto de instrucciones para SPSS, el usuario se ve, ocasionalmente, en la necesidad de repetir una o varias instrucciones determinado número de veces. SPSS cuenta con la opción DO REPEAT-END REPEAT, que permite reducir el número de tarjetas necesarias así como el tiempo de su preparación.

A partir de la columna 1 de una tarjeta se escribe DO REPEAT y a partir de la 16, el nombre de una variable que tendrá la fun-

ción de parámetro y que sustituirá a las variables que, finalmente, serán transformadas, a continuación un signo igual y la lista de variables que serán transformadas.

Los nombres de los parámetros siguen los mismos lineamientos que cualquier nombre de variable en SPSS, son variables temporales, es decir, sólo existen durante la ejecución de DO REPEAT - END REPEAT, deben ser distintas a cualquier otra variable del archivo y de los demás parámetros de un mismo bloque DO REPEAT - END REPEAT. Se pueden describir todos los parámetros necesarios, separados entre sí por un slash ("/") y utilizando a partir de la columna 16 de las siguientes tarjetas.

Las variables que serán sustituidas por los parámetros pueden ser variables ya existentes en el archivo o variables nuevas que serán creadas. Durante la ejecución SPSS crea duplicados del bloque, sustituyendo cada vez los parámetros por las variables indicadas. Cuando se usan más de 2 parámetros, sus listas deben contener el mismo número de variables ya que la sustitución se llevará a cabo en cada parámetro, por su variable correspondiente, para cada bloque ejecutado.

Una vez descritos los parámetros y sus listas de variables, se colocan las instrucciones de transformación, colocando los parámetros en el lugar de las variables que van a sustituir. Puede haber variables que se queden como constantes. Y finalmente, la instrucción END REPEAT cerrando el bloque.

Ejemplo:

```
RUN NAME      ANALISIS DE UNA SOCIEDAD DE CONSUMO
VARIABLE LIST VAR01 TO VAR05
INPUT FORMAT  FIXED (SF3,0)
INPUT MEDIUM  DISK
N OF CAGES    50
DO REPEAT     A=CONS1 TO CONS7/ B=1 2 3 4 5 6 7/
COUNT       A= VAR01 TO VAR05 (B)
END REPEAT
VAR LABELS    VAR01 BEBIDAS Q TOMA EN EL DESAYUNO/ VAR02 BEBIDAS Q TOMA A MEDIA MAÑANA/
              VAR03 BEBIDAS Q TOMA EN LA COMIDA/ VAR04 BEBIDAS Q TOMA A MEDIA TARDE/
              VAR05 BEBIDAS Q TOMA EN LA CENA/   CONS1 CONSUMO DIARIO DE LECHE/
              CONS2 CONSUMO DIARIO DE BEB-LECHE/  CONS3 CONSUMO DIARIO DE REF-EMBOT/
              CONS4 CONSUMO DIARIO DE AGUAS PREP/ CONS5 CONSUMO DIARIO DE JUGOS NAT/
              CONS6 CONSUMO DIARIO DE BEB-ALCOHOL/ CONS7 CONSUMO DIARIO DE OTRAS BEBIDAS/

VALUE LABELS  VAR01 TO VAR05
              (1)LECHE (2)BEBIDAS CON LECHE (3)REFRESCOS EMBOTELLADOS
              (4)AGUAS PREPARADAS (5)JUGOS NAT (6)BEBIDAS CON ALCOHOL (7)OTROS
```

La instrucción anterior abrevia la preparación de 7 instrucciones COUNT. Crea 7 indicadores de consumo denominados CONS1 a CONS7 en base al conteo de los valores 1 a 7 en las variables VAR01 a VAR05.

Las instrucciones que pueden aparecer entre las tarjetas DO REPEAT - END REPEAT son las siguientes:

```
IF           *IF
COMPUTE      *COMPUTE
RECODE       *RECODE
COUNT       *COUNT
SELECT IF    *SELECT IF
```

Los bloques que contengan transformaciones permanentes deben ir antes del primer procedimiento, los que contienen transformaciones temporales, antes del procedimiento que utilizará esas transformaciones. Tratar de colocar un bloque DO REPEAT - END REPEAT dentro de otro no está permitido. Deberán darse en forma independiente.



### XIII. DESCRIPCION DE SUBROUTINAS

#### 13.1) CONDESCRIPTIVE

Sintaxis:       1                   16  
          CONDESCRIPTIVE <lista de variables>  
                                  ALL

Al empezar a estudiar los datos es común tratar de conocer las características distribucionales de las variables, ya que éstas ayudarán al investigador a diseñar los siguientes análisis estadísticos.

SPSS cuenta con la subrutina CONDESCRIPTIVE que obtiene información de la distribución, variabilidad, medidas de tendencia central y otras estadísticas descriptivas de las variables indicadas. También permite al usuario obtener en un archivo de salida los valores estandarizados de cualquier conjunto de variables.

Se coloca la palabra CONDESCRIPTIVE en el área de control de una tarjeta y, a partir de la columna 16, aquellas variables de las que se desea obtener las estadísticas. Si se desea para todas las variables del archivo, se da simplemente la palabra ALL.

Las estadísticas disponibles para esta subrutina están asociadas a la instrucción STATISTICS y son:

1- Media	8- Skewness**
2- Error Estándar	9- Rango
5- Desviación Estándar	10- Mínimo
6- Varianza	11- Máximo
7- Kurtosis*	

\* Kurtosis da el grado de apertura de la curva definida por la distribución, si es cero la curva es normal, si mayor que cero la curva es más alta y delgada que si fuera normal, si menor que cero la curva es más baja y ancha.

\*\* Skewness mide la aproximación de la distribución a una normal desde el punto de vista de la simetría con respecto a su media, si es cero la curva es simétrica y normal, si mayor que cero los casos se agrupan a la izquierda de la media y si menor que cero a la derecha.

Si esta subrutina no tiene asociada la STATISTICS, se llevará a cabo el cálculo de **todas** las estadísticas.

Las opciones disponibles para CONDESCRIPTIVE son:

- 1- Para los cálculos toma en cuenta todos los valores, incluso aquellos definidos como missing.
- 2- Suprime la impresión de las etiquetas de las variables.
- 3- Ordena la impresión de los valores estandarizados en el archivo definido por la instrucción RAW OUTPUT UNIT.\*
- 4- Después de la última hoja de resultados de CONDESCRIPTIVE se imprime un directorio de referencia que informa al usuario la página donde se imprimieron las estadísticas de cada variable.

Si no se especifica la instrucción OPTIONS, las etiquetas de las variables serán impresas y los valores definidos como missing serán excluidos de los cálculos.

LIMITACION: no pueden ser nombradas o estar implícitas más de 500 variables en una instrucción CONDESCRIPTIVE.

- \* Los valores estandarizados son aquellos que le restan al valor original su media y lo dividen entre su desviación estándar, obteniendo así una nueva variable de media 0 y desviación estándar de 1. Estas nuevas variables son muy útiles si se desea observar diferencias entre medias y desviaciones estándar con otras variables, las que, en su forma original, no habrían sido comparables por diferencias en unidades y, por lo tanto, en media y desviación estándar.

La impresión será en registros de 80 columnas y en el orden en que se nombraron en la tarjeta CONDESCRIPTIVE. Si los valores no alcanzan en una tarjeta, se utilizarán todas las necesarias. Las variables SERIAL, número de registro, SUBFILE y CASNOY se imprimirán en todas las tarjetas de las columnas 1 a la 16 con los formatos F6.0, F2.0, M, F4.2 respectivamente. Los valores estandarizados estarán a continuación en las columnas 17 a 80 con un formato F8.5 sin punto decimal. El manejo de los valores missing será según la opción 1 de esta subrutina, si no es usada, los missing se representarán como 999.00000, si es seleccionada, entrarán a los cálculos como el resto de los valores.

RUN NAME EJEMPLO DE RESULTADOS UTILIZANDO LA OPCION 3  
 VARIABLE LIST VAR023 VAR026  
 INPUT FORMAT FIXED (F2.0,I,1,F1.0)  
 INPUT MEDIUM DISK  
 N OF CASES UNKNOWN  
 RECODE ALL (BLANK=9999)  
 MISSING VALUES VAR023 VAR026 (9999)

VAR LABELS VAR023 EDAD REAL DEL ENTREVISTADO/ VAR026 ESCOLARIDAD/  
 VALUE LABELS VAR026 (0)PRIMARIA INCOMP (1)PRIMARIA COMPLETA  
 (2)SECUNDARIA INCOMP (3)SECUNDARIA COMP (4)PREPA INCOMP  
 (5)PREPA COMPLETA (6)CARRERA TECNICA (7)LICENCI INCOMP  
 (8)LICENCI COMPLETA (9)POSGRADO

RAW OUTPUT UNIT 15  
 CONDESCRIPTIVE VAR023 VAR026  
 OPTIONS 3  
 READ INPUT DATA

VARIABLE	VAR023	EDAD REAL DEL ENTREVISTADO				
MEAN	29.864		STD ERROR	0.536	STD DEV	13.169
VARIANCE	173.418		KURTOSIS	1.969	SKEWNESS	1.336
RANGE	69.000		MINIMUM	14.000	MAXIMUM	83.000

VALID OBSERVATIONS - 604 MISSING OBSERVATIONS - 0

VARIABLE	VAR026	ESCOLARIDAD				
MEAN	3.816		STD ERROR	0.107	STD DEV	2.622
VARIANCE	6.873		KURTOSIS	-1.066	SKEWNESS	0.339
RANGE	9.000		MINIMUM	0.0	MAXIMUM	9.000

VALID OBSERVATIONS - 604 MISSING OBSERVATIONS - 0

2 Z-SCORES WERE WRITTEN ON LOGICAL UNIT 15 FOR 604 UNWEIGHTED CASES. 1 RECORDS OUTPUT PER ONE.  
 FORMAT IS (F6.0,F2.0,M,4,F4.2,0FB.5). TO INPUT Z-SCORES ONLY, WITHOUT RESEQUENCING DATA, USE (164,0FB.5).  
 MISSING Z-SCORES ARE OUTPUT AS 999.0. NON-MISSING BUT EXTREME Z-SCORES ARE TRUNCATED TO +99.0 OR -99.0

OUTPUT	UNWEIGHTED	NUMBER OF	NUMBER	RECORD	MISSING
VARIABLE	PER CASE	COLLAPSE	CASES		
SEQUEN	1	1	1-6	1	1
RECORD N	1	1	7-8	1	1
SUBFILE	1	1	9-12	1	1
CHRMST	1	1	13-16	1	1
VAR023	1	1	17-24	1	0
VAR026	1	1	25-32	1	0

## 13.2)

## FREQUENCIES

```

1
Syntaxis: FREQUENCIES      16
                                GENERAL = <lista de variables>
                                ALL
1
FREQUENCIES      16
                                INTEGER = <variable o lista> (min,max)
                                ALL (min,max)

```

Subprograma de SPSS que proporciona información básica sobre la distribución de las variables del archivo, así como tablas de distribución de frecuencias las que, además de constituir un documento de referencia, permiten al usuario analizar la variabilidad y verificar la correcta codificación y captura de las variables.

FREQUENCIES cuenta con dos modos de operación: General y Entero. El modo general trabaja con todo tipo de variables mientras que el modo entero sólo con variables numéricas enteras.

Para utilizar el modo general se coloca la palabra GENERAL a partir de la columna 16, el signo igual y la lista de variables que serán procesadas. Las variables pueden ser alfanuméricas, numéricas reales o enteras. Si se desea procesar todas las variables del archivo se puede dar simplemente la palabra ALL.

Para el modo entero se coloca la palabra INTEGER a partir de la columna 16, el signo igual, las variables a las que se obtendrá su distribución de frecuencias y su valor mínimo y máximo, entre paréntesis, a continuación de cada variable, o de una lista de variables si es que comparten el mismo rango. También se puede hacer uso de la palabra ALL para hacer referencia a todas las variables del archivo que deben ser enteras y compartir el mismo rango, el cual se colocará a continuación entre paréntesis. Si se

dan las condiciones necesarias, se recomienda ampliamente usar el modo INTEGER que representa un ahorro significativo de recursos ya que agiliza notablemente el procesamiento.

Las estadísticas que podrán ser solicitadas mediante una instrucción STATISTICS asociada a la subrutina FREQUENCIES son:

- |                        |             |
|------------------------|-------------|
| 1- Media               | 7- Kurtosis |
| 2- Error Estándar      | 8- Skewness |
| 3- Mediana             | 9- Rango    |
| 4- Moda                | 10- Mínimo  |
| 5- Desviación Estándar | 11- Máximo  |
| 6- Varianza            |             |

Si FREQUENCIES no lleva asociada la tarjeta STATISTICS, no se producirá ninguna estadística.

Las opciones disponibles para esta subrutina son:

- 1- Se incluirán todos los valores en el cálculo de tablas, estadísticas e histogramas, aun los definidos missing.
- 2- Omite la impresión de las etiquetas de los valores de las variables.
- 3- Ordena que la impresión de resultados y títulos se realice en 8 1/2 x 11 pulg. (tamaño carta).
- 4- La impresión se dirigirá al archivo definido mediante RAW OUTPUT UNIT, por lo que en impresora no se tendrá ningún resultado para esta subrutina. La impresión se realizará a tamaño carta.
- 5- Realizará la impresión de todas las tablas solicitadas en formato condensado. De esta manera se pueden imprimir hasta 132 valores por página, en lugar de los 20 del formato normal, ya que omite la impresión de las etiquetas de los valores, la frecuencia relativa (que incluye los missing en los porcentajes) y redondea los porcentajes al entero más cercano.
- 6- Si alguna tabla ocupa más de una página en formato normal, se imprimirá automáticamente en formato condensado, de lo contrario, utilizará el formato normal.
- 7- Omite la impresión de las tablas de frecuencias y genera únicamente las estadísticas. La impresión generada es del tipo que produce la subrutina CONDESCRIPTIVE pero en la que, además, se obtiene el cálculo de las estadísticas Mediana y Moda.

- 8- Para cada variable en FREQUENCIES se imprimirá su representación gráfica mediante un histograma.
- 9- Después de la última hoja de resultados de FREQUENCIES, se imprimirá un directorio de referencia que informa al usuario la página donde se imprimió la tabla de frecuencias y estadísticas de cada variable.

Si no se especifica OPTIONS, los valores missing serán excluidos del proceso así como las etiquetas de valores y variables; las tablas de frecuencias y estadísticas serán impresas y los resultados serán enviados a impresora y en formato normal.

LIMITACION: no pueden ser nombradas o estar implícitas más de 500 variables en una instrucción FREQUENCIES.

```

RUN NAME      EJEMPLO DE RESULTADOS EN FORMATO CONDENSADO Y NORMAL.
VARIABLE LIST EDREAL SEXO
INPUT FORMAT  FIXED ( F2.0,F1.0 )
INPUT MEDIUM  DISK
N OF CASES    UNKNOWN
RECODE        EDREAL (15 THRU 19=1) (20 THRU 24=2) (25 THRU 29=3)
              (30 THRU 34=4) (35 THRU 39=5) (40 THRU 44=6)
              (45 THRU 50=7) (ELSE= 99)

MISSING VALUES EDREAL (99)
VAR LABELS     EDREAL EDAD REAL
VALUE LABELS   EDREAL (1)15 A 19 ANOS (2)20 A 24 ANOS (3)25 A 29 ANOS
              (4)30 A 34 ANOS (5)35 A 39 ANOS (6)40 A 44 ANOS
              (7)45 A 50 ANOS/
              SEXO (1)MASCULINO (2)FEMENINO

COMMENT        ***** EN FORMATO CONDENSADO *****
              ***** OPCION 5 *****

FREQUENCIES   INTEGER= EDREAL (1,7)
OPTIONS       5
READ INPUT DATA

```

EDREAL	EDAD REAL		ADJ CUM				ADJ CUM			
	CODE	FREQ PCT PCT	CODE	FREQ PCT PCT	CODE	FREQ PCT PCT	CODE	FREQ PCT PCT		
1	524	19 19	4	428	15 70	7	271	10 100		
2	517	19 37	5	346	12 82					
3	472	17 54	6	228	8 90					

MISSING DATA

CODE	FREQ	CODE	FREQ	CODE	FREQ
WILD	27				

VALID CASES 2786 MISSING CASES 27

```

COMMENT        ***** EN FORMATO NORMAL *****
FREQUENCIES   GENERAL= EDREAL

```

CATEGORY LABEL	CODE	ABSOLUTE		RELATIVE	ADJUSTED	CUM
		FREQ	(PCT)	FREQ (PCT)	FREQ (PCT)	FREQ (PCT)
15 A 19 ANOS	1.	524	18.6	18.8	18.8	18.8
20 A 24 ANOS	2.	517	18.4	18.6	37.4	37.4
25 A 29 ANOS	3.	472	16.8	16.9	54.3	54.3
30 A 34 ANOS	4.	428	15.2	15.4	69.7	69.7
35 A 39 ANOS	5.	346	12.3	12.4	82.1	82.1
40 A 44 ANOS	6.	228	8.1	8.2	90.3	90.3
45 A 50 ANOS	7.	271	9.6	9.7	100.0	100.0
	99.	27	1.0	MISSING	100.0	100.0
TOTAL		2813	100.0	100.0		

VALID CASES 2786 MISSING CASES 27

RUN NAME EJEMPLO DE RESULTADOS  
 COMMENT UTILIZANDO EL FORMATO NORMAL, LA OPCION 8 QUE GENERA UN HISTOGRAMA Y LA IMPRESION DE TODAS LAS ESTADISTICAS.

VARIABLE LIST VAROZS  
 INPUT FORMAT FIXED (F1.0)  
 INPUT MEDIUM DISK  
 N OF CASES UNKNOWN  
 VAR LABELS VAROZS ESTADO CIVIL  
 VALUE LABELS VAROZS (1)SOLTERO (2)CASADO (3)VIUDO (4)DIVORCIADO (5)UNION LIBRE/

FRECUENCIES GENERAL= VAROZS  
 STATISTICS ALL  
 OPTIONS 8  
 READ INPUT DATA

VAROZS ESTADO CIVIL

CATEGORY LABEL	CODE	ABSOLUTE FREQ	RELATIVE	ADJUSTED	CUM
			FREQ (PCT)	FREQ (PCT)	FREQ (PCT)
SOLTERO	1.	276	45.7	45.7	45.7
CASADO	2.	288	47.7	47.7	93.4
VIUDO	3.	22	3.6	3.6	97.0
DIVORCIADO	4.	8	1.3	1.3	98.3
UNION LIBRE	5.	10	1.7	1.7	100.0
TOTAL		604	100.0	100.0	

VAROZS ESTADO CIVIL  
 CODE

```

1 |
1 | ***** ( 276)
1 | SOLTERO
2 | ***** ( 288)
1 | CASADO
3 | *** ( 22)
1 | VIUDO
4 | ** ( 8)
1 | DIVORCIADO
5 | ** ( 10)
1 | UNION LIBRE
1 | .....|.....|.....|.....|.....|
0 | 100 200 300 400 500
FREQUENCY

```

MEAN	1.656	STD ERR	0.031	MEDIAN	1.590
MODE	2.000	STD DEV	0.760	VARIANCE	0.578
KURTOSIS	5.112	SKEWNESS	1.760	RANGE	4.000
MINIMUM	1.000	MAXIMUM	5.000		

VALID CASES 604 MISSING CASES 0



## 13.3)

## AGGREGATE

1	16
Sintaxis: AGGREGATE	GROUPVARS= <variables o lista de var>/ VARIABLES= <variables o lista de var>/ AGGSTATS= <lista de estadísticas>/ RMISS= <valor>/ VARIABLES=.../AGGSTATS=.../RMISS=.../

Al convertirse la investigación social en algo más sofisticado y complejo, los estudios van requiriendo análisis en múltiples niveles, donde las características más importantes a medir son las que describen y resumen la distribución de los valores de una variable entre los miembros de un grupo. AGREGATE permite producir características de grupos creando un archivo en disco de estadísticas descriptivas obtenidas de variables provenientes de un archivo de casos individuales. El archivo en disco será creado en formato binario y estará listo para leerse por SPSS y utilizarse en un nuevo análisis (ver instrucción RAW OUTPUT UNIT).

El subprograma AGGREGATE requiere que el archivo de entrada se encuentre de manera que todos los casos pertenecientes a cada grupo se encuentren **contiguos** en el archivo. Si los grupos están definidos por una sola variable, debe haber tantos grupos como valores diferentes tenga la variable, si es más de una, cada grupo consistirá de aquellos casos donde los valores de todas las variables de agrupamiento sean **idénticos**.

El primer parámetro de AGGREGATE es la palabra "GROUPVARS=" que se coloca a partir de la columna 16 de la tarjeta y a continuación el nombre de hasta 4 variables que definirán los grupos en los cuales las estadísticas serán calculadas. Este parámetro es obligatorio y sólo puede aparecer una vez, finalizando con un slash "/". El siguiente parámetro es "VARIABLES=" al que le sigue

la lista de variables numéricas para las cuales las estadísticas serán calculadas. Se finaliza también con un slash. Le sigue "AGGSTATS=" que especifica las estadísticas descriptivas que serán producidas para cada variable de la lista anterior. Las estadísticas disponibles para AGGSTATS son:

VALIDN	Número de casos de cada grupo que no tengan valores missing en la variable.
SUM	Sumatoria de los valores de la variable en todos los casos del grupo.
MEAN	Media de la variable en el grupo.
SD	Desviación estándar de la variable en el grupo.
MAX	Valor máximo de la variable en el grupo.
MIN	Valor mínimo de la variable en el grupo.
SKEW	Asimetría de la variable en el grupo.
KURT	Kurtosis de la variable en el grupo.
PCTGT (n)	Porcentaje de casos en el grupo para los cuales los valores de la variable son mayores a "n".
PCTLT (n)	Porcentaje de casos en el grupo para los cuales los valores de la variable son menores a "n".
PCTBTN (n1,n2)	Porcentaje de casos en el grupo para los cuales los valores de la variable son mayores o iguales a "n1" o menores o iguales al valor "n2".

Si alguna estadística fuera incalculable SPSS le asigna el valor cero, el cual podría ser el resultado válido de alguna estadística y no habría forma de distinguirlos. Para evitar esto, se tiene el parámetro opcional "RMISS=" el cual especifica el valor a ser impreso en el lugar de una estadística incalculable. Se sugiere utilizar un número negativo o uno positivo grande que no sea un posible resultado de alguna estadística.

AGGREGATE permite repetir los parámetros VARIABLES, AGGSTATS y RMISS para producir diferentes estadísticas descriptivas para diferentes tipos de variables en el archivo.

AGGREGATE puede crear dos tipos diferentes de archivos dependiendo del tipo de análisis posterior que se contemple. El primer tipo imprime un registro por grupo conteniendo las esta-

dísticas solicitadas. Este tipo de archivo puede leerse posteriormente por SPSS o por otro programa estadístico y así poder generar análisis subsecuentes. A estos archivos se les denomina "verdaderos". El segundo tipo de archivo contiene la misma información que el primero pero repetida tantas veces como casos tenga el archivo original de datos y, en una corrida posterior, se puede agregar esta información a las variables originales del archivo de datos para generar análisis composicionales. A estos archivos se les denomina "composicionales".

Ambos tipos de archivos son escritos en registros binarios de 80 columnas y 20 datos por registro. El contenido de cada registro y el número de registros por caso, dependerá del tipo y número de las estadísticas solicitadas y del número de variables en VARIABLES. Un reporte siempre es impreso por AGGREGATE informando al usuario el contenido exacto y la posición secuencial de cada una de las variables en el archivo en disco.

Para leer un archivo 'verdadero' mediante SPSS se utilizan las siguientes instrucciones:

```
1          16
VARIABLE LIST <lista de variables>
INPUT FORMAT BINARY ( <n>V )
```

donde "n" es el número de variables a leer en formato binario.

Para agregar las variables de un archivo 'composicional' al archivo original:

```
1          16
GET FILE <nombre del archivo>
ADD VARIABLES <lista de variables>
INPUT FORMAT BINARY ( <n>V )
INPUT MEDIUM DISK
```

la instrucción N OF CASES no debe ser preparada ya que el número de casos debe coincidir al predefinido por el archivo accesado.

#### Estadísticas disponibles con la instrucción STATISTICS:

- 1- Ordena la impresión de un reporte conteniendo la identificación del grupo, el número total de casos y los valores de las variables de agrupamiento para cada grupo.
- 2- Imprime la información definida en la estadística 1 más el contenido exacto de cada caso en cada grupo. Hay que notar que si se trabaja con muchos grupos, será grande la información producida.
- 3- Imprime la información definida en la estadística 2 pero sólo para los primeros 10 casos por grupo.

#### Opciones disponibles con AGGREGATE:

- 1- Los valores missing de las variables definidas en alguna lista VARIABLES entrarán al cálculo de las estadísticas como cualquier otro valor.
- 2- Si un caso tiene algún valor missing para alguna variable en una lista VARIABLES, ese caso no será considerado para el cálculo de las estadísticas de ese grupo. Por esto, todas las estadísticas calculadas en ese grupo, serán en base al mismo número de casos válidos.
- 3- AGGREGATE identifica cada grupo con un número único y progresivo, mediante esta opción es posible usar como identificación adicional, el valor de las variables de agrupamiento.
- 4- El archivo en disco será creado tipo "COMPOSICIONAL".

Si no se utiliza OPTIONS, un caso que contenga para una variable valores missing, será excluido en el cálculo de las estadísticas para esa variable, pero será incluido para aquellas variables con valores no missing. Todos los valores de las variables de agrupamiento son válidos, si el investigador quisiera excluir aquellos considerados missing, deberá borrar del archivo de datos estos grupos. Los valores de las variables de agrupamiento no serán impresos y el archivo en disco creado será tipo "VERDADERO".

**LIMITACIONES:**

- 1- Hasta 4 variables de agrupamiento pueden definirse.
- 2- Cualquier variable numérica del archivo puede entrar a VARIABLES excepto aquellas usadas en GROUPVARS.
- 3- Se permiten hasta 25 listas de VARIABLES, AGGSTATS y RMISS.
- 4- Hasta 500 variables pueden ser definidas en total en todas las listas VARIABLES y hasta 100 en cada lista individual.

```

COMMENT      EJEMPLO DE RESULTADOS
VAR LABELS  MEN11 CONSIDERA CANAL 11 CANAL CULTURAL/
            CAL11 CALIFICACION CANAL 11 ENTRE 0 Y 10/

VALUE LABELS MEN11 (0)NO (1)SI/
RAW OUTPUT UNIT 9
AGGREGATE    GROUPVARS= MEN11/ VARIABLES=CAL11/
            AGGSTATS= VALIDN MEAN SD MAX MIN PCTGT(6)/ RMISS=9999/

STATISTICS  2
OPTIONS     3
READ INPUT DATA
    
```

GROUP VARIABLES..		AGGREGATE									
		MEN11 CONSIDERA CANAL 11 CANAL CULTURA									
GROUP-ID	1 TOTAL N	213	GROUP-VALUES		0.						
VARIABLE	VALID N		SUM	MEAN	ST-DEV	MAXIMUM	MINIMUM	SKEWNESS	KURTOSIS	PCT-GT	PCT-LT
CAL11	213			4.36	2.19	10.00	0.0			24.41	
GROUP-ID	2 TOTAL N	391	GROUP-VALUES		1.						
VARIABLE	VALID N		SUM	MEAN	ST-DEV	MAXIMUM	MINIMUM	SKEWNESS	KURTOSIS	PCT-GT	PCT-LT
CAL11	391			5.12	2.27	10.00	0.0			31.97	

**THE CONTENTS OF EACH CASE ON THE AGGREGATED OUTPUT FILE.,**  
 SEQUENTIAL POSITION      CREATED VARIABLE

- 1      AGGREGATION GROUP NUMBER
- 2      TOTAL NUMBER OF CASES IN AGGREGATION GROUP
- 3      MEN11    - GROUPING VARIABLE VALUE
- 4      CAL11    - VALIDN
- 5      CAL11    - MEAN
- 6      CAL11    - SD
- 7      CAL11    - MAX
- 8      CAL11    - MIN
- 9      CAL11    - PCTGT

A TRUE AGGREGATED OUTPUT FILE WAS WRITTEN ON LOGICAL UNIT 9 IT CONTAINS      2 AGGREGATED CASES  
 THE VARIABLES ON THE FILE GENERATED CAN BE READ INTO SPSS IN A SUBSEQUENT RUN USING THE 'BINARY' FORMAT SPECIFICATION

### 13.3.1)

### SORT CASES.

Sintaxis: 1 SORT CASES 16  
<lista de variables> (A o D)

SORT CASES permite al usuario grabar en disco un archivo del sistema SPSS, en donde la secuencia original de los casos ha sido reordenada según los valores de una o más variables del archivo.

Tiene dos objetivos principales: auxiliar a los usuarios a preparar archivos de entrada para AGGREGATE y poder definir una nueva estructura de subarchivos por medio de SUBFILE LIST.

El campo de especificación contiene los nombres de hasta 6 variables para cuyos valores los casos serán sorteados, le sigue el caracter A o D, entre paréntesis, que indica si el sorteo será en orden ascendente o descendente respectivamente.

Se permite que una lista de variables contenga algunas en orden ascendente y otras en orden descendente. Si todas las variables en SORT CASES van a ser sorteadas en orden ascendente, (A) puede omitirse.

Al sortear, los valores missing son tratados como cualquier otro valor y son sorteados en comparación con los demás valores de la variable. El utilizar variables alfanuméricas como criterio de ordenamiento puede crear confusión, por lo que se recomienda recodificarlas previamente en numéricas.

La tarjeta SORT CASES se coloca inmediatamente antes de SAVE FILE, y no tiene el menor efecto durante la corrida en la cual es ejecutado el ordenamiento. El sorteo es activado sólo cuando se encuentra la tarjeta SAVE FILE. Si el archivo tenía estructura de subarchivo, esta será borrada a partir del ordenamiento.

## 13.4)

## CROSSTABS

Sintaxis:	1 CROSSTABS	16 TABLES=<variable o lista de var> BY .. BY <var o lista>/ <var o lista> BY ...
	1 CROSSTABS	16 VARIABLES=<var o lista> (min,max) ... <var o lista> (min,max)/ TABLES=<var o lista> BY ... BY <var o lista>/ <var o lista> BY ...

El objetivo primordial para el investigador es estudiar conjuntos de relaciones entre dos o más variables. Para auxiliarlo en esta tarea, SPSS cuenta con la subrutina CROSSTABS que calcula e imprime tablas cruzadas de dos o más variables de clasificación (numéricas o alfanuméricas). Una tabla cruzada es una distribución conjunta de frecuencias, la cual puede ser estadísticamente analizada. CROSSTABS dispone, opcionalmente, de pruebas estadísticas de significancia, así como de medidas de asociación nominales, ordinales y la correlación de Pearson.

Como FRECUENCIAS, CROSSTABS trabaja en dos modos: Entero y General. El modo general trabaja con todo tipo de variables (numéricas o alfanuméricas) mientras que el modo entero únicamente con variables numéricas enteras que, aunque requiere de mayor preparación en sus tarjetas de control, si se dan las condiciones necesarias, se recomienda ampliamente sobre el modo general, ya que, representa un ahorro significativo de recursos, además de mayor rapidez en el procesamiento.

La especificación para el modo entero consta de dos partes: la primera define las variables que entrarán al proceso CROSSTABS y su rango de valores, la segunda define las tablas deseadas. El modo general sólo, requiere la definición de las tablas cuyo formato es idéntico en ambos modos.

En el modo entero, en la columna 16 de la tarjeta CROSSTABS, se coloca la palabra "VARIABLES=" seguida de una o más variables con sus respectivos valores mínimo y máximo entre paréntesis y separados por coma. Se puede dar un mismo rango para un conjunto de variables, colocándolo simplemente al final de la lista de variables. Para finalizar se coloca un slash ("/").

Para el modo general se coloca la palabra "TABLES=" a partir de la columna 16 de la tarjeta CROSSTABS y a continuación la especificación de las tablas deseadas haciendo referencia a cualquier variable previamente leída o generada. Para el modo entero TABLES va a continuación de VARIABLES y sólo puede hacer referencia a las variables definidas en VARIABLES. Para definir una tabla de 2x2 se colocan dos nombres de variables separadas por la palabra BY. Si se desean tablas de más dimensiones se agregan los nombres de las variables separadas por un BY. La primera variable mencionada será la variable renglón, la segunda la variable columna y después de la tercera serán variables control que irán controlando los valores de cada tabla. Si las especificaciones de las tablas no alcanzan en una tarjeta, se puede continuar en las siguientes ocupando únicamente de los campos 16 a 72 inclusive. Las tablas deben separarse una de otra con un slash "/". Varias tablas pueden definirse colocando más de una variable a la izquierda y derecha de cualquier BY, produciéndose así una tabla por cada variable a la izquierda del BY por cada variable a la derecha del BY.

Debido a que las hojas de impresión son de dos dimensiones, tablas de más dimensiones se presentan como varias subtablas de dos dimensiones. Las estadísticas que se soliciten serán impresas



a continuación de la tabla a la que se apliquen. Cuando se trate de tablas para 3 o más variables, las estadísticas serán condicionales a las dos variables de la subtabla bajo el nivel especificado por las variables control.

Ejemplo:

```
1          16
CROSSTABS  TABLES= VAR1 BY VAR2/
           X01 TO X15 BY EDAD SEXO/
           X BY Y BY Z W/
```

Para este ejemplo CROSSTABS produciría 33 tablas, 31 de dos dimensiones y dos de tres dimensiones. La primera será un cruce de VAR1 contra VAR2, la segunda será X01 contra EDAD, después la variable X01 contra SEXO, X02 contra EDAD, X02 contra SEXO y así hasta la tabla X15 contra SEXO, por último CROSSTABS generará las tablas de tres dimensiones X contra Y contra Z y X contra Y contra W.

#### LIMITACIONES:

- \* Hasta 20 listas separadas por slash ("/") puede contener una instrucción TABLES.
- \* Sólo serán impresos los primeros 16 caracteres, de los 20 permitidos en las etiquetas de los valores de las variables.
- \* Hasta 200 variables pueden nombrarse o estar implícitas, en modo general, en la tarjeta TABLES.
- \* Hasta 250 valores puede tomar una variable en modo general.
- \* Tablas de hasta 10 dimensiones pueden pedirse en modo general.
- \* Hasta 100 variables se pueden nombrar o estar implícitas en modo entero en la instrucción VARIABLES y por lo tanto en TABLES.
- \* Hasta 200 valores puede tomar una variable en modo entero.
- \* Tablas de hasta 8 dimensiones pueden producirse en modo entero.

**Estadísticas que podrán ser seleccionadas:**

- 1- Ji-cuadrada. Para tablas de 2x2 se aplicará la corrección de Yates y, con menos de 21 casos, se aplicará la prueba Fisher.
- 2- Phi para tablas de 2x2 y Cramers'V para tablas más grandes.
- 3- Coeficiente de contingencia.
- 4- Lambda simétrica y asimétrica.
- 5- Coeficiente de incertidumbre simétrico y asimétrico.
- 6- Tau B de Kendall.
- 7- Tau C de Kendall.
- 8- Gamma parcial y de orden cero para tablas de 3xn
- 9- D de Somers simétrica y asimétrica.
- 10- Eta (disponible sólo para variables numéricas).
- 11- Coeficiente de correlación de Pearson.

Si CROSSTABS no lleva asociada la instrucción STATISTICS, no se producirá ninguna estadística.

Es responsabilidad del usuario aplicar aquellas estadísticas que sean apropiadas a los datos ya que el tipo de análisis dependerá de las características de las variables.

Las opciones disponibles para esta subrutina son:

- 1- Incluye los valores declarados como missing en las tablas y también en el cálculo de las estadísticas.
- 2- Suprime la impresión de las etiquetas de las variables y las de los valores de las variables.
- 3- Suprime la impresión de los porcentajes por renglón.
- 4- Suprime la impresión de los porcentajes por columna.
- 5- Suprime la impresión del porcentaje total.
- 6- Suprime únicamente la impresión de las etiquetas de los valores de las variables.

- 7- Incluye los missing en las tablas más no en el cálculo de las estadísticas ni en los porcentajes por renglón y por columna.
- 8- La variable renglón se imprimirá en orden descendente.
- 9- Ordena la impresión de un directorio de las tablas producidas y la página en la que se encuentran.

Si CROSSTABS no lleva asociada la tarjeta OPTIONS, los valores missing se excluirán de las tablas y del cálculo de las estadísticas, las etiquetas de variables y valores serán impresos, todas las tablas llevarán porcentajes por renglón, columna y total, la variable renglón será impresa en orden ascendente y el directorio no se imprimirá.

```

RUN NAME      EJEMPLO DE TABLA CRUZADA PRODUCIDA POR CROSSTABS.
VARIABLE LIST REGION SEXO OCUP ESCO EMBD EMO CIVIL HIJOS SUELDO
INPUT FORMAT  FIXED (F1.0,F2.0,S6X,F7.0)
INPUT MEDIUM DISK
N OF CASES   UNKNOWN
  
```

```

MISSING VALUES REGION (3) / SEXO EMBD (99)
  
```

```

VALUE LABELS  REGION (1)VALLE DE MEXICO (2)AURAL/
              SEXO (1)HOMBRES (2)MUJERES/
              OCUP (1)CAMPO (2)JEFES (3)GERENOS (4)OFICIOS
              (5)EMPLEADO (6)VENDEDOR COMERC (7)SUREPL
              (8)HOGAR (9)ESTUDIANTE/
              ESCO (1)SIN INSTRUCCION (2)PRIM INICOP
              (3)PRIM COM (4)SECUND O PRIM/
              EMBD (1)15-19 (2)20-24 (3)25-29 (4)30-34
              (5)35-39 (6)40-44 (7)45 O MAS/
              EMO CIVIL (1)ACT UNIDO (0)NO UNIDO/
              HIJOB (1) SI (0)NO/
  
```

```

PAGESIZE      86
CROSSTABS     TABLES= SEXO BY EMBD
STATISTICS    ALL
READ INPUT DATA
  
```

\*\*\*\*\* CROSSTABULATION OF \*\*\*\*\*  
 SEXO BY EDAD  
 \*\*\*\*\* PAGE 1 OF 1

		EDAD							
COUNT		115-19	20-24	25-29	30-34	35-39	40-44	45 O MAS	ROW
COL. PCT									TOTAL
TOT. PCT		1.1	2.1	3.1	4.1	5.1	6.1	7.1	
SEXO	1.	295	245	210	213	153	118	135	1389
		21.2	17.6	15.1	15.3	11.0	8.5	11.2	49.4
		55.9	47.5	44.5	49.8	44.3	31.8	52.9	
		10.5	8.7	7.5	7.6	5.4	4.2	5.5	
MUJERES	2.	233	271	262	215	192	110	138	1421
		16.4	19.1	18.4	15.1	13.5	7.7	9.7	50.6
		44.1	52.5	55.5	50.2	55.7	48.2	47.1	
		8.3	9.6	9.3	7.7	6.8	3.9	4.9	
COLUMN		528	516	472	428	345	228	293	2810
TOTAL		18.8	18.4	16.8	15.2	12.3	8.1	10.4	100.0

CHI SQUARE = 19.64243 WITH 6 DEGREES OF FREEDOM SIGNIFICANCE = 0.0032  
 CRAMER'S V = 0.06361  
 CONTINGENCY COEFFICIENT = 0.08332  
 LAMBDA (ASYMMETRIC) = 0.06263 WITH SEXO DEPENDENT. = 0.01645 WITH EDAD DEPENDENT.  
 LAMBDA (SYMMETRIC) = 0.03405  
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.00506 WITH SEXO DEPENDENT. = 0.00184 WITH EDAD DEPENDENT.  
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.00269  
 KENDALL'S TAU B = 0.01836 SIGNIFICANCE = 0.1334  
 KENDALL'S TAU C = 0.02389 SIGNIFICANCE = 0.1334  
 GINVA = 0.02818  
 SOMERS'S D (ASYMMETRIC) = 0.01410 WITH SEXO DEPENDENT. = 0.02390 WITH EDAD DEPENDENT.  
 SOMERS'S D (SYMMETRIC) = 0.01774  
 ETA = 0.06361 WITH SEXO DEPENDENT. = 0.01290 WITH EDAD DEPENDENT.  
 PEARSON'S R = 0.01291 SIGNIFICANCE = 0.2449

NUMBER OF MISSING OBSERVATIONS = 3

CROSSTABS TABLES- SEXD BY EADG BY REGION

\*\*\*\*\* CROSSTABULATION OF \*\*\*\*\*  
 SEXD  
 CONTROLLING FOR.. BY EADG  
 REGION VALLE = 1. VALLE DE MEXICO  
 \*\*\*\*\* PAGE 1 OF 1 \*\*\*\*\*

		EADG							
		COUNT							
		115-19	20-24	25-29	30-34	35-39	40-44	45 O MS	ROW
		COL PCT							TOTAL
SEXD	TOT PCT	1.1	2.1	3.1	4.1	5.1	6.1	7.1	
HOMBRES	1.	52	34	36	32	20	13	14	201
		25.9	16.9	17.9	15.9	10.0	6.5	7.0	53.7
		67.5	48.6	52.2	51.6	46.5	52.0	50.0	
		13.9	9.1	9.6	8.6	5.3	3.5	3.7	
MUJERES	2.	25	36	33	30	23	12	14	173
		14.5	20.8	19.1	17.3	13.3	6.9	8.1	46.3
		32.5	51.4	47.8	48.4	53.5	48.0	50.0	
		6.7	9.6	8.8	8.0	6.1	3.2	3.7	
COLUMN		77	70	69	62	43	25	28	374
TOTAL		20.6	18.7	18.4	16.6	11.5	6.7	7.5	100.0

\*\*\*\*\* CROSSTABULATION OF \*\*\*\*\*  
 SEXD  
 CONTROLLING FOR.. BY EADG  
 REGION VALLE = 2. RURAL  
 \*\*\*\*\* PAGE 1 OF 1 \*\*\*\*\*

		EADG							
		COUNT							
		115-19	20-24	25-29	30-34	35-39	40-44	45 O MS	ROW
		COL PCT							TOTAL
SEXD	TOT PCT	1.1	2.1	3.1	4.1	5.1	6.1	7.1	
HOMBRES	1.	134	114	108	116	77	67	104	720
		18.6	15.8	15.0	16.1	10.7	9.3	14.4	49.3
		53.6	46.2	42.4	52.0	45.3	52.8	55.3	
		9.2	7.8	7.4	7.9	5.3	4.6	7.1	
MUJERES	2.	116	133	147	107	93	60	84	740
		15.7	18.0	19.9	14.5	12.6	8.1	11.4	50.7
		46.4	53.8	57.6	48.0	54.7	47.2	44.7	
		7.9	9.1	10.1	7.3	6.4	4.1	5.8	
COLUMN		250	247	255	223	170	127	188	1440
TOTAL		17.1	16.9	17.5	15.3	11.6	8.7	12.9	100.0

NUMBER OF MISSING OBSERVATIONS = 979

### 13.5)

### BREAKDOWN

```
1
Sintaxis: BREAKDOWN          16
                                TABLES=<variable o lista de var> BY ..
                                BY <var o lista>/ <var o lista> BY ...

1
                                16
BREAKDOWN                    VARIABLES=<var o lista> (min,max) ...
                                <var o lista> (min,max)/
                                TABLES=<var o lista> BY ... BY
                                <var o lista>/ <var o lista> BY ...
```

Técnica simple para examinar las medias y varianzas de una variable de criterio o dependiente, entre una muestra o toda la población. Cruces donde cada media y desviación estándar resumen la distribución de un renglón o columna de una tabla de contingencia. Dada una variable numérica dependiente (continua o discreta) para la cual la media es una medida razonable de tendencia central, BREAKDOWN permite obtener medias, desviación estándar y varianzas de la variable de clasificación permitiendo de una hasta 5 variables independientes (nominales u ordinales).

Como FREQUENCIES y CROSSTABS, BREAKDOWN trabaja en dos modos: Entero y General. El modo general trabaja con todo tipo de variables independientes (numéricas o alfanuméricas) mientras que el modo entero únicamente con variables numéricas enteras. En ambos modos las variables dependientes deben ser numéricas. Si se dan las condiciones, se recomienda usar el modo entero sobre el modo general porque aunque requiere de mayor preparación en sus tarjetas de control, es notablemente más rápido en su proceso.

La especificación para el modo entero consta de dos partes: la primera define las variables que entrarán al proceso BREAKDOWN y su rango de valores, la segunda define las tablas deseadas. El modo general sólo requiere la definición de las tablas cuyo formato es idéntico en ambos modos.

En el modo entero, en la columna 16 de la tarjeta BREAKDOWN, se coloca la palabra "VARIABLES=" seguida de una o más variables con sus respectivos valores mínimo y máximo entre paréntesis y separados por coma. Se puede dar un mismo rango para un conjunto de variables, colocándolo simplemente al final de la lista de variables. Para finalizar se coloca un slash ("/").

En modo entero las variables independientes deben ser enteras y tener un rango claramente especificado, pero las dependientes usualmente son continuas por lo que BREAKDOWN permite usar las palabras LOWEST y HIGHEST (o sus abreviaciones LO y HI) en la definición de sus rangos.

Para el modo general se coloca la palabra "TABLES=" en la columna 16 de la tarjeta BREAKDOWN y a continuación la especificación de las tablas deseadas haciendo referencia a cualquier variable previamente leída o generada. Para el modo entero TABLES va a continuación de VARIABLES y sólo puede hacer referencia a las variables definidas en VARIABLES. Para definir una tabla se colocan los nombres de las variables separadas por la palabra BY. La primera variable mencionada será siempre la variable dependiente o de criterio. Las variables independientes irán entrando en la tabla en el orden en que se definan procediendo de izquierda a derecha. Los valores de la última variable serán los que cambien más rápido. Hay que notar que el orden de entrada de las variables a BREAKDOWN es en orden contrario al definido en CROSSTABS en donde los valores de la primera variable cambian más rápido. Si las especificaciones de las tablas no alcanzan en una tarjeta, se puede continuar en las siguientes ocupando únicamente de los campos 16 a 72 inclusive. Las tablas deben separarse una





Las opciones disponibles para esta subrutina son:

- 1- Considera válidos todos los datos a entrar en las tablas BREAKDOWN, incluso los definidos como missing.
- 2- Excluye los valores que solamente son missing para la variable dependiente p de criterio.
- 3- Suprime la impresión de las etiquetas tanto de las variables como de los valores.
- 4- Ordena que las tablas sean impresas en con formato de diagrama de árbol (disponible sólo para modo general).

Si CROSSTABS no lleva asociada la tarjeta OPTIONS, un caso se omitirá de los cálculos de alguna tabla si alguno de los valores de alguna variable de esa tabla es missing, las etiquetas serán impresas y las tablas se imprimirán en formato estándar.

#### LIMITACIONES:

- \* Hasta 30 listas separadas por slash ("/") puede contener una instrucción TABLES.
- \* El máximo número de dimensiones para BREAKDOWN es seis.
- \* Hasta 200 etiquetas de valores pueden imprimirse por variable, si fueran más, simplemente no se imprimen.
- \* Hasta 200 variables pueden nombrarse o estar implícitas, en modo general, en la tarjeta TABLES.
- \* Hasta 250 tablas pueden solicitarse en modo general.
- \* Hasta 100 variables se pueden nombrar o estar implícitas en modo entero en la instrucción VARIABLES y por lo tanto en TABLES.
- \* Hasta 100 tablas pueden solicitarse en modo entero.

COMMENT EJEMPLO DE RESULTADOS CON FORMATO NORMAL  
Y TODAS LAS ESTADISTICAS.

VAR LABELS MEN11 CONSIDERA CANAL 11 CANAL CULTURAL/  
CAL11 CALIFICACION CANAL 11 ENTRE 0 Y 10/

VALUE LABELS MEN11 (0)NO (1)SI/  
SEXO (1)HOMBRES (2)MUJERES/

BREAKDOWN VARIABLES=CAL11(0,10) MEN11(0,1)/  
TABLES=CAL11 BY MEN11

STATISTICS ALL

----- DESCRIPTION OF SUBPOPULATIONS -----  
 CRITERION VARIABLE CAL11 CALIFICACION CANAL 11 ENTRE 0 Y 10  
 BROKEN DOWN BY MEN11 CONSIDERA CANAL 11 CANAL CULTURAL  
 -----

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			2930.3880	4.8520	2.2681	5.1443	( 604)
MEN11	0	NO	928.2352	4.3579	2.1881	4.7879	( 213)
MEN11	1	SI	2002.3528	5.1211	2.2683	5.1452	( 391)
TOTAL CASES =			604				

CRITERION VARIABLE CAL11

----- ANALYSIS OF VARIANCE -----

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	SUM OF SQ	N
MEN11	0	NO	928.2352	4.3579	2.1881	1915.0377	( 213)
MEN11	1	SI	2002.3528	5.1211	2.2683	2006.6354	( 391)
WITHIN GROUPS TOTAL			2930.3880	4.8520	2.2404	3021.6751	( 604)

```

*****
*
*           ANALYSIS OF VARIANCE
*
*****
* SOURCE           SUM OF SQUARES  D.F.  MEAN SQUARE      F      SIG.
* BETWEEN GROUPS           88.314      1      88.314      16.008  0.0001
*
* WITH FEWER THAN THREE GROUPS, THE RELATIONSHIP IS LINEAR
* WITHIN GROUPS           3021.675      602      5.019
*           ETA = 0.1609  ETA SQUARED = 0.0239
*
*****

```

COMMENT EJEMPLO CON FORMATO DE DIAGRAMA DE ARBOL.

BREAKDOWN TABLES-CAL11 BY SEXO BY MEN11  
OPTIONS 4

----- DESCRIPTION OF SUBPOPULATIONS -----  
CRITERION VARIABLE CAL11 CALIFICACION CANAL 11 ENTRE 0 Y 10  
BROKEN DOWN BY SEXO  
BY MEN11 CONSIDERA CANAL 11 CANAL CULTURAL  
-----

FOR ENTIRE POPULATION  
SUM 2930.588  
MEAN 4.852  
STD DEV 2.268  
VARIANCE 5.144  
N ( 604)

VARIABLE	SEXO	VARIABLE	MEN11
CODE	1.	CODE	0.
HOMBRES		NO	
SUM	1578.235	SUM	481.176
MEAN	4.994	MEAN	4.583
STD DEV	2.341	STD DEV	2.320
VARIANCE	5.482	VARIANCE	5.381
N	( 316)	N	( 105)

CODE	1.
SI	
SUM	1077.099
MEAN	5.199
STD DEV	2.331
VARIANCE	5.432
N	( 211)

CODE	2.	CODE	0.
MUJERES		NO	
SUM	1352.353	SUM	447.099
MEAN	4.696	MEAN	4.139
STD DEV	2.178	STD DEV	2.039
VARIANCE	4.744	VARIANCE	4.158
N	( 288)	N	( 108)

CODE	1.
SI	
SUM	905.294
MEAN	5.029
STD DEV	2.196
VARIANCE	4.822
N	( 180)

TOTAL CASES = 604

BREAKDOWN cuenta con la facilidad CROSSBREAK que proporciona las tablas en un formato idéntico al de CROSSTABS reemplazando los porcentajes por renglón, columna y total por las medias, casos válidos, sumas y desviaciones estándar de la variable dependiente. Sólo se reemplaza la palabra CROSSBREAK por TABLES dentro del formato del modo entero. Su principal finalidad es proporcionar una tabla visualmente más clara en la relación de la variable dependiente con las independientes.

1	16
Sintaxis: BREAKDOWN	VARIABLES=<var o lista> (min,max) ... <var o lista> (min,max)/ CROSSBREAK=<var o lista> BY ... BY <var o lista>/ <var o lista> BY ...

Adicionalmente al formato normal de BREAKDOWN, CROSSBREAK cuenta con las siguientes opciones:

- 3- Omite la impresión de la frecuencia de cada celda.
- 6- Omite la impresión de las sumas.
- 7- Omite la impresión de la desviación estándar.
- 8- Omite la impresión de las etiquetas de los valores, imprimiendo las etiquetas de las variables.

Sólo la opción 4 de Breakdown no está disponible ya que está en clara contradicción con la facilidad CROSSBREAK.

Las estadísticas de BREAKDOWN no están disponibles para CROSSBREAK pero en cambio se tienen las siguientes diez:

- 3- Ji-cuadrada. Para tablas de 2x2 se aplicará la corrección de Yates y, con menos de 21 casos, se aplicará la prueba Fisher.
- 4- Phi para tablas de 2x2 y Cramers'V para tablas más grandes.
- 5- Coeficiente de contingencia.
- 6- Lambda simétrica y asimétrica.

- 7- Coeficiente de incertidumbre simétrico y asimétrico.
- 8- Tau B de Kendall.
- 9- Tau C de Kendall.
- 10- Gamma parcial y de orden cero para tablas de 3xn.
- 11- D de Somers simétrica y asimétrica.
- 12- Eta (disponible sólo para variables numéricas).

que son las mismas de CROSSTABS con la numeración recorrida en dos. Es importante hacer notar que las estadísticas anteriores sólo miden el grado de relación entre las variables.

COMMENT EJEMPLO DE RESULTADOS CON FORMATO CROSSBREAK  
 BREAKDOWN VARIABLES=CAL11(0,10) SEXO(1,2) MEN11(0,1)/  
 CROSSBREAK=CAL11 BY SEXO BY MEN11

\*\*\*\*\* CROSS-BREAKDOWN OF \*\*\*\*\*  
 SEXO BY MEN11 CONSIDERA CANAL 11 CANAL CULTURAL  
 \*\*\*\*\*  
 VARIABLE AVERAGED... CAL11 CALIFICACION CANAL 11 ENTRE 0 Y 10  
 \*\*\*\*\* PAGE 1 OF 1 \*\*\*\*\*

		MEN11			
		MEN 1		NON	
		COUNT	NO	SI	TOTAL
		SUM			
		STD DEV	0	1	
SEXO					
	1	4.38	5.20	4.99	
HOMBRES		105	211	316	
		481.18	1097.06	1578.24	
		2.32	2.33	2.34	
	2	4.14	5.03	4.70	
MUJERES		108	180	288	
		447.06	905.29	1352.35	
		2.04	2.20	2.18	
COLUMN TOTAL		4.36	5.12	4.85	
		213	391	604	
		928.24	2002.35	2930.59	
		2.19	2.27	2.27	

## 13.6)

## T-TEST

```

1
Syntax: T-TEST      16
                   GROUPS= <especificación de grupo>/
                   VARIABLES= <variable o lista de var>

T-TEST      PAIRS=<var o lista> WITH <var o lista>/

```

T-TEST lleva a cabo la prueba de la significancia de la diferencia entre medias; pueden realizar dos tipos de pruebas:

- a) Para muestras independientes. Los casos están clasificados en dos subgrupos independientes de la población y la prueba es ejecutada para variables específicas.
- b) Para muestras apareadas. Se realiza una prueba-t correlacionada entre dos variables, cada una medida en el mismo conjunto de datos.

Para pruebas entre grupos (muestras independientes), se debe definir cómo serán formados los grupos y a cuales variables será aplicada la prueba. Se coloca la palabra "GROUPS=" a partir de la columna 16 de la tarjeta T-TEST y a continuación la especificación de grupo definiéndola de alguna de las siguientes formas:

```
GROUPS= <variable> (<valor>)/
```

el grupo 1 estará formado por los casos donde el valor de la 'variable' sea mayor o igual al 'valor'. El grupo 2 lo formarán el resto de los casos,

```
GROUPS= <variable> (<valor 1>,<valor 2>)
```

el grupo 1 lo formarán aquellos casos donde el valor de la 'variable' sea igual al 'valor 1', el segundo grupo aquellos iguales al 'valor 2'. El resto de los casos serán ignorados. El caso especial donde 'valor 1'=1 y 'valor 2'=2 puede especificarse como GROUPS= <variable>/,

```
GROUPS= <n1>,<n2>/
```

los grupos 1 y 2 serán formados tomando los primeros 'n1' casos y 'n2' casos del archivo de datos respectivamente.

Después de GROUPS, sigue la especificación "VARIABLES=" y la o las variables que entrarán a la prueba.

Para pruebas entre variables (muestras apareadas) se coloca la palabra "PAIRS=" en el campo de especificación de la tarjeta T-TEST y a continuación una lista de variables si se desea realizar la prueba entre todas las posibles parejas formadas en las variables de la lista. Opcionalmente, se puede dar otra lista de variables, separadas de la primera por la palabra "WITH" y entonces la prueba se realizará para las parejas formadas entre la primera y la segunda lista de variables.

Ambas pruebas pueden realizarse en una misma instrucción T-TEST, colocando primero las especificaciones para muestras independientes GROUPS y VARIABLES y, al último, la de muestras apareadas PAIRS.

Esta prueba no cuenta con la instrucción STATISTICS.

Las opciones disponibles para T-TEST son:

- 1- Todos los valores entrarán a los cálculos, incluso los definidos como missing.
- 2- Si un caso contiene un valor missing para alguna variable en VARIABLES o PAIRS, será excluido del análisis.
- 3- Suprimirá la impresión de las etiquetas de las variables.

Si la instrucción OPTIONS no es utilizada para pruebas entre grupos, un caso será excluido de los cálculos si contiene un valor missing en la variable, para pruebas entre variables un caso será excluido si contiene un valor missing para alguna de las dos variables. Las etiquetas de las variables serán impresas.

LIMITACION: Sólo puede haber una especificación GROUPS, VARIABLES y PAIRS, debiendo aparecer PAIRS al último.

En el caso de grupos al azar, el SPSS, a partir de un análisis de varianza de varianzas, estima un coeficiente T para cuando no existen diferencias significativas entre las varianzas de los grupos (supuesto de homoscedasticidad) POOLED VARIANCE ESTIMATED y un coeficiente para cuando si existen diferencias significativas entre las varianzas (supuesto de heteroscedasticidad) SEPARATE VARIANCE ESTIMATED. El investigador debe elegir aquella prueba que indique el análisis de varianza (F value).



COMMENT EJEMPLO DE RESULTADOS TRABAJANDO CON 3 ESCALAS QUE MIDEN EL CONCEPTO CULTURA ASOCIÁNDOLO CON OTROS CONCEPTOS, LAS ESCALAS TIENEN UN RANGO DE 1 A 5.

VAR LABELS ESCAL1 CALIF CULTURA SEGUN CONDICIONADO/  
 ESCAL2 CALIF CULTURA SEGUN PRODUCTIVIDAD/  
 ESCAL3 CALIF CULTURA SEGUN CIVISMO/  
 VAR022 SEXO DEL ENTREVISTADO/  
 VALLE LABELS VAR022 (1)MASCULINO (2)FEMENINO/  
 T-TEST GROUPS=VAR022/ VARIABLES=ESCAL1 ESCAL2

T-TEST

GROUP 1 - VAR022 EQ	1.	* POOLED VARIANCE ESTIMATE * SEPARATE VARIANCE ESTIMATE										
GROUP 2 - VAR022 EQ	2.											
VARIABLE	NUMBER OF CASES	MEAN	STANDARD DEVIATION	STANDARD ERROR	F VALUE	2-TAIL PROB.	T VALUE	DEGREES OF FREEDOM	T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.	
ESCAL1 CALIF CULTURA SEGUN CONDICIONADO												
GROUP 1	316	3.7097	0.726	0.041	1.14	0.291	-0.84	602	0.404	-0.83	596.94	0.405
GROUP 2	288	3.7608	0.777	0.046								
ESCAL2 CALIF CULTURA SEGUN PRODUCTIVIDAD												
GROUP 1	316	2.3525	1.069	0.061	1.15	0.218	-0.76	602	0.445	-0.76	596.33	0.445
GROUP 2	288	2.4227	1.169	0.069								

T-TEST PAIRS=ESCAL1 ESCAL2 WITH ESCAL3

T-TEST

VARIABLE	NUMBER OF CASES	MEAN	STANDARD DEVIATION	STANDARD ERROR	(DIFFERENCE) MEAN	STANDARD DEVIATION	STANDARD ERROR	2-TAIL CORR. PROB.	T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.	
ESCAL1 CALIF CULTURA SEGUN CONDICIONADO												
	604	3.7340	0.751	0.031	0.2683	0.772	0.031	0.595	0.000	8.54	603	0.000
		3.4657	0.929	0.038								
ESCAL3 CALIF CULTURA SEGUN CIVISMO												
ESCAL2 CALIF CULTURA SEGUN PRODUCTIVIDAD												
	604	2.3860	1.127	0.046	-1.0797	1.075	0.044	0.467	0.000	-21.67	603	0.000
		3.4657	0.929	0.038								
ESCAL3 CALIF CULTURA SEGUN CIVISMO												



Las opciones disponibles para esta subrutina son:

- 1- Ordena que el subprograma incluya todos los casos en el cálculo de los coeficientes de correlación no importando que algunos valores se hayan definido missing.
- 2- Los casos con valores missing serán eliminados LISTWISE, en donde, si un caso contiene un valor missing para alguna variable en la lista, ese caso será omitido de los cálculos de todos los coeficientes. Opción particularmente importante cuando el investigador desea realizar análisis multivariados en las correlaciones y requiere que los coeficientes sean obtenidos de los mismos casos.

El default o manejo normal de los valores missing es eliminación PAIRWISE, mediante la cual, un caso se omite de los cálculos de un coeficiente si el valor de cualquiera de las dos variables involucradas es missing. La eliminación PAIRWISE tiene la ventaja de utilizar todos los datos posibles en el cálculo de cada coeficiente y tiene la desventaja de (bajo algunas circunstancias) producir coeficientes que se basan en un número diferente de casos y quizás, en subpoblaciones completamente diferentes.

- 3- Ordena la aplicación de una prueba de dos colas de significancia estadística a cada coeficiente solicitado. Una prueba de dos colas es generalmente usada cuando el investigador no tiene una hipótesis explícita acerca del comportamiento esperado del coeficiente, es decir, positivo o negativo.
- 4- Produce la impresión de una matriz de coeficientes, para todas las listas especificadas en forma matricial, en el archivo que el usuario seleccione. La palabra WITH nunca debe ser usada en las listas para las cuales se desea la impresión de la matriz. Todas las matrices serán producidas en un formato SF10.7, cada renglón de la matriz inicia en una nueva tarjeta y el renglón continúa por tantas tarjetas como se requiera. Ver la instrucción RAW OUTPUT UNIT.
- 5- Los coeficientes de correlación serán impresos en formato matricial, pero se omitirá la impresión del número de casos y la significancia para cada coeficiente, imprimiendo así más coeficientes por página.
- 6- Sólo los coeficientes no redundantes junto con su número asociado de casos y significancia serán impresos, es decir, sólo aquellos elementos del triángulo superior de la matriz. Esta opción no afecta la opción 4.

Si la instrucción OPTIONS se omite, las opciones default son las siguientes: los valores missing serán excluidos PAIRWISE, para todos los coeficientes solicitados será calculada una prueba de significancia estadística de una cola, no se imprimirán matrices en archivos del sistema, se imprimirá el número de casos y la significancia para cada coeficiente de correlación y los coeficientes serán impresos en forma matricial.

Las estadísticas disponibles para PEARSON CORR son:

- 1- Será calculada e impresa la media y desviación estándar de cada variable referenciada en PEARSON CORR.
- 2- Se imprimirá la multiplicación de las desviaciones de las puntuaciones con respecto a la media y la covarianza para cada pareja de variables para las cuales fue solicitado el coeficiente de correlación.

Si no se asocia la instrucción STATISTICS, sólo los coeficientes, el número de casos y su nivel de significancia serán reportados (a menos que la opción 5 sea seleccionada).

```

RUN NAME      EJEMPLO DE RESULTADOS DE PEARSON CORR
VARIABLE LIST EDOCVIL HIJOS EDAD ESCO
INPUT FORMAT  FIXED (4(F1.0,X))
INPUT MEDIUM DISK
N OF CASES   UNKNOWN
VALUE LABELS  ESCO (1)SIN INSTRUCCION (2)PRIM INCOMP
              (3)PRIM COMP (4)SECUND O MAS/
              EDAD (1)15-19 (2)20-24 (3)25-29 (4)30-34
              (5)35-39 (6)40-44 (7)45 O MAS/
              EDOCVIL (1)ACT UNIDO (0)NO UNIDO/
              HIJOS (1) SI (0)NO/

```

```

PEARSON CORR EDOCVIL HIJOS EDAD ESCO
READ INPUT DATA

```

----- PEARSON CORRELATION COEFFICIENTS -----

	EDOCVIL	HIJOS	EDAD	ESCO
EDOCVIL	1.0000 ( 0) P=****	0.8319 ( 2813) P=0.000	0.4075 ( 2813) P=0.000	-0.3054 ( 2813) P=0.000
HIJOS	0.8319 ( 2813) P=0.000	1.0000 ( 0) P=****	0.4927 ( 2813) P=0.000	-0.3478 ( 2813) P=0.000
EDAD	0.4075 ( 2813) P=0.000	0.4927 ( 2813) P=0.000	1.0000 ( 0) P=****	-0.4206 ( 2813) P=0.000
ESCO	-0.3054 ( 2813) P=0.000	-0.3478 ( 2813) P=0.000	-0.4206 ( 2813) P=0.000	1.0000 ( 0) P=****

(COEFFICIENT / (CASES) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

```

PEARSON CORR EDOCVIL HIJOS EDAD ESCO
OPTIONS      6

```

----- PEARSON CORRELATION COEFFICIENTS -----

VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR		
EDOCVIL	0.8319	EDOCVIL	0.4075	EDOCVIL	-0.3054	HIJOS	0.4927	HIJOS	-0.3478	EDAD	-0.4206
WITH N( 2813)		WITH N( 2813)		WITH N( 2813)		WITH N( 2813)		WITH N( 2813)		WITH N( 2813)	
HIJOS	SIG .000	EDAD	SIG .000	ESCO	SIG .000	EDAD	SIG .000	ESCO	SIG .000	ESCO	SIG .000

A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

### 13.8) NONPAR CORR.

Sintaxis:           1                           16  
NONPAR CORR       <lista de var> WITH <lista de var>/

El subprograma NONPAR CORR obtiene coeficientes no paramétricos de correlación de Kendall y/o Spearman. Ambos coeficientes requieren que las variables sean al menos de escala ordinal y numérica (al ser no paramétricos no dependen de una distribución normal ni de escalas de intervalo), además del uso de rangos sobre los valores absolutos de las variables. Debido a las semejanzas entre las subrutinas PEARSON CORR y NONPAR CORR, se recomienda leer el capítulo referente a PEARSON CORR.

Las principales diferencias entre los coeficientes de Spearman y Kendall es que los coeficientes de Kendall son algo más explicativos cuando los datos contienen un número grande de rangos empatados. Por otro lado, los coeficientes de Spearman parecen tener mejor aproximación al coeficiente de correlación producto-momento cuando los datos son continuos, es decir, si no se caracterizan por un número grande de empates en cada rango.

En resumen, se recomienda usar Kendall cuando un número suficientemente grande de casos son clasificados en un número relativamente pequeño de categorías y, a Spearman, cuando la relación de casos y categorías es pequeña. Ambos coeficientes varían de +1 a -1 pero, en general, el valor absoluto de Kendall tiende a ser más pequeño que el de Spearman.

Las opciones 1 a 4 disponibles en este subprograma son idénticas a las de PEARSON CORR, sin embargo, el uso de la opción 1 ó 2 para procesamiento de datos missing, mejorará significativamente el tiempo de procesamiento con respecto al default que

elimina los valores missing PAIRWISE.

Las opciones 5 y 6 controlan la selección de Spearman y/o Kendall, si no se especifica ninguna se obtendrán únicamente coeficientes de Spearman. La opción 5 produce sólo correlaciones Kendall y la 6 ambos coeficientes. Si se usan las opciones 4 y 6, simultáneamente, las matrices de correlación de Kendall precederán a las de Spearman.

Los coeficientes de correlación, los niveles de significancia estadística y el número de casos sobre los cuales cada correlación fué calculada siempre se imprimen. Este procedimiento no cuenta con otras estadísticas por lo que no requiere una instrucción STATISTICS asociada.

RUM NAME EJEMPLO DE RESULTADOS DE NONPAR CORR.  
 VARIABLE LIST EDOCCIVIL HIJOS EDAD SEXO  
 INPUT FORMAT FIXED (4(F1.0,X))  
 INPUT MEDIUM DISK  
 N OF CASES UNKNOWN  
 VALUE LABELS SEXO (1)HOMBRES (2)MUJERES/  
 EDAD (1)15-19 (2)20-24 (3)25-29 (4)30-34  
 (5)35-39 (6)40-44 (7)45 0 MAS/  
 EDOCCIVIL (1)NCT UNIDO (0)NO UNIDO/  
 HIJOS (1) SI (0)NO/  
  
 NONPAR CORR EDOCCIVIL HIJOS EDAD SEXO  
 OPTIONS 6  
 READ INPUT DATA

----- KENDALL CORRELATION COEFFICIENTS -----

VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR
EDOCIVIL 0.8319	EDOCIVIL 0.4028	EDOCIVIL -0.2967	HIJOS 0.4739	HIJOS -0.3317	EDAD -0.4063
WITH N( 2813)	WITH N( 2813)	WITH N( 2813)	WITH N( 2813)	WITH N( 2813)	WITH N( 2813)
HIJOS SIG .001	EDAD SIG .001	ESCO SIG .001	EDAD SIG .001	ESCO SIG .001	ESCO SIG .001

A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

----- SPEARMAN CORRELATION COEFFICIENTS -----

VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR	VARIABLE PAIR
EDOCIVIL 0.8320	EDOCIVIL 0.4599	EDOCIVIL -0.3147	HIJOS 0.3433	HIJOS -0.3592	EDAD -0.4099
WITH N( 2813)	WITH N( 2813)	WITH N( 2813)	WITH N( 2813)	WITH N( 2813)	WITH N( 2813)
HIJOS SIG .001	EDAD SIG .001	ESCO SIG .001	EDAD SIG .001	ESCO SIG .001	ESCO SIG .001

A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.



### 13.9) SCATTERGRAM.

```
1          16
Sintaxis: SCATTERGRAM <var o lista> (mínimo,máximo)
          LOWEST HIGHEST

          WITH <var o lista> (mínimo,máximo)
          LOWEST HIGHEST
```

SCATTERGRAM imprime una gráfica de dos dimensiones y obtiene, además, estadísticas asociadas con regresión simple. La gráfica requiere 55 líneas por página y no la afecta la instrucción PAGESIZE.

Cada punto graficado es representado por '\*' si sólo un caso cae en esa posición, si caen más de 2 y hasta 9 casos en la misma posición, el número de casos será el impreso, 9 o más casos se representan por el número 9.

Para ejecutar esta subrutina se coloca la palabra SCATTERGRAM a partir de la columna 1 de la tarjeta y los nombres de las variables a graficar a partir de la 16. El campo de especificación puede tener dos diferentes formatos:

- 1- Únicamente una lista de variables. La primera gráfica será formada con la primera variable en la lista como variable renglón y la segunda como variable columna, la siguiente con la primera y la tercera variables, ..., la primera con la última variable de la lista, la siguiente gráfica será formado por la segunda y tercera variables y así hasta terminar con la lista.
- 2- En el segundo formato, la palabra WITH puede proporcionarse para definir las gráficas entre las variables antes y después de la palabra WITH.

Ambos formatos no pueden mezclarse en una sola tarjeta SCATTERGRAM.

Las siguientes instrucciones son completamente similares, generando en ambos casos 6 gráficas.

SCATTERGRAM X1 TO X4

SCATTERGRAM X1 WITH X2 TO X3/ X2 WITH X3 X4/ X3 WITH X4

Si no se especifica lo contrario, la escala es determinada por los valores menor y mayor de cada variable. SCATTERGRAM cuenta con la opción de especificar el rango de valores y así, modificar la escala de cada variable a ser impresa pudiendo mejorar las gráficas al dar el rango deseado. Los casos cuyos valores no estén dentro del rango especificado, serán excluidos de las gráficas y reportados bajo la gráfica.

Para la definición del rango se puede utilizar la palabra LOWEST o HIGHEST o sus abreviaciones, LO o HI respectivamente.

Las opciones disponibles para esta subrutina son:

- 1- Se incluirán los valores missing en las gráficas y en el cálculo de las estadísticas.
- 2- Un caso no será incluido en la gráfica ni en el cálculo de las estadísticas si alguna de las variables en la lista de SCATTERGRAM contiene un valor missing.
- 3- Se suprimirá la impresión de las etiquetas de las variables.
- 4- Normalmente se imprimen líneas paralelas a los ejes vertical y horizontal de la gráfica dividiendo la gráfica en 9 segmentos iguales para mejor visualización de los puntos; mediante esta opción se suprimirán estas líneas.
- 5- Imprimirá además, líneas divisorias diagonales.
- 6- Se realizará una prueba estadística de dos colas si se elige además, la estadística 3.
- 7- Escalamiento automático. Al rango menor se le asignará el valor entero menor encontrado en los datos y al mayor, el entero mayor encontrado. Si se define el rango de una variable y además se especifica la opción 7, el rango se utilizará sobre la opción 7.
- 8- Si no se tiene el espacio suficiente en memoria para el procesamiento de todos los casos del archivo, la gráfica será producida para los primeros 'n' casos, donde n es el número máximo que puede ser procesado.

Si la instrucción OPTIONS no es asociada a SCATTERGRAM las opciones default serán las siguientes: un caso será eliminado de la gráfica si uno de los dos valores en la gráfica es missing; se imprimirán las etiquetas de las variables así como las líneas divisorias rectangulares; se llevará a cabo una prueba de significancia de una cola si es elegida la estadística 3; no se hará un escalamiento automático.

Las estadísticas disponibles para SCATTERGRAM son:

- 1- Coeficiente de correlación de Pearson ('r').
- 2- Coeficiente de determinación ('r' al cuadrado).
- 3- Prueba de significancia de 'r'.
- 4- Error estándar del estimador.
- 5- Ordenada al origen.
- 6- Pendiente.

Quando algún caso es excluido de la gráfica, también lo es del cálculo de las estadísticas.

```

RUN NAME      EJEMPLO DE GRAFICA PRODUCIDA POR SCATTERGRAM
VARIABLE LIST EDADPA EDADMA
INPUT FORMAT  FIXED (4X,F2.0,7I,F2.0)

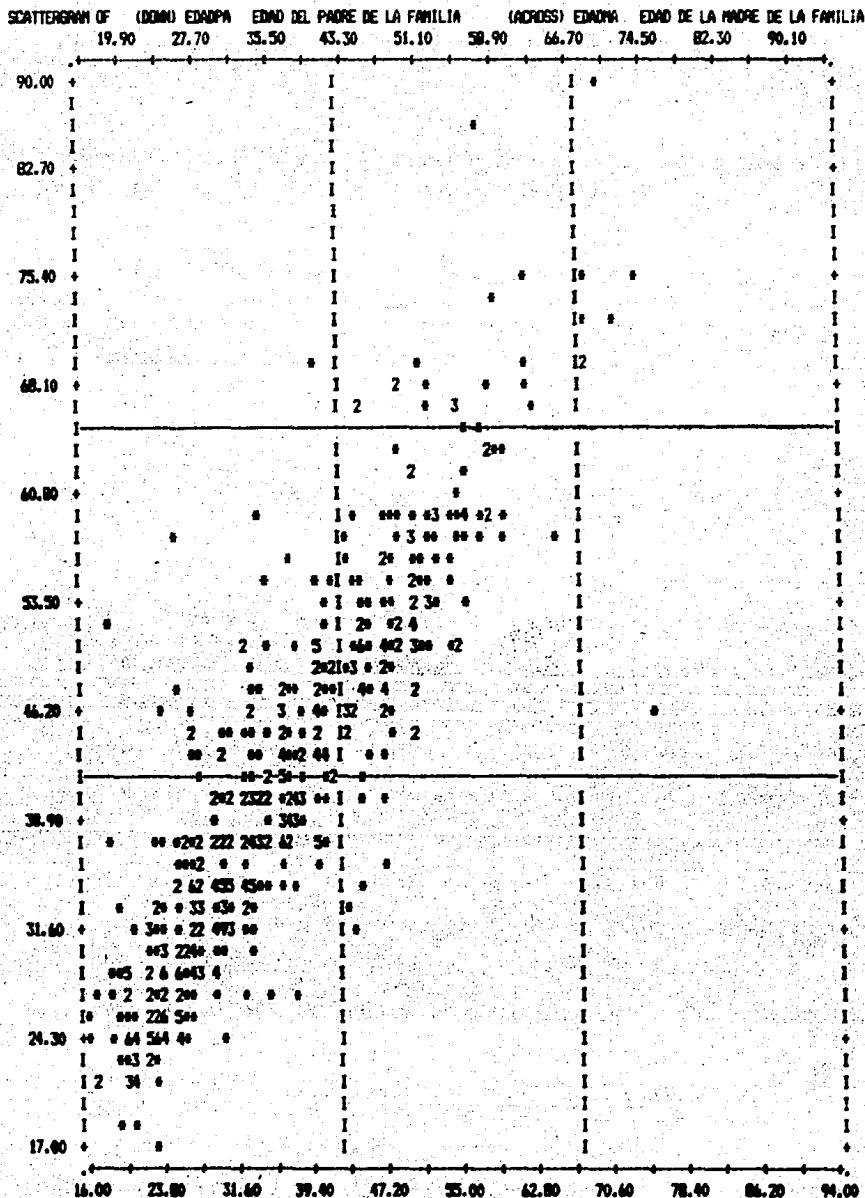
INPUT MEDIUM DISK
N OF CASES    UNKNOWN
VAR LABELS    EDADPA EDAD DEL PADRE DE LA FAMILIA/
              EDADMA EDAD DE LA MADRE DE LA FAMILIA
RECORD        EDADPA EDADMA (BLANK=9999)
MISSING VALUES EDADPA EDADMA (9999)

SCATTERGRAM   EDADPA WITH EDADMA
STATISTICS    ALL
  
```

STATISTICS..

CORRELATION (R)-	0.88659	R SQUARED	-	0.78604	SIGNIFICANCE	-	0.00000
STD ERR OF EST -	5.79087	INTERCEPT (A) -		6.67582	SLOPE (B)	-	0.95251
PLOTTED VALUES -	582	EXCLUDED VALUES-		0	MISSING VALUES -		75

'\*\*\*\*\*' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.



## 13.10)

## PARTIAL CORR.

Syntax:           1                           16  
 PARTIAL CORR   <lista de var> WITH <lista de var>  
   BY <lista de var> (<val de orden>)/

La correlación parcial proporciona al investigador una medida simple de asociación que describe la relación entre dos variables considerando los efectos de una o más variables adicionales.

La correlación parcial permite al investigador eliminar el efecto de la variable control de la relación entre las variables dependientes e independientes, sin un manejo físico de los datos. En la correlación parcial el efecto de la o las variables control supone ser lineal; y es esta suposición lo que hace posible la correlación parcial.

La correlación parcial es ampliamente usada para auxiliar al investigador en la comprensión y claridad de relaciones entre 3 o más variables. Cuando se emplea apropiadamente, llega a ser una excelente técnica para localizar relaciones aparentes (espúreas), variables que parecían no tener relación alguna y hasta para hacer cierto tipo de inferencias causales.

Para ejecutar la subrutina se coloca PARTIAL CORR a partir de la columna 1 de una tarjeta y a partir de la 16 los siguientes datos: una o más parejas de variables independientes y dependientes para las cuales los coeficientes serán calculados (LISTA DE CORRELACION), una o más variables control (LISTA CONTROL) y por último los VALORES DE ORDEN que indican el orden de parcialidad deseado de las correlaciones y de la lista control.

La LISTA DE CORRELACION que especifica las parejas de variables a ser correlacionadas y controladas por las variables en la

LISTA CONTROL, se dan de dos formas dependiendo si el usuario desea los coeficientes para todos los posibles combinaciones de variables en la lista o sólo para combinaciones específicas.

Como en PEARSON CORR, combinaciones específicas de variables se definen con la palabra WITH. Así, se calculará al menos un coeficiente (dependiendo de los VALORES DE ORDEN y de la LISTA CONTROL) para cada variable a la izquierda de WITH pareada con cada variable a su derecha.

Si el usuario desea uno o más coeficientes para todas las posibles combinaciones de variables en la lista de correlación, se omite la palabra WITH, calculando así la matriz cuadrada y completa que será impresa para cada variable control y/o combinaciones de ellas. Sin embargo, sólo serán calculados los coeficientes no redundantes.

La última variable en la LISTA DE CORRELACION es seguida por la palabra BY y a continuación, la lista de variables a ser usadas como control para cada pareja de variables especificadas en la LISTA DE CORRELACION. Sin embargo, los VALORES DE ORDEN son los que indicarán la forma en que estas variables control serán aplicadas a las parejas de variables.

La última variable en la LISTA CONTROL es seguida de uno hasta 5 valores entre paréntesis y separados por comas. Estos valores denominados VALORES DE ORDEN especifican el orden de parcialidad que será calculado de la lista de correlación y control. Esto significa que cada coeficiente será producido para cada combinación única de variables control para cada pareja de correlaciones especificadas en la lista de correlación.

Ejemplo:

PARTIAL CORR NIVEL WITH SEXO BY EDAD ESCO (1,2)

Calculará dos parciales de orden 1 y una parcial de orden 2. NIVEL y SEXO controlado primero por EDAD y luego por ESCO y NIVEL y SEXO controlado por EDAD y ESCO simultáneamente.

Relación entre los valores de orden y la lista control:

- 1- Cualquier valor de orden puede tomar desde el valor 1 hasta el número de variables control.
- 2- Un conjunto completo de coeficientes serán producidos por cada valor de orden en la lista.
- 3- Si se especifica valor de orden=1, un coef. parcial de primer orden será producido para cada variable control para cada pareja de variables especificada en la lista de correlación.
- 4- Cuando se especifica un valor igual al número de variables control, un sólo coeficiente (de orden n, donde n es el número de variables control) será producido para cada pareja de variables especificado en la lista de correlación usando todas las variables control simultáneamente.
- 5- Cuando es dado un valor entre 1 y n, un coeficiente será calculado para cada combinación única posible de variables control para cada pareja de variables en la lista de correlación.

Los valores de orden serán seguidos por un slash ('/'), y hasta 25 listas de correlación, de control y valores de orden pueden ir a continuación.

El subprograma PARTIAL CORR, como REGRESSION y FACTOR, permite al usuario dar los datos en la forma usual como casos u observaciones o, alternativamente, en la forma de una matriz de coeficientes de correlación, evitando al subprograma el primero y más costoso paso al generar estos coeficientes.

Dar una matriz es completamente diferente a dar un archivo de casos, por lo que se deben seguir las siguientes reglas:

las correlaciones deben estar en formato matricial estándar de BF10.7; puede residir en tarjetas, disco o cinta; cada renglón de la matriz debe iniciar en un nuevo registro físico; los coeficientes positivos no necesitan ser precedidos por el signo pero los negativos deben serlo por '-'; siempre debe ser proporcionada la matriz cuadrada completa.

Hay dos métodos para especificar la estructura de matrices de entrada para PARTIAL CORR. Con el método default, cada LISTA DE CORRELACION requiere su propia matriz entre todas las variables nombradas tanto en la lista de correlación cuanto en la de control. El orden de las variables en cada matriz está determinado por el orden de las variables en ambas listas.

El siguiente ejemplo muestra la matriz requerida para una tarjeta PARTIAL CORR definida de la siguiente manera:

PARTIAL CORR A, B WITH C, D BY E, F (1,2)

La matriz de correlación requerida tomaría la sig. forma:

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

La siguiente instrucción generaría la matriz anterior:

PEARSON CORR A, B, C, D, E, F

El segundo método para especificar la estructura de la matriz es mediante la opción 6, que indica que sólo una matriz será dada para el procedimiento, la cual deberá incluir todas las variables especificadas en VARIABLE LIST, se hayan o no referenciado en PARTIAL CORR. Así, el usuario puede preparar una matriz y luego usarla en varios procedimientos. Cuando se complica prepa-



rar las matrices de correlación por querer obtener una variedad de coeficientes parciales para un gran número de variables, se recomienda ampliamente producir una sola matriz.

Cualquiera sea el método usado para especificar la estructura y contenido de la matriz de entrada, el usuario debe incluir las siguientes tarjetas:

- 1.- VARIABLE LIST.
- 2.- INPUT MEDIUM que especifica la localización de la matriz.
- 3.- N OF CASES, para indicar el número de casos en los que se basan las correlaciones de las matrices.
- 4.- OPTIONS, con la opción 4 que indica la lectura de una matriz. Si se elige dar una sola, se dará además la opción 6. Las opciones 1, 2 y 5 no deben usarse junto con la 4 y/o 6.
- 5.- READ MATRIX que debe seguir a OPTIONS y STATISTICS para informar que una o más matrices de coeficientes van a continuación. La instrucción READ MATRIX tiene la misma función que READ INPUT DATA. Si hay más de una matriz, las matrices se colocan una detrás de otra en el orden que PARTIAL CORR las solicite y nada debe haber entre ellas.
- 6.- Si se ha producido matrices para un archivo con estructura de subarchivo, se omitirá N OF CASES, y en su lugar se colocará SUBFILE LIST y RUN SUBFILES precediendo a PARTIAL CORR. La matriz para el primer subarchivo a ser procesado aparece primero (inmediatamente después de READ MATRIX si las matrices se encuentran en tarjetas), seguido del segundo, etc.

Las opciones disponibles para esta subrutina son:

- 1.- Todos los datos entrarán en los cálculos de las correlaciones aun y cuando hubieran sido definidos como missing.
- 2.- Exclusión PAIRWISE de valores missing. Un caso será omitido de los cálculos para un coeficiente dado si el valor para cualquiera de las dos variables consideradas es missing.

La exclusión normal que PARTIAL CORR da a los valores missing es LISTWISE, la cual ordena la omisión de un caso en todos los cálculos si contiene un valor missing para cualquier variable, asegurando así que las correlaciones parciales serán obtenidas de la misma muestra de la población.

- 3.- Cada coeficiente de correlación parcial será acompañado por una prueba de significancia estadística de dos colas.

- 4.- Informa al subprograma PARTIAL CORR que una matriz de correlación será leída en vez de un archivo de datos.
- 5.- Será impresa la matriz de correlación para todas las variables en el análisis, en el dispositivo que el usuario indique, con formato de 8F10.7. Cuando se selecciona esta opción y se especifica más de un conjunto de listas de correlación y control, se imprimirá una matriz para cada conjunto en PARTIAL CORR. Ver instrucción RAW OUTPUT UNIT.
- 6.- El número y orden de las variables en la matriz de entrada será definido por el número y orden de las variables en VARIABLE LIST. Debe usarse junto con la opción 4.
- 7.- Sólo los coeficientes de correlación parcial serán impresos, se omitirán los grados de libertad y significancia imprimiendo así más coeficientes por página.
- 8.- Sólo los coeficientes parciales no-redundantes serán impresos, junto con sus grados de libertad y significancia.

Si la instrucción OPTIONS no se utiliza, los missings se excluirán LISTWISE, se obtendrá una prueba de significancia estadística de una cola para cada parcial, aceptará como entrada un archivo de datos, las matrices de correlación no se imprimirán, los grados de libertad y significancia serán impresos para cada coeficiente los que serán impresos en formato matricial.

Las estadísticas disponibles para PARTIAL CORR son

- 1.- Se imprimirán las correlaciones simples usadas en los cálculos de las correlaciones parciales cada una con sus grados de libertad y significancia estadística.
- 2.- Las medias y desviaciones estándar de las variables en PARTIAL CORR se imprimirán junto con el número de casos válidos para cada variable. No está disponible si se dio de entrada una matriz de correlación.
- 3.- Imprimirá los coef. de correlación de orden cero sólo si no son calculables. Si se seleccionan las estadísticas 1 y 3 simultáneamente, la 1 tendrá precedencia.

Los coeficientes de correlación parcial, grados de libertad y niveles de significancia estadística son siempre reportados.

RUN NAME EJEMPLO DE RESULTADOS DE PARTIAL CORR.

VARIABLE LIST REGION ESDO HIJOS EDREAL  
 INPUT FORMAT FIXED ( 3F1.0,F2.0 )  
 INPUT MEDIUM DISK  
 N OF CASES UNKNOWN  
 VAR LABELS EDREAL EDAD REAL DEL ENTREVISTADO  
 VALUE LABELS REGION (0)VALLE DE MEXICO (1)RURAL/  
 ESDO (1)SIN INSTRUCCION (2)PRIMARIA INCOMP  
 (3)PRIMARIA COMP (4)SECUNDARIA O MAS/  
 HIJOS (0) NO (1) SI

COMMENT \*\*\* ES IMPORTANTE PEDIR LA ESTADISTICA 1,\*\*\*  
 \*\*\* PARA PODER VALORAR LA PARCIALIZACION \*\*\*

PARTIAL CORR ESDO HIJOS BY REGION EDREAL (1,2)  
 STATISTICS 1

-----PARTIAL CORRELATION COEFFICIENTS-----

ZERO ORDER PARTIALS

	ESDO	HIJOS	REGION	EDREAL
ESDO	1.0000 ( 0 ) P=*****	-0.3313 ( 1833 ) P=0.000	-0.2015 ( 1833 ) P=0.000	-0.4391 ( 1833 ) P=0.000
HIJOS	-0.3313 ( 1833 ) P=0.000	1.0000 ( 0 ) P=*****	0.0125 ( 1833 ) P=0.297	0.4749 ( 1833 ) P=0.000
REGION	-0.2015 ( 1833 ) P=0.000	0.0125 ( 1833 ) P=0.297	1.0000 ( 0 ) P=*****	0.0692 ( 1833 ) P=0.002
EDREAL	-0.4391 ( 1833 ) P=0.000	0.4749 ( 1833 ) P=0.000	0.0692 ( 1833 ) P=0.002	1.0000 ( 0 ) P=*****

(COEFFICIENT / (D.F.) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

-----PARTIAL CORRELATION COEFFICIENTS-----

CONTROLLING FOR.. REGION

	ESDO	HIJOS
ESDO	1.0000 ( 0 ) P=*****	-0.3357 ( 1832 ) P=0.000
HIJOS	-0.3357 ( 1832 ) P=0.000	1.0000 ( 0 ) P=*****

(COEFFICIENT / (D.F.) / SIGNIFICANCE) (A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

-----PARTIAL CORRELATION COEFFICIENTS-----

CONTROLLING FOR..	EDREAL	
	ESCO	HIJOS
ESCO	1.0000 ( 0) P=*****	-0.1553 ( 1832) P=0.000
HIJOS	-0.1553 ( 1832) P=0.000	1.0000 ( 0) P=*****

(COEFFICIENT / (D.F.) / SIGNIFICANCE)

(A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)

-----PARTIAL CORRELATION COEFFICIENTS-----

CONTROLLING FOR..	REGION	EDREAL
	ESCO	HIJOS
ESCO	1.0000 ( 0) P=*****	-0.1628 ( 1831) P=0.000
HIJOS	-0.1628 ( 1831) P=0.000	1.0000 ( 0) P=*****

(COEFFICIENT / (D.F.) / SIGNIFICANCE)

(A VALUE OF 99.0000 IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED)



```

1
Sintaxis: REGRESSION
16
          VARIABLES= <lista de variables>/
          REGRESSION= <var.dep.> (parámetros)
          WITH <lista var.indep.>
          (<nivel inclusión>) RESID= <0 ó prv>/
          REGRESSION= .../ VARIABLES= .../

```

El subprograma REGRESSION, a través de la regresión múltiple estándar y de la capacidad de transformación de variables de SPSS, permite realizar análisis multivariados, regresiones polinómicas, regresiones DUMMY y análisis de varianza y covarianza. Además permite examinar residuales y valores Y'.

La entrada para REGRESSION puede consistir de un archivo de datos o de una matriz de correlación. Si se utiliza la segunda opción, será un ahorro considerable de tiempo de proceso.

La regresión múltiple es una técnica estadística que permite analizar la relación entre una variable dependiente o de criterio y un conjunto de variables independientes. En general, requiere que las variables sean intervalares y la relación entre ellas sea lineal y aditiva. Puede verse como un instrumento descriptivo por el cual la dependencia lineal de una variable con otras es resumida y analizada, o como un instrumento inferencial por el cual las relaciones en la población son evaluadas de datos de la muestra, tratando de responder a la pregunta de inferencia estadística, si se puede generalizar los datos de una muestra observada, al universo.

Los usos más importantes de ésta técnica como instrumento descriptivo son: (1) encontrar la mejor ecuación lineal y evaluar su exactitud en la predicción, (2) controlar factores confusos para evaluar la contribución de una o un conjunto de variables y

(3) encontrar relaciones estructurales y proporcionar explicación a relaciones multivariadas aparentemente complejas.

La subrutina se ejecuta colocando REGRESSION en las columnas 1 a 15 de una tarjeta, una lista VARIABLES y al menos una instrucción REGRESSION en el campo de especificación de la tarjeta. VARIABLES nombra las variables que entrarán al análisis de regresión, REGRESSION especifica la variable dependiente, las independientes y el modo de análisis a ser ejecutado. Las variables nombradas en REGRESSION deben estarlo en VARIABLES, sin embargo, puede haber más variables en VARIABLES si hay más de una instrucción REGRESSION.

Cuando SPSS encuentra una lista VARIABLES, construye una matriz de coeficientes de correlación con las variables nombradas. Al encontrar REGRESSION, realiza el análisis seleccionando los coeficientes apropiados, si es VARIABLES, construye una matriz de coeficientes de correlación nueva.

La sintaxis de la instrucción REGRESSION toma la siguiente forma: la variable dependiente, la palabra WITH, las variables independientes, un número de inclusión entre 1 y 99, entre paréntesis y, opcionalmente, si serán calculados los residuales.

Si el número de inclusión es PAR, indica una regresión múltiple ordinaria, si es NON, una inclusión STEPWISE en la que entrarán las variables a la ecuación de regresión de la mejor a la peor según un criterio estadístico establecido.

En una regresión múltiple ordinaria, la solución de regresión es ejecutada por pasos (las variables independientes entran en la ecuación en un orden predeterminado); es útil cuando existe un ordenamiento causal entre las variables, cuando el usuario

desea descomponer la varianza explicada en componentes ortogonales STEPDOWN y para verificar multicolinealidad entre las var. independientes. Cuando se solicitan tales soluciones, el usuario puede especificar el orden de inclusión dando diferentes números para cada variable en forma individual o conjunto de ellas.

Ejemplo:

REGRESSION= A WITH B TO D(8) E(2) F 6(4)

Se resolverán tres ecuaciones de regresión. En el primer paso será A con las variables B a D, en el segundo A con B a D, F y G y, como tercer paso, incluirá la variable E.

Para algunos problemas de investigación es apropiado dar las variables independientes una por una en la base de algún criterio estadístico pre-establecido. Hay varios procedimientos:

- (1) Inclusión FORWARD (STEPWISE). Las variables independientes entran a la ecuación dependiendo si su contribución a la  $r$  cuadrada es significativa. El orden de inclusión es determinado por la contribución en la explicación de la varianza.
- (2) Eliminación BACKWARD. Las var. independientes son eliminadas una por una de la ecuación que inicialmente incluye a todas.
- (3) Solución STEPWISE. La inclusión FORWARD se combina con la eliminación BACKWARD, si ya no cumplen el criterio pre-establecido en cada paso.
- (4) Solución COMBINATORIAL. Se examinan todas las combinaciones.

La subrutina REGRESSION cuenta con la primera opción, inclusión FORWARD (STEPWISE), la que se ejecuta al dar un número de nivel de inclusión NON. En este caso, las variables entrarán a la ecuación de la mejor a la peor según el criterio estadístico establecido en la sección de parámetros. La variable que explique la mayor cantidad de varianza de la variable dependiente entrará primero, la que en combinación con la primera explique la mayor cantidad de varianza entrará en segundo lugar, etc.



La inclusión STEPWISE puede ejecutarse en combinación con una jerarquía preestablecida, por ejemplo:

**REGRESSION= A WITH B TO D(5) E TO X(1)**

hace que las variables B a D entren de una en una si cumplen el criterio STEPWISE y, como segundo paso, las variables E a X. Las variables que tengan un nivel de inclusión mayor que no entraron a la ecuación por no cumplir el criterio, seguirán tomándose en cuenta aunque sean elegibles las de un nivel de inclusión menor.

**REGRESSION= A WITH B TO D(6) E TO G(5)**

Las variables B a D entrarán simultáneamente a la ecuación en el primer paso mientras que las variables E a G entrarán posteriormente según la inclusión STEPWISE.

Se pueden especificar, opcionalmente, 3 criterios estadísticos para decidir la entrada de variables. Se dan en la sección de parámetros de REGRESSION y su forma general es (n,F,T) donde:

- n - Máximo número de variables indep. que entran a la ecuación. Se elegirán las mejores 'n' variables, aunque podrían ser menos si no se satisfacen los demás criterios estadísticos.
- F - En cada paso del análisis se calcula la estadística F, para las variables que no están en la ecuación, relacionando la F calculada con una prueba de significancia. El valor dado de F en la sección de parámetros, es el mínimo valor de F que el usuario aceptará.
- T - Tolerancia. La tolerancia de una variable independiente es la proporción de la varianza no especificada por las que ya se encuentran en la ecuación. Tiene un rango entre 0 y 1. Una tolerancia de 0 indica una combinación lineal perfecta, de 1, que la variable no está correlacionada con las otras.

Cada parámetro tiene asignado por SPSS un valor default que son: n=80, F=0.01 y T=0.001, los que permiten obtener resultados más completos pero no reducen la ecuación a los mejores 'n' predictores. Los parámetros sólo están en efecto cuando se solicita inclusión STEPWISE mediante un número de nivel de inclu-

sión NON. Cuando ambos números, pares y nones, son usados los parámetros sólo se aplicarán a variables con número non.

El SPSS obtiene 'B' que es la contribución de la var. indep. a la dependiente medida en sus unidades naturales de medición y, 'BETA' en sus unidades estandarizadas. Si las unidades de las variables independientes son diferentes, usar 'BETA', si no 'B'.

El subprograma REGRESSION calcula e imprime los residuales y valores estimados de Y cuando el usuario coloca, opcionalmente, RESID=<prvm> al final de la lista de REGRESSION. <prvm> es un número entre 0 y 1 y es la 'PROPORCION DE REEMPLAZO DE LOS VALORES MISSING'. Cuando se elige eliminación de valores missing LISTWISE (default), <prvm> debe ser igual a cero. Esta opción debe darse para cada lista REGRESSION para la cual se deseen residuales.

Ejemplo:

```
REGRESSION= Y WITH X1 TO X7(2) RESID=.40/
```

Se calcularán los residuales y estimados para todo caso en el cual 40% o menos de las var. indep. tengan valores missing.

REGRESSION puede aceptar como entrada medias, desviaciones estándar y coeficientes de correlación de orden cero en lugar del archivo normal de datos. Este tipo de entrada ocasiona un ahorro considerable de tiempo de proceso.

La matriz puede generarse con las subrutinas PEARSON CORR, PARTIAL CORR o incluso REGRESSION. Debe ser cuadrada, simétrica y con unos en la diagonal. Con un formato de BF10.7, iniciando cada renglón de la matriz en una nueva tarjeta. Cuando se procese para subarchivos (mediante RUN SUBFILES) debe darse una matriz para cada subarchivo, en el orden que sean procesados. Similarmente, cuando haya más de una lista VARIABLES, se debe dar una matriz para cada lista, a menos que se indique la opción 9.

Hay dos metodos para especificar la estructura de la matriz (y de medias y desviaciones). En el default, el número y orden de las variables en la matriz debe corresponder al de la lista VARIABLES en REGRESSION. Especificando la opción 9, además de la 4, la estructura de la matriz la definirá el número y orden de las variables en la instrucción VARIABLE LIST. Así, sólo una matriz (y medias y desviaciones asociadas) tendrá que prepararse.

Cuando se indiquen las opciones 4 y 5 simultáneamente, se deben incluir las medias y desviaciones estándar al archivo de datos. Deben colocarse las medias, seguidas por las desviaciones, precediendo siempre a la matriz de correlación. El orden debe ser el mismo que el de las variables en la matriz utilizando un formato de BF10.0.

Si el usuario utiliza las opciones 4 y 5, debe preparar, adicionalmente, las siguientes tarjetas:

1. VARIABLE LIST.
2. INPUT MEDIUM, que define donde se encuentra la matriz.
3. N OF CASES para uso en cálculos de pruebas de significancia.
4. READ MATRIX que debe seguir a las tarjetas STATISTICS u OPTIONS teniendo la misma función que READ INPUT DATA.

Las opciones disponibles para esta subrutina son:

- 1.- Todos los casos serán incluidos en los cálculos aún y cuando hayan sido definidos como missing.
- 2.- Eliminación PAIRWISE de valores missing. Un valor missing para una variable causa que todo el caso sea eliminado de los cálculos que envuelven esa variable. Esta opción debe usarse sólo cuando se tengan pocos valores missing ya que puede originar resultados basados, quizás, en subpoblaciones completamente diferentes.

Por default, se utiliza eliminación LISTWISE. Los casos con valores missing son eliminados de todos los cálculos, obteniendo así, resultados basados en los mismos datos.

- 3.- Se suprimirá la impresión de las etiquetas de las variables.
- 4.- Especifica que el usuario da como entrada una matriz de coeficientes de correlación.
- 5.- Indica que las medias y desviaciones estándar se leerán en forma previa a la matriz de correlación. La opción 5 sólo puede especificarse junto con la opción 4.
- 6.- Se suprimirá la impresión de resultados paso por paso, imprimiendo únicamente la tabla final.
- 7.- Se suprimirá la impresión de la tabla final, imprimiendo únicamente los resultados paso por paso.
- 8.- La matriz o matrices de correlación usadas en los cálculos, serán impresas en un archivo definido por el usuario. Ver instrucción RAW OUTPUT UNIT.
- 9.- Indica que el usuario dará sólo una matriz de correlación, según el orden de las variables en VARIABLE LIST. La opción 9 debe especificarse junto con la opción 4.
- 10.- Imprimirá los residuales y los valores Y', ambos estandarizados, en un archivo definido por el usuario, con el sig. formato: el número de secuencia del caso de las columnas 1 a 8, el número de registro en las columnas 9 y 10, la identificación del subarchivo de las columnas 12 a 15 y 6 residuales o valores Y' de las columnas 21 a 80 con un formato 6F10.6. Ver instrucción RAW OUTPUT UNIT.
- 11 Y 12.- Si se señala únicamente la opción 11, se imprimirán los residuales estandarizados en un archivo definido por el usuario. Si la opción 12, los valores Y' estandarizados. Si se especifican ambas opciones, tanto los residuales como los valores Y' serán impresos con un formato de 8F10.6. Ver instrucción RAW OUTPUT UNIT.
- 13.- Cuando las variables independientes toman valores missing, SPSS estima esos valores (para poder realizar los cálculos) según dos métodos. El método default es reemplazándolo por la media de la variable. El otro método se acciona mediante la opción 13 y produce valores estimados en base a productos ponderados de los datos ya existentes. Esta opción sólo tendrá efecto si se utiliza la opción 2 y RESID con prvm mayor que cero (prvm>0).
- 14.- Suprimirá la impresión de los ejes en las gráficas producidas por la estadística 6.
- 15.- Las medias y desviaciones se imprimirán en un archivo con formato de 8F10.4. Ver instrucción RAW OUTPUT UNIT.

Las estadísticas disponibles para REGRESSION son:

- 1.- Imprimirá la matriz o matrices de correlación para cada lista VARIABLES en el subprograma REGRESSION.
- 2.- Las medias, desviaciones estándar y número de casos válidos serán impresos para cada lista VARIABLES.
- 3.- Fuerza la impresión de la matriz de correlación aun y cuando alguno de sus elementos no pueda ser calculado. Si la estadística 3 se usa sin la 1 ó 7, la matriz se imprimirá sólo cuando uno o más coeficientes sean incalculables. Si se usa junto con la 1 ó 7, la matriz de correlación siempre será impresa.
- 4.- Se imprimirá una gráfica de residuales estandarizados contra la secuencia de los casos en el archivo. Esta gráfica sólo se obtendrá para la última lista REGRESSION que utilice la opción RESID=0. Se listarán además, los valores observados de Y, los valores Y' y los residuales. Se graficarán los primeros 500 casos del archivo. Para que tenga verdadero sentido esta estadística, el archivo debe estar ordenado por alguna variable significativa.
- 5.- Calcula la estadística DURBIN-WATSON para residuales.
- 6.- Obtiene una gráfica de los residuales estandarizados contra los valores estandarizados de Y'. Se imprimirá una gráfica para cada lista REGRESSION que utilice la opción RESID.
- 7.- Imprimirá la matriz de correlación y el número de casos si se especifica la opción 2. Si se utiliza eliminación LISTWISE (default) o la opción 1, se sustituirá por la estadística 1.

```

RUN NAME:      EJEMPLO DE RESULTADOS DE REGRESSION.
VARIABLE LIST  SUELDO SEXO ESCO EDREAL
INPUT FORMAT  FIXED (F7.0,F1.0,F2.0)
INPUT METHOD   DISK
N OF CASES    UNKNOWN

VAR LABELS    EDREAL EDNO REAL
VALUE LABELS  SEXO (0)HOMBRES (1)MUJERES/
              ESCO (1)SIN INSTRUCCION (2)PRIM INCOOP
              (3)PRIM COMP (4)SEGUNDO O MAS/

REGRESSION    VARIABLES= SUELDO SEXO ESCO EDREAL/
              REGRESSION= SUELDO WITH SEXO ESCO EDREAL
              RESID=0/

STATISTICS    4,5,6
READ INPUT DATA
```

\*\*\*\*\* MULTIPLE REGRESSION \*\*\*\*\* VARIABLE LIST 1  
REGRESSION LIST 1

DEPENDENT VARIABLE.. SUELDO  
VARIABLE(S) ENTERED ON STEP NUMBER 1.. ESCO

MULTIPLE R	0.17680	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.03126	REGRESSION	1.	235914886723.60100	235914886723.60100	90.69983
ADJUSTED R SQUARE	0.03091	RESIDUAL		2811.7311534776216.29300	2801051147.71124	
STANDARD ERROR	51000.50145					

----- VARIABLES IN THE EQUATION -----

VARIABLE	B	BETA	STD ERROR B	F
ESCO	8805.498	0.17680	932.99340	90.700
(CONSTANT)	28393.04			

----- VARIABLES NOT IN THE EQUATION -----

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
SEXO	-0.07073	-0.07149	0.98987	14.436
EDREAL	0.06235	0.05633	0.79056	8.944

\*\*\*\*\*  
VARIABLE(S) ENTERED ON STEP NUMBER 2.. SEXO

MULTIPLE R	0.19029	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.03621	REGRESSION	2.	273285472091.71140	136642736045.85570	52.78476
ADJUSTED R SQUARE	0.03552	RESIDUAL		2810.727184190948.18300	258667147.98868	
STANDARD ERROR	50879.04920					

----- VARIABLES IN THE EQUATION -----

VARIABLE	B	BETA	STD ERROR B	F
ESCO	8327.736	0.16968	935.52228	83.092
SEXO	-7296.099	-0.07073	1920.28077	14.434
(CONSTANT)	33051.49			

----- VARIABLES NOT IN THE EQUATION -----

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
EDREAL	0.05938	0.05374	0.78938	9.135

\*\*\*\*\* MULTIPLE REGRESSION \*\*\*\*\* VARIABLE LIST 1  
REGRESSION LIST 1

DEPENDENT VARIABLE.. SUELDO  
VARIABLE(S) ENTERED ON STEP NUMBER 3.. EDREAL

MULTIPLE R	0.19746	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.03899	REGRESSION	3.	294291774079.26030	9809728293.08462	37.99096
ADJUSTED R SQUARE	0.03777	RESIDUAL		2809.725317788860.63400	2382128999.64539	
STANDARD ERROR	50814.57448					

----- VARIABLES IN THE EQUATION -----

VARIABLE	B	BETA	STD ERROR B	F
ESCO	9903.861	0.19706	1051.35294	88.705
SEXO	-7084.133	-0.06867	1919.28643	13.621
EDREAL	302.2250	0.03938	165.96045	8.133
(CONSTANT)	20257.42			

----- VARIABLES NOT IN THE EQUATION -----

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
----------	---------	---------	-----------	---

MAXIMUM STEP REACHED  
STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.

\*\*\*\*\* MULTIPLE REGRESSION \*\*\*\*\* VARIABLE LIST 1  
REGRESSION LIST 1

DEPENDENT VARIABLE.. SUELDO

SUMMARY TABLE

VARIABLE	MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
ESCO	0.17680	0.03126	0.03126	0.17680	9903.861	0.19706
SEXO	0.19029	0.03621	0.00495	-0.08780	-7084.135	-0.06867
EBREAL (CONSTANT)	0.19746	0.03899	0.00278	-0.03162	302.2280	0.05938
					20257.42	

\*\*\*\* NOTE CHANGE IN FORMULA FOR STANDARDIZED RESIDUALS AS OF 17 DEC 79 \*\*\*\*

IT WAS (RESIDUAL/STD. DEV. OF DEP. VARIABLE)

IT IS NOW (RESIDUAL/STD. ERROR OF REGRESSION)

\*\*\*\*\* MULTIPLE REGRESSION \*\*\*\*\*

DEPENDENT VARIABLE: SUELDO FROM VARIABLE LIST 1  
REGRESSION LIST 1

SECNUM	OBSERVED SUELDO	PREDICTED SUELDO	RESIDUAL	PLOT OF STANDARDIZED RESIDUAL				
				-2.0	-1.0	0.0	1.0	2.0
1	32000.00	49231.59	-17231.59			*	1	
2	20000.00	46714.09	-26714.09		*	1		
3	5000.000	40536.63	-35536.63		*	1		
4	50000.00	35770.59	14229.41			*	1	*
5	20000.00	57624.32	-37624.32		*	1		
6	50006.00	65312.91	-15312.91			*	1	
7	50000.00	72566.25	-22566.25		*	1		
8	45000.00	39437.67	5437.68		*	1		
9	25000.00	61855.25	-36855.25		*	1		
10	-90000.00	61755.77	28244.22		*	1	*	
11	16000.00	45674.45	-29674.45		*	1		
12	50000.00	61855.47	-11855.47			*	1	
13	30000.00	44465.35	-14465.35		*	1		
14	80000.00	60344.35	19655.65			*	1	*
15	20000.00	67126.25	-47126.26	*		1		
16	20000.00	56182.76	36182.76	*		1		
17	48000.00	42047.75	5952.242			*	1	*
18	30000.00	60546.87	-30546.88		*	1		
19	50000.00	67126.25	-17126.26		*	1		
20	30000.00	57624.32	-27624.32		*	1		
21	30000.00	39035.75	29035.75		*	1		
22	90000.00	72566.25	17433.69			*	1	*
23	40000.00	50440.49	-10440.49		*	1		
24	24000.00	47720.46	-23720.46	*		1		
25	12000.00	65312.91	-53312.91	*		1		

DURBIN-WATSON TEST OF RESIDUAL DIFFERENCES COMPARED BY CASE ORDER (SECNUM).  
VARIABLE LIST 1, REGRESSION LIST 1. DURBIN-WATSON TEST 1.73589





### 13. 12)

### ANOVA.

Sintaxis:     1                             16  
ANOVA           <lista var.dep.> BY <lista var.indep.>  
                 (min,max) WITH <lista covariables>/  
                 <lista var.dep.> ...

El subprograma ANOVA de SPSS permite al usuario obtener análisis de varianza para diseños factoriales, permitiendo hasta 5 factores para cada diseño. Aunque ANOVA no cuenta con un análisis de covarianza completo, permite ajustar hasta 5 covariables. Adicionalmente, produce una tabla de Análisis de Clasificación Múltiple (STATISTICS 1).

ANOVA puede realizar análisis de varianza de una vía, sin embargo se recomienda al usuario utilizar el subprograma ONEWAY, específicamente diseñado para manejar este tipo de análisis, que cuenta con una amplia variedad de estadísticas adicionales al análisis de varianza, como contrastes a-priori y a-posteriori, siete diferentes pruebas estadísticas y una prueba de tendencia.

ANOVA acepta como entrada archivos normales de datos. No pretende ser un programa de análisis de varianza y covarianza para todo propósito. En particular, no maneja variables dependientes múltiples, interacción entre factores y covariables, ni diseños anidados o no-factoriales.

El análisis simple de varianza y covarianza debe tener una variable dependiente, también llamada de criterio, que debe ser de por lo menos una escala intervalar. Las variables independientes se denominan factores. Si se está interesado en los posibles efectos de un sólo factor, el análisis se denomina Análisis de Varianza de una vía, si en los efectos simultáneos de n-factores, el análisis es de n-vías.

El orden en el que las covariables y factores entran al análisis es controlado por el usuario; ANOVA es muy flexible en este aspecto permitiendo producir resultados según las 3 aproximaciones de análisis: Clásica, Jerárquica y Regresión. Además, puede trabajar los efectos de las covariables antes, con, o después de los principales efectos de los factores.

Para ejecutar la subrutina, se coloca la palabra ANOVA en las columnas 1 a 15 de una tarjeta y a partir de la 16, una lista de variables dependientes, una de independientes y, opcionalmente, una de covariables. Hasta 5 conjuntos de listas pueden aparecer en la misma instrucción ANOVA, separadas por un slash (/).

La lista de variables dependientes puede nombrar hasta 5 variables. A continuación, separada por la palabra BY, la lista de variables independientes nombra hasta 5 variables enteras, debe contener la especificación '(min,max)' indicando los valores menor y mayor de las variables entre paréntesis y separados por una coma, al final o después de cada variable en la lista. Los valores mínimo y máximo pueden no corresponder exactamente a los datos, pero si algún valor excede al mínimo o máximo especificado, el caso será ignorado del análisis. Por otro lado, si el rango especificado es mayor al real, el espacio de memoria para el proceso crece significativamente, por lo tanto, el costo. Mediante RECODE pueden convertirse variables no-enteras en variables enteras adecuadas para usar como factores en ANOVA.

Las opciones disponibles para esta subrutina son:

- 1.- Incluirá todos los casos en los cálculos, aun y cuando se hayan definido como missing.
- 2.- Suprime la impresión de las etiquetas de los valores.

Si se trabaja con muchos factores, la interpretación de las interacciones puede llegar a ser problemático o de poco interés. Mediante las opciones 3 a 6, se permite eliminar los efectos de interacciones de orden-n de un análisis. El default es examinar todos los efectos de interacciones.

- 3.- Instruye al programa ignorar interacciones de 2 o más vías entre los factores, trabajando sólo los efectos principales junto con los de las covariables.
- 4.- Ignora interacciones de 3 o más vías entre los factores, considerando interacciones de hasta 2-vías.
- 5.- Ignora interacciones de 4 o más vías entre los factores, considerando interacciones de hasta 3-vías.
- 6.- Ignora interacciones de 5-vías entre los factores, considerando interacciones de hasta 4-vías.

Las opciones 7 y 8 controlan el orden en que las covariables entrarán al análisis. El default es introducir las covariables antes que los factores. Las opciones 7 y 8 no pueden seleccionarse simultáneamente.

- 7.- Procesa las covariables junto con los efectos principales de los factores.
- 8.- Se procesan las covariables después de los efectos principales de los factores.

Las opciones 9 y 10 determinan la aproximación a ser usada. El default es usar la aproximación CLASICA EXPERIMENTAL en la cual, el efecto principal de cada factor es ajustado por los demás factores y similarmente, el efecto de cada covariable es ajustado por las demás covariables. Las opciones 9 y 10 no pueden seleccionarse simultáneamente.

- 9.- Se usará la aproximación REGRESSION, en donde todos los efectos, incluyendo los efectos principales, de covariables y cualquier efecto de interacción se fijará simultáneamente como una regresión múltiple. Si se usa la opción 9, las opciones 7 y 8 se ignorarán (si se especifican).
- 10.- Se usará la aproximación JERARQUICA, en donde los principales efectos de factores y covariables se fijan jerárquicamente.

MÉTODOS DE ANÁLISIS PRODUCIDOS POR LAS OPCIONES 7 A 10.

	DETERMINACION ENTRE EFECTOS	DETERMINACION EN EL MISMO TIPO DE EFECTO.
Default	1o COVARIABLES 2o FACTORES 3o INTERACCIONES	COVARIABLES- Ajuste para las demás covariables. FACTORES- Ajuste para covariables y los demás factores. INTERACCIONES- Ajuste para covariables, factores y las demás interacciones del mismo orden o menor.
Opción 7	1o FACTORES y COVARIABLES 2o INTERACCIONES	COVARIABLES- Ajuste para factores y las demás covariables. FACTORES- Ajuste para covariables y los demás factores. INTERACCIONES- Ajuste para covariables, factores y las demás interacciones del mismo orden o menor.
Opción 8	1o FACTORES 2o COVARIABLES 3o INTERACCIONES	FACTORES- Ajuste para los demás factores. COVARIABLES- Ajuste para factores y las demás covariables. INTERACCIONES- Ajuste para covariables, factores y las demás interacciones del mismo orden o menor.
Opción 9	COVARIABLES, FACTORES e INTERACCIONES simultáneamente	COVARIABLES- Ajuste para factores, interacciones y las demás covariables. FACTORES- Ajuste para covariables y los demás factores. INTERACCIONES- Ajuste para covariables, factores y demás interacciones.
Opción 10	1o COVARIABLES 2o FACTORES 3o INTERACCIONES	COVARIABLES- Ajuste para las covariables precedentes en la lista. FACTORES- Ajuste para covariables y factores precedentes en la lista. INTERACCIONES- Ajuste para covariables, factores y las demás interacciones del mismo orden o menor.
Opción 7 y 10	1o FACTORES y COVARIABLES 2o INTERACCIONES	FACTORES- Ajuste sólo para factores precedentes. COVARIABLES- Ajuste para factores y covariables precedentes. INTERACCIONES- Ajuste para interacciones, factores y las demás interacciones del mismo orden o menor.
Opción 8 y 10	1o FACTORES 2o COVARIABLES 3o INTERACCIONES	FACTORES- Ajuste sólo para factores precedentes. COVARIABLES- Ajuste para factores y covariables precedentes. INTERACCIONES- Ajuste para interacciones, factores y las demás interacciones del mismo orden o menor.

Las estadísticas disponibles para ANOVA son:

- 1.- Produce una tabla de Análisis de Clasificación Múltiple, que consiste en la gran media de la variable dependiente y una tabla de medias por categoría para cada factor, expresada como desviación de la gran media, así, las medias reflejan la magnitud del efecto de cada categoría de un factor.
- 2.- Imprime para las covariables los coeficientes parciales de regresión no estandarizados. Como estos coeficientes son calculados cuando las covariables entran a la ecuación, sus valores dependerán del tipo de diseño especificado según las opciones 7 a 10 o su default.

Ejemplos:

ANOVA A, B BY C, D (0,4)

Se realizarán 2 análisis de varianza de dos vías, la variable A con C y D, ambos factores de 5 categorías y B con C y D.

Si no se especifican opciones, ANOVA ejecuta un análisis de varianza CLASICO. Esto es, los principales efectos para C serán fijados con D constante y los de D con C constante. El efecto de interacción de CD será entonces fijado con los principales efectos de C y D constantes. Si se especifica la opción 9, se obtendrá una solución tipo REGRESION, en donde cada efecto se considera como una contribución adicional a la explicación de la varianza. Con la opción 10, se ejecutará un análisis JERARQUICO, donde el orden en que se fijan las variables, es determinado por el orden en la lista de variables independientes o covariables.

ANOVA Y BY A, B (0,4) WITH X, Z

Si no se especifican las opciones 7 u 8, las covariables X y Z se procesarán primero, ignorando los factores. Los factores A y B se procesarán después con las covariables constantes. Las interacciones AB se fijarán por último, con ajustes para covariables y factores.

Con la opción 7, las covariables y los principales efectos aditivos se procesan como un solo bloque y, una vez constantes, se fijará la interacción AB. Con la opción 8, los principales efectos para A y B se fijan ignorando el bloque de covariables. Entonces, la suma de cuadrados de X y Z será calculada con ajustes para A y B. La interacción AB será calculada de nuevo después de ajustar A, B, X y Z.

```

RUN NAME      EJEMPLO DE RESULTADOS DE ANOVA.
VARIABLE LIST SEXO ESCO EDADREAL WR262
INPUT FORMAT  FIXED (ZF1.0,F2.0,F1.0)
INPUT MEDIUM  DISK
N OF CASES    UNCOMMON
WR LABELS     WR262 TIENE CONFIANZA EN LA TELEVISION/
VALUE LABELS  WR262 (1)MINEJERA (2)POCA(3)SUFICIENTE(4)MUCHO/
              SEXO (0)MASCULINO (1)FEMENINO/
              ESCO (1)ANALF Y PRIM. INC (2)PRIMARIA (3)SECUNDA-RIA
              (4)PREPARATORIA (5)SUPERIOR/
ANOVA         WR262 BY SEXO(0,1) ESCO(1,5) WITH EDADREAL
STATISTICS    1

```

\*\*\*\*\* ANALYSIS OF VARIANCE \*\*\*\*\*

WR262 TIENE CONFIANZA EN LA TELEVISION

BY SEXO  
ESCO  
WITH EDADREAL

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF
CONTRASTES	0.290	1	0.290	0.403	0.525
EDADREAL	0.290	1	0.290	0.403	0.525
MAIN EFFECTS	94.053	5	18.811	26.161	0.000
SEXO	0.232	1	0.232	0.323	0.570
ESCO	90.075	4	22.519	31.319	0.000
2-WAY INTERACTIONS	2.392	4	0.598	0.832	0.505
SEXO ESCO	2.392	4	0.598	0.832	0.505
EXPLAINED	94.735	10	9.473	13.454	0.000
RESIDUAL	1261.877	1765	0.719		
TOTAL	1358.612	1765	0.770		

\*\*\* MULTIPLE CLASSIFICATION ANALYSIS \*\*\*

WR262 TIENE CONFIANZA EN LA TELEVISION

BY SEXO  
ESCO  
WITH EDADREAL

GRAND MEAN = 2.68

VARIABLE + CATEGORY	N	ADJUSTED FOR INDEPENDENTS			
		UNADJUSTED DEV'N	ETA	BETA	INDEPENDENTS + CONTRASTES DEV'N BETA
SEXO					
0 MASCULINO	898	-0.05			-0.01
1 FEMENINO	868	0.05			0.01
		0.05			0.01
ESCO					
1 ANALF Y PRIM. INC	339	0.33			0.33
2 PRIMARIA	239	0.14			0.15
3 SECUNDA-RIA	544	0.03			0.02
4 PREPARATORIA	392	-0.23			-0.23
5 SUPERIOR	252	-0.34			-0.33
		0.26			0.26
MULTIPLE R SQUARED					0.049
MULTIPLE R					0.264

```

1
Syntax:  ONEWAY      16
                <lista var dep> BY <var indep> (min,max)/
                POLYNOMIAL= n/
                CONTRAST= <lista de coeficientes>/
                RANGES= <nombre de prueba> ( <alfa> )

```

ONEWAY resuelve problemas de análisis de varianza que envuelven sólo una variable independiente. Adicionalmente, proporciona pruebas de : tendencia entre categorías de la variable independiente, contrastes a-priori y a-posteriori. La entrada de esta subrutina puede ser en forma de archivo de datos o vectores de frecuencias por categoría, medias y desviaciones estándar.

Hasta 20 variables pueden aparecer en la lista de variables dependientes, pero sólo una variable independiente puede especificarse, debiendo ser una variable entera. La especificación '(min,max)' indica el valor menor y mayor de la variable independiente que será utilizada en el análisis. Los casos se dividen en grupos según sus valores en la variable independiente. El grupo 'n' se compone por todos los casos con el valor 'n' en la variable. Cuando dos o más variables dependientes son nombradas, una tabla ANOVA se obtendrá para cada variable nombrada.

Si se desea una prueba de tendencia, se coloca la palabra POLYNOMIAL en la tarjeta ONEWAY, que indica a SPSS particionar la suma de cuadrados entre grupos en: lineal, cuadrático, cúbico o un orden mayor de componentes de tendencia. La 'n' especifica el grado máximo polinomial a ser usado, debe ser un entero positivo menor o igual a 5 y menor al número de grupos. Sólo una instrucción POLYNOMIAL puede darse en una tarjeta ONEWAY.

La partición de sumas de cuadrados entre grupos en componentes de tendencia, envuelve una regresión polinomial de medias de grupo en los valores de la variable independiente. Por lo que una prueba de tendencia no tiene sentido a menos que las categorías de la variable independiente tengan algún orden de magnitud.

Si se desea una prueba de contrastes, se coloca la palabra CONTRAST seguida por una lista de coeficientes que el usuario desee asignar a las medias de grupo. Se pueden especificar hasta 10 contrastes.

La notación 'N\*C' se puede usar al especificar coeficientes, donde 'N' es el factor de repetición y 'C' el coeficiente. Los coeficientes cero a la derecha pueden omitirse.

Ejemplo:

```
ONEWAY      ENDO BY NLEVEL (1,5)/  
            CONTRAST= 1 0 -1 /  
            CONTRAST= 1.5 2*0 -.5/
```

El primer CONTRAST especifica una prueba t para la diferencia entre las medias del primer y tercer grupo, el segundo CONTRAST define un contraste entre el grupo 1 y la combinación de los grupos 2 y 5. Para aplicaciones más sofisticadas, el usuario debe asegurarse que cada conjunto de coeficientes sumen cero.

Para obtener una prueba de contrastes a-posteriori, se coloca la palabra RANGES en la tarjeta ONEWAY que permite al usuario elegir entre 7 diferentes pruebas. La opción ALFA es el nivel de significancia deseado. Si no se especifica, el valor default es .05. Hasta 10 pruebas y sus alfas asociadas pueden listarse en una tarjeta ONEWAY.

Está permitido usar las especificaciones POLYNOMIAL, CONTRAST y RANGES en la misma tarjeta ONEWAY, siempre y cuando



POLYNOMIAL se encuentre primero, después todas las instrucciones CONTRAST y finalmente todas las instrucciones RANGES.

NOMBRE	PRUEBA	ALFA PERMITIDA
LSD	Diferencia Mínima Significativa	cualquiera
DUNCAN	Prueba Duncan de Rango Múltiple	.10, .05 y .01
SNK	Student - Newman - Keuls	.05
TUKEYB	Prueba alternativa de Tukey	.05
TUKEY	Diferencia significativamente verdadera	.05
LSDMOD	Prueba LSD Modificada	cualquiera
SCHEFFE	Prueba de Scheffe	cualquiera

Las opciones disponibles para esta subrutina son:

- 1.- Entrarán a los cálculos todos los casos, aun y cuando hayan sido definidos como missing.
- 2.- Eliminación LISTWISE de valores missing. Un caso será omitido del análisis si su valor en la variable independiente o cualquier dependiente es missing. El default es eliminación PAIRWISE, donde un caso será eliminado de todos los cálculos que envuelven esa variable.
- 3.- Suprime la impresión de las etiquetas de las variables.
- 4.- Ordena la impresión del número de casos, medias y desviaciones estándar para cada categoría en el archivo definido por el usuario. Para cada variable dependiente, un vector de frecuencias por categoría (formato 8F10.0), un vector de medias (formato 8F10.4) y un vector de desviaciones estándar (formato 8F10.4). Ver instrucción RAW OUTPUT UNIT.
- 6.- Usa los primeros 8 caracteres de las etiquetas de los valores de la variable independiente como etiqueta de grupo. Las etiquetas por default son de la forma GRP1, GRP2, etc.
- 7.- La entrada será frecuencias por categoría, medias y desviaciones estándar en vez de un archivo de datos. Las frecuencias se leen en formato 8F10.0, las medias y desviaciones estándar en formato 8F10.4, dando cada vector en una nueva tarjeta. La tarjeta READ MATRIX debe preceder a los datos.
- 8.- La entrada será frecuencias por categoría, medias, varianza combinada y grados de libertad para la varianza combinada. Las frecuencias y medias se dan como en la opción 7. La varianza combinada y los grados de libertad siguen en una nueva tarjeta con un formato de 2F10.4. Si se omiten los grados de libertad, se supondrá igual a 'N-K', donde 'N' es el total de número de casos y 'K' el número de categorías de la variable independiente. Los únicos análisis que pueden realizarse bajo esta opción son contrastes a-priori y a-posteriori.

Las estadísticas disponibles para ONEWAY son:

- 1.- Se imprimirá el número de casos para cada categoría, medias, desviaciones estándar, error estándar, mínimo, máximo e intervalo de confianza al 95% para la media.
- 2.- Se imprimirán efectos fijos y aleatorios.
- 3.- Se imprimirá la homocedasticidad de las estadísticas de varianza (Cochran's C, Bartlett-Box F y Hartley's F max).

```

RUN NAME      EJEMPLO DE RESULTADOS DE ONEWAY.
VARIABLE LIST ESCO VAR262
INPUT FORMAT  FIXED (2F1.0)
INPUT MEDIUM  DISK
N OF CASES    UNKNOWN

VAR LABELS    VAR262 TIENE CONFIANZA EN LA TELEVISION/
VALUE LABELS  VAR262 (1)NINGUNA (2)POCA(3)SUFICIENTE(4)MUCHA/
              ESCO (1)ANALF Y PRIM. INC (2)PRIMARIA (3)SECUNDA-RIA
              (4)PREPARATORIA (5)SUPERIOR/

ONEWAY        VAR262 BY ESCO(1,5)/
              POLYNOMIAL=3/ RANGES= LSD/ RANGES=SNK
OPTIONS       6
STATISTICS    ALL
READ INPUT DATA
    
```

----- O N E W A Y -----

VARIABLE VAR262 TIENE CONFIANZA EN LA TELEVISION  
 BY VARIABLE ESCO

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	4	93.1850	23.2962	32.416	0.0000
LINEAR TERM (UNWEIGHTED)	1	84.3327	84.3327	117.347	0.0000
LINEAR TERM (WEIGHTED)	1	91.1840	91.1840	126.880	0.0000
DEVIATION FROM LINEAR	3	2.0010	0.6670	0.928	0.4263
QUAD. TERM (UNWEIGHTED)	1	0.0063	0.0063	0.009	0.9253
QUAD. TERM (WEIGHTED)	1	0.0489	0.0489	0.068	0.7942
DEVIATION FROM QUAD.	2	1.9521	0.9760	1.358	0.2574
CUBIC TERM (UNWEIGHTED)	1	0.1090	0.1090	0.152	0.6970
CUBIC TERM (WEIGHTED)	1	0.2769	0.2769	0.385	0.5349
DEVIATION FROM CUBIC	1	1.6752	1.6752	2.331	0.1270
WITHIN GROUPS	1761	1265.5658	0.7187		
TOTAL	1765	1358.7507			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN	
ANALF Y	359	3.0167	0.9030	0.0477	1.0000	4.0000	2.9230	TO 3.1104
PRIMARIA	239	2.8201	0.8532	0.0532	1.0000	4.0000	2.7114	TO 2.9288
SECUNDA	544	2.7132	0.8829	0.0379	1.0000	4.0000	2.6389	TO 2.7876
PREPARAT	392	2.4515	0.7553	0.0382	1.0000	4.0000	2.3765	TO 2.5265
SUPERIOR	232	2.3405	0.8168	0.0536	1.0000	4.0000	2.2349	TO 2.4462
TOTAL	1766	2.6823	0.8774	0.0209	1.0000	4.0000	2.6414	TO 2.7233
		FIXED EFFECTS MODEL	0.8477	0.0202			2.6428	TO 2.7219
		RANDOM EFFECTS MODEL		0.1221			2.3432	TO 3.0215

RANDOM EFFECTS MODEL - ESTIMATE OF BETWEEN COMPONENT VARIANCE 0.0657

TESTS FOR HOMOGENEITY OF VARIANCES

COCHRAN'S C = MAX. VARIANCE/SUM(VARIANCES) = 0.2290, P = 0.090 (APPROX.)  
 BARTLETT-BOX F = 3.775, P = 0.005  
 MAXIMUM VARIANCE / MINIMUM VARIANCE = 1.429

----- O N E W A Y -----

VARIABLE VAR262 TIENE CONFIANZA EN LA TELEVISION  
 BY VARIABLE ESCO

MULTIPLE RANGE TEST

LSD PROCEDURE  
 RANGES FOR THE 0.050 LEVEL -

2.77 2.77 2.77 2.77  
 THE RANGES ABOVE ARE TABLE RANGES. THE VALUE ACTUALLY COMPARED WITH MEAN(J)-MEAN(I) IS:  
 $0.5994 * \text{RANGE} * \text{SQRT}(1/N(I) + 1/N(J))$

(\*) DENOTES PAIRS OF GROUPS SIGNIFICANTLY DIFFERENT AT THE 0.050 LEVEL

		S P S P A
		U R E R N
		P E C I A
		E P U M L
		R A N A F
		I R D R
		O A A I Y
MEAN	GROUP	R T - A
2.3405	SUPERIOR	
2.4515	PREPARAT	
2.7132	SECUNDA	***
2.8201	PRIMARIA	***
3.0167	ANALF Y	****

----- ONEWAY -----

VARIABLE VAR242 TIENE CONFIANZA EN LA TELEVISION  
 BY VARIABLE ESCD

MULTIPLE RANGE TEST

STUDENT-NEWMAN-KUELS PROCEDURE  
 RANGES FOR THE 0.050 LEVEL -

2.81 3.33 3.45 3.87

THE RANGES ABOVE ARE TABLE RANGES. THE VALUE ACTUALLY COMPARED WITH  $MEAN(I) - MEAN(J)$  IS.

$0.5994 * RANGE * \sqrt{(1/N(I) + 1/N(J))}$

HOMOGENEOUS SUBJECTS (SUBSETS OF GROUPS, WHOSE HIGHEST AND LOWEST MEANS DO NOT DIFFER BY MORE THAN THE SHORTEST SIGNIFICANT RANGE FOR A SUBSET OF THAT SIZE)

SUBSET 1

GROUP	SUPERIOR	PREPARAT
MEAN	2.3465	2.4515

SUBSET 2

GROUP	SECUNDA	PRIMARIA
MEAN	2.7132	2.8201

SUBSET 3

GROUP	ANLF Y
MEAN	3.0167

## 13.14)

## DISCRIMINANT

```

1
Sintaxis:  DISCRIMINANT  16
                GROUPS= <variable entera>(min,max)/
                SUBFILES.
                VARIABLES= <lista de variables>/
                ANALYSIS= <lista de variables>(nivel)/
                METHOD= DIRECT/ PRIORS= EQUAL/
                WILKS           SIZE
                MAHAL           <lista>
                MAXMINF
                MINRESID
                RAO
                FUNCTIONS= nf,cp,sig/ TOLERANCE= t/
                MAXSTEPS= m/ FIN= fi/ FOUT= fo/
                PIN= pi/ POUT= po/ VIN= vi

```

El subprograma DISCRIMINANT ejecuta un análisis discriminante. Este puede realizarse dando todas las variables discriminantes directamente al análisis o a través de 5 métodos STEPWISE que seleccionan el mejor conjunto de variables discriminantes.

El análisis discriminante sirve para distinguir estadísticamente entre dos o más grupos de casos, para lo cual, el investigador selecciona un conjunto de variables discriminantes que midan las características en las cuales se espera que los grupos difieran.

El criterio por el cual las variables independientes son seleccionadas durante el análisis es controlado por el usuario e indicado en la especificación METHOD. La técnica default es METHOD=DIRECT, mediante la cual todas las variables independientes entran al análisis simultáneamente. Las funciones discriminantes son creadas directamente del conjunto completo de las variables independientes, sin considerar el poder discriminante de cada una de ellas. El método directo es apropiado cuando por razones técnicas se desean tener todas las variables independientes dentro del análisis y no son de interés resultados interme-

dios basados en subconjuntos de las variables independientes. El método directo requiere menos tiempo de computadora y menos espacio en memoria en comparación a los métodos STEPWISE.

La alternativa al método directo es usar una selección por método STEPWISE. Las variables independientes serán seleccionadas para entrar al análisis según su poder discriminante. Como cada criterio hace énfasis en un aspecto diferente de "separación", debe tenerse cuidado en seleccionar el criterio apropiado.

Es común que el conjunto completo de variables independientes contenga exceso de información acerca de las diferencias entre grupos. Seleccionando en cada paso "el siguiente mejor discriminador", se elige un subconjunto de las variables, el cual es casi tan bueno, y algunas veces mejor, que el conjunto completo.

Con METHOD=WILKS, el criterio es el promedio de la F multivariada global para la prueba de diferencias entre los centroides de grupos. La variable que maximiza la F promedio, minimiza la Lambda de Wilks (medida de discriminación de grupos). Esta prueba toma en consideración las diferencias entre todos los centroides y la cohesión (homogeneidad) entre grupos. METHOD=MAHAL busca maximizar la distancia Mahalanobis entre los dos grupos más cercanos. Bajo METHOD=MAXINF es seleccionada la variable que maximiza la menor F promedio entre parejas de grupos. Aunque esta F promedio es la prueba de la distancia Mahalanobis calculada en METHOD=MAHAL, los dos criterios no producen necesariamente los mismos resultados cuando los grupos son de diferente tamaño. METHOD=MINRESID es otro criterio que tiende a separar grupos minimizando la variación de los residuales. METHOD=RAO se basa en la V de Rao, medida general de distancia. La variable selecciona-

da es aquella que contribuye al mayor incremento en V. Cuando hay en número grande de casos, el cambio en V es una distribución Ji-cuadrada con un grado de libertad, susceptible de ser probada para significancia estadística.

El uso de un procedimiento STEPWISE da como resultado un conjunto óptimo de variables. El resultado es sólo óptimo, y no global (MAXIMAL), debido a que no todos los posibles conjuntos son considerados, usualmente hay demasiadas variables para ejecutar esto en un tiempo razonable. Cuando el usuario desea que ciertas variables entren antes que otras, pueden especificarse niveles de inclusión para cada variable. Se debe hacer notar que la secuencia en que las variables son seleccionadas no es necesariamente la misma a su importancia relativa como discriminadores.

La especificación GROUPS define los grupos a ser usados en el análisis y VARIABLES el conjunto total de variables discriminantes. Estas especificaciones deben ir en primero y segundo lugares, respectivamente, y sólo una de cada una en una tarjeta DISCRIMINANT. A continuación va una especificación ANALYSIS que indica las variables discriminantes a ser usadas en el subanálisis, junto con algunas especificaciones opcionales asociadas. La especificación ANALYSIS y sus opciones pueden repetirse para realizar diferentes subanálisis, si se omiten, se ejecutará un análisis utilizando todas las variables en VARIABLES.

El usuario especifica los grupos colocando la palabra GROUPS seguida por el nombre de una variable entera y su rango de valores entre paréntesis, o por la palabra SUBFILES. Cuando le sigue una variable, los valores dentro del rango especificado son los que definen los grupos, valores fuera del rango, no entrarán

al análisis. Se recomienda al usuario recodificar la variable (si es necesario) de manera que sus valores sean consecutivos. Cuando la palabra SUBFILES sigue a GROUPS, indica que los grupos a ser discriminados son los subarchivos previamente definidos.

Cuando es procesado el archivo entero, el número de grupos discriminantes será igual al número de subarchivos en el archivo. Cuando se utiliza la instrucción RUN SUBFILES, el número de grupos en cada conjunto de archivos será igual al número de subarchivos en el grupo. Deben haber al menos 2 grupos discriminantes para poder ejecutar el análisis.

Ejemplo:

```
RUN SUBFILES (SA1,SA2,SA3)
DISCRIMINANT GROUPS= SUBFILES/ VARIABLES= VAR01 TO VAR20
```

Dará por resultado un análisis con todas las variables en VARIABLES como discriminantes, con todos los casos de los subarchivos SA1, SA2 y SA3 indicando 3 grupos discriminantes.

La especificación ANALYSIS es seguida por una lista de variables y, opcionalmente, por niveles de inclusión. El análisis será ejecutado usando sólo las variables dadas en esa lista. El usuario puede solicitar tantos subanálisis como desee definiendo cada uno por una lista ANALYSIS. Los niveles de inclusión sólo se usarán cuando se utilice un método STEPWISE para controlar el orden en el cual las variables serán consideradas para entrar al análisis. Un nivel de inclusión entre paréntesis puede seguir a cada variable o a una lista de variables. Si sigue a una lista, se aplicará a todas las variables en ella. El nivel de inclusión es un número entero entre 1 y 99, si ninguno se especifica, es asignado el valor 1.



Variables con nivel mayor entrarán antes que las de nivel menor. Si el nivel es un número FAR, todas las variables de ese nivel entrarán simultáneamente, si es NON, entrarán una por una en la secuencia determinada por el criterio de entrada. Las variables con nivel mayor que 1 permanecerán en el análisis; una vez adentro, con nivel 1, serán probadas a cada paso después de entrar para determinar si continúa haciendo suficiente contribución a la discriminación, si no, será eliminada del análisis.

Pueden colocarse parámetros adicionales después de cada instrucción ANALYSIS, para mayor control en el subanálisis, en cualquier orden y sus valores default serán asignados si se omiten. Los parámetros TOLERANCE, MAXSTEPS, FIN FOUT, PIN, POUT y VIN, sólo se usarán cuando se indique un método STEPWISE.

Antes que la variable sea probada durante el proceso STEPWISE, se verifica su nivel de tolerancia. La tolerancia mínima por default es .001, la que se puede modificar utilizando la opción TOLERANCE seguida por el número decimal deseado.

La especificación MAXSTEPS limita el número de pasos que ejecuta el procedimiento STEPWISE. El valor default es dos veces el número de variables listadas en ANALYSIS. Si se utiliza esta especificación, no debe ser mayor que el doble del número total de variables en VARIABLES.

Como ya se ha indicado, una variable es seleccionada sólo si su F promedio parcial es suficientemente grande, para lo cual, se realiza una prueba de significancia estadística de la cantidad de separación del centroide agregada por esta variable a la producida por las variables que ya entraron. Esta F promedio es llamada 'F-de-entrada', cuyo valor mínimo es controlado por la especi-

cación FIN con un valor default de 1.

En cada paso se prueba si las variables ya seleccionadas con nivel 1, todavía agregan una cantidad significativa a la separación, dadas las variables en la ecuación. Esta contribución es medida por la 'F-de-salida', la que es controlada por la especificación FOUT. Si la variable toma un valor de F menor que FOUT, se eliminará de la ecuación. FOUT tiene un valor default de 1.

La asignación estadística de una F promedio depende de sus grados de libertad. Para ambas, 'F-de-entrada' y 'F-de-salida', los grados de libertad dependen del número total de casos, el número de grupos y el número de variables que entraron. Si se quiere tener la mínima F en un nivel de significancia fijo y conocido, se puede hacer con las especificaciones PIN y POUT. Si un valor es dado para PIN, el nivel de significancia de la 'F-de-entrada' se calcula y compara contra PIN. Acción similar se toma con 'F-de-salida' cuando POUT es especificado. Cuando no se especifican PIN y POUT, las pruebas se hacen contra FIN y FOUT.

Cuando las variables se prueban para selección, deben pasar primero la prueba FIN o PIN para ser elegibles. Así la variable con el mejor criterio de entrada es seleccionada. No se verifica más el criterio de entrada, excepto con el método V de Rao, donde se realiza una prueba para el mínimo incremento. La magnitud de este incremento es controlado por la especificación VIN con un valor default de 0. Como es posible que el cambio en V sea negativo (indicando una separación decreciente), un valor negativo puede ser necesario y así forzar a seleccionar todas las variables. En realidad, si el usuario desea que el procedimiento STEPWISE continúe hasta que todas las variables especificadas

sean seleccionadas, no importando su contribución negativa o positiva, PIN y POUT deben igualarse a cero.

Cuando hay más de dos grupos, algunas veces es deseable limitar el número de funciones discriminantes que se deriven. El máximo número son: el número de variables discriminantes o una menos que el número de grupos. Si hay 'g' grupos, siempre hay un máximo potencial de (g-1) funciones, por razones estadísticas y técnicas, el investigador puede desear un número menor. El número de funciones puede ser impuesto a través de la especificación FUNCTIONS. Están disponibles tres maneras para fijar el límite:

- 'nf' Número máximo de funciones.
  - 'cp' Mínimo porcentaje acumulado de la suma de los valores EIGEN.
  - 'sig' Máximo nivel de significancia tolerado. Apropiado para datos obtenidos de una muestra aleatoria.
- Con valores default de (g-1), 100 y 1, respectivamente.

Las probabilidades de miembros de grupo pueden ajustarse por un conocimiento a-priori de la distribución de casos, lo que no es siempre conocido. Cuando el usuario desea hacer este ajuste, lo informa al programa a través de la opción PRIORS. El default es PRIORS=EQUAL mediante la cual todos los grupos se consideran de igual probabilidad y no se realiza ningún ajuste al calcular la probabilidad de los miembros del grupo. Con PRIORS=SIZE, la proporción de casos que caen en cada grupo se toma de la probabilidad a-priori de los grupos. Alternativamente, el usuario puede proporcionar una lista de probabilidades, una para cada grupo, debiendo sumar la lista 1.

Las opciones disponibles para esta subrutina son:

- 1.- Serán incluidos todos los casos durante el análisis, aun y cuando hayan sido declarados como missing, siempre y cuando cumplan con la especificación GROUPS.
- 2.- Durante la fase de clasificación todos los casos con datos missing serán procesados. Si el código de grupo es missing, el caso no será clasificado. Si en las variables discriminantes los datos son missing, la variable será substituida por la media total. Esta opción es incompatible con la 1.
- 3.- Suprime la impresión de resultados intermedios.
- 4.- Suprime la impresión de la tabla resumen STEPWISE.
- 5.- Imprime una tabla de resultados de clasificación, la cual indica para cada grupo, el número de casos y el porcentaje correcto de clasificación.
- 6.- Imprime para cada caso: identificación de subarchivo, número de valores missing, grupo al que pertenece, grupo más cercano (G), probabilidad de pertenecer al grupo G, y cinco puntajes de funciones discriminantes, si hay más de 5, se imprimirán líneas múltiples.
- 7.- Si sólo hay una función discriminante, graficará un histograma de la distribución de casos en la función. Si hay 2 o más funciones, se imprimirá una gráfica basada en las dos primeras funciones discriminantes. Cada caso es representado por un símbolo de grupo. Graficando, los centroides de grupo se imprimen como un asterisco.
- 8.- Igual a la opción 7, sólo que una gráfica será producida para cada grupo. Esta opción requiere un paso a través de los datos para cada grupo, lo que puede consumir tiempo excesivo si se tienen muchos casos en el archivo.
- 9.- Casos no incluidos en la especificación GROUPS, no entrarán a la fase de clasificación. Lo que también se aplica al cálculo de puntajes discriminantes y gráficas.
- 10.- Imprime un mapa territorial. Esta opción es aplicable sólo cuando hay una o dos funciones discriminantes.
- 11.- Imprime los coeficientes no estandarizados de las funciones discriminantes, así como la constante de ajuste a la media de la variable.
- 12.- Imprime las funciones de clasificación, así como la constante de ajuste a la media de la variable.
- 13.- Es ejecutada una rotación VARIMAX de las funciones discriminantes e imprime coeficientes estandarizados. Si se usa la

opción 11, los coeficientes rotados no-estandarizados también son impresos.

- 14.- Usa matrices de covarianza de grupos individuales para clasificación. Estas, serán empleadas en lugar de la matriz de covarianza combinada entre-grupos al calcular las probabilidades de pertenencia de grupo.
- 15.- DISCRIMINANT produce una salida de tipo matricial en el archivo definido por RAW OUTPUT UNIT. Esta matriz puede darse como entrada en análisis posteriores, en los que no puede realizarse fase de clasificación. La matriz se encuentra en registros de 80 columnas, utiliza un formato E13.6 y contiene el número de casos de cada grupo, las medias totales para todas las variables discriminantes listadas en VARIABLES, la matriz de covarianza combinada entre grupos y la matriz total de covarianza, iniciando cada uno de los 4 conjuntos en un nuevo registro. Si se especifica la opción 14, también serán impresas las medias y matrices de covarianza para cada grupo.
- 16.- DISCRIMINANT acepta una entrada de tipo matricial. Las tarjetas de control se colocan de la manera normal. El número y la secuencia de las variables en VARIABLES debe coincidir con las variables en la matriz. Si se especifican, las estadísticas 6, 7, y 8, y las opciones 1, 5, 6, 7, 8, 9, 14, 17, 18 y 19, serán ignoradas. Para leerla se coloca la tarjeta READ MATRIX después de las de DISCRIMINANT, OPTIONS y STATISTICS. Si se encuentra la matriz en tarjetas, debe seguir inmediatamente a READ MATRIX, si no, se leerá del dispositivo indicado por INPUT MEDIUM. El formato de la matriz debe ser idéntico al producido por la opción 15.

Ver instrucción RAW OUTPUT UNIT para las siguientes opciones:

- 17.- Imprime los puntajes discriminantes.
- 18.- Imprime probabilidades de pertenencia para todos los grupos.
- 19.- Imprime el grupo actual y números de clasificación de grupo.

Si no se utiliza OPTIONS, se llevará a cabo una eliminación LISTWISE de valores missing; si fue seleccionado un método STEPWISE, imprime todos los resultados intermedios; no entra a la fase de clasificación; si alguna opción de clasificación es seleccionada, entrará a la fase de clasificación pero producirá sólo los resultados solicitados; sólo imprimirá los coeficientes estandarizados de las funciones discriminantes.

Las estadísticas disponibles para DISCRIMINANT son:

- 1.- Serán impresas para cada grupo, para el conjunto total de casos, las medias de cada una de las variables discriminantes, que incluye todos los casos clasificados que no han sido eliminados por valores missing.
- 2.- Serán impresos para cada grupo y el conjunto total de casos, desviaciones estándar para las variables discriminantes
- 3.- Matriz de covarianza combinada entre grupos.
- 4.- Matriz de correlación combinada entre grupos.
- 5.- Produce una matriz PAIRWISE de F promedio, que consiste en una F para cada pareja de grupos. Esta F es la prueba de significancia para la distancia Mahalanobis entre grupos. Bajo las suposiciones de muestreo aleatorio y distribución normal multivariada, estas F pueden ser usadas para probar igualdad de parejas de centroides. Esta estadística está disponible si se seleccionó un método STEPWISE.
- 6.- Realiza una prueba de análisis de varianza de una vía para igualdad de medias de grupo en una sola variable discriminante. Se imprime una F para cada variable.
- 7.- Prueba igualdad de matrices de grupos de covarianza. Si se eligió la opción 14, también se calcula para las matrices de covarianza basadas en las funciones discriminantes.
- 8.- Se imprimirán matrices de covarianza de grupo para las variables discriminantes. Si se eligió la opción 14, también se imprimirán las matrices basadas en las funciones discriminantes.
- 9.- Se imprimirá la matriz total de covarianza.

NOTA: Seleccionar las estadísticas 7 u 8 (y/u opción 14) incrementa notablemente la cantidad de memoria y tiempo empleado.

Las estadísticas disponibles para DISCRIMINANT son:

- 1.- Serán impresas para cada grupo, para el conjunto total de casos, las medias de cada una de las variables discriminantes, que incluye todos los casos clasificados que no han sido eliminados por valores missing.
- 2.- Serán impresos para cada grupo y el conjunto total de casos, desviaciones estándar para las variables discriminantes
- 3.- Matriz de covarianza combinada entre grupos.
- 4.- Matriz de correlación combinada entre grupos.
- 5.- Produce una matriz PAIRWISE de F promedio, que consiste en una F para cada pareja de grupos. Esta F es la prueba de significancia para la distancia Mahalanobis entre grupos. Bajo las suposiciones de muestreo aleatorio y distribución normal multivariada, estas F pueden ser usadas para probar igualdad de parejas de centroides. Esta estadística está disponible si se seleccionó un método STEPWISE.
- 6.- Realiza una prueba de análisis de varianza de una vía para igualdad de medias de grupo en una sola variable discriminante. Se imprime una F para cada variable.
- 7.- Prueba igualdad de matrices de grupos de covarianza. Si se eligió la opción 14, también se calcula para las matrices de covarianza basadas en las funciones discriminantes.
- 8.- Se imprimirán matrices de covarianza de grupo para las variables discriminantes. Si se eligió la opción 14, también se imprimirán las matrices basadas en las funciones discriminantes.
- 9.- Se imprimirá la matriz total de covarianza.

NOTA: Seleccionar las estadísticas 7 u 8 (y/u opción 14) incrementa notablemente la cantidad de memoria y tiempo empleado.

COMMENT EJEMPLO DE RESULTADOS DE DISCRIMINANT.

VAR LABELS NIVEL NIVEL SOCIOECONOMICO DEL ENTREVISTADO/  
NOTIC PREFERENCIA DE PROGRAMAS TIPO NOTICIEROS

VALUE LABELS NIVEL (1)ALTO (2)MEDIO (3)BAJO/  
SEXO (0)HOMBRES (1)MUJERES/  
EDAD (1)7-11 (2)12-14 (3)15-17 (4)18-22 (5)23-30  
(6)31-44 (7)45 Y MAS/  
ESTADO (0)SOLTERO (1)CASADO/  
ESCO (1)PRIMARIA (2)SECUNDARIA (3)PREPA (4)LICEN-POSS/  
ESCUELA (0)PUBLICA (1)PRIVADA/  
NOTIC (1)1ER A 3ER LUGAR (0)LOS DEMAS/

DISCRIMINAT GROUPS= NOTIC(0,1)/  
VARIABLES= NIVEL SEXO EDAD ESTADO ESCO ESCUELA/  
METHOD=WILKS

STATISTICS 7  
OPTIONS 5

SINCE ANALYSIS= WAS OMITTED FOR THE FIRST ANALYSIS, ALL VARIABLES  
ON THE VARIABLES= LIST WILL BE ENTERED AT LEVEL 1.

----- DISCRIMINANT ANALYSIS -----  
ON GROUPS DEFINED BY NOTIC PREFERENCIA DE PROGRAMAS TIPO NOTICIEROS

1197 (UNWEIGHTED) CASES WERE PROCESSED.  
0 OF THESE WERE EXCLUDED FROM THE ANALYSIS.  
1197 (UNWEIGHTED) CASES WILL BE USED IN THE ANALYSIS.

NUMBER OF CASES BY GROUP

NOTIC	NUMBER OF CASES		LABEL
	UNWEIGHTED	WEIGHTED	
0	809	809.0	LOS DEMAS
1	388	388.0	1ER A 3ER LUGAR
TOTAL	1197	1197.0	

----- DISCRIMINANT ANALYSIS -----  
ON GROUPS DEFINED BY NOTIC PREFERENCIA DE PROGRAMAS TIPO NOTICIEROS

ANALYSIS NUMBER 1  
STEPWISE VARIABLE SELECTION

SELECTION RULE: MINIMIZE WILKS' LAMBDA

MAXIMUM NUMBER OF STEPS.....	12
MINIMUM TOLERANCE LEVEL.....	0.00100
MINIMUM F TO ENTER.....	1.0000
MAXIMUM F TO REMOVE.....	1.0000

CANONICAL DISCRIMINANT FUNCTIONS

MAXIMUM NUMBER OF FUNCTIONS.....	1
MINIMUM CUMULATIVE PERCENT OF VARIANCE...	100.00
MAXIMUM SIGNIFICANCE OF WILKS' LAMBDA....	1.0000



PRIOR PROBABILITY FOR EACH GROUP IS 0.50000

VARIABLES NOT IN THE ANALYSIS AFTER STEP 0				
VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MILKS' LAMBDA
NIVEL	1.0000000	1.0000000	0.85274	0.9992869
SEXO	1.0000000	1.0000000	3.9465	0.9967083
EDAD	1.0000000	1.0000000	75.258	0.9407540
ESTADO	1.0000000	1.0000000	34.379	0.9720351
ESCD	1.0000000	1.0000000	9.6405	0.9919972
ESQUELA	1.0000000	1.0000000	3.5676	0.9970235

.....

AT STEP 1, EDAD WAS INCLUDED IN THE ANALYSIS.				
MILKS' LAMBDA	DEGREES OF FREEDOM	SIGNIFICANCE	BETWEEN GROUPS	
0.9407540	1 1	1195.0		
EQUIVALENT F	75.25771	1	1195.0	0.0000

VARIABLES IN THE ANALYSIS AFTER STEP 1				
VARIABLE	TOLERANCE	F TO REMOVE	MILKS' LAMBDA	
EDAD	1.0000000	75.258		

VARIABLES NOT IN THE ANALYSIS AFTER STEP 1				
VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MILKS' LAMBDA
NIVEL	0.9994660	0.9994660	0.49157	0.9403660
SEXO	0.9993731	0.9993731	4.5481	0.9371685
ESTADO	0.4718965	0.4718965	0.38716	0.9404490
ESCD	0.9864750	0.9864750	4.1862	0.9374672
ESQUELA	0.9887795	0.9887795	0.90923	0.9408381

.....

AT STEP 2, SEXO WAS INCLUDED IN THE ANALYSIS.				
MILKS' LAMBDA	DEGREES OF FREEDOM	SIGNIFICANCE	BETWEEN GROUPS	
0.9371685	2 1	1195.0		
EQUIVALENT F	40.02323	2	1194.0	0.0000

VARIABLES IN THE ANALYSIS AFTER STEP 2				
VARIABLE	TOLERANCE	F TO REMOVE	MILKS' LAMBDA	
SEXO	0.9993731	4.5481	0.9407540	
EDAD	0.9993731	75.857	0.9967083	

----- VARIABLES NOT IN THE ANALYSIS AFTER STEP 2 -----

VARIABLE	TOLERANCE	MINIMUM		F TO ENTER	WILKS' LAMBDA
		TOLERANCE			
NIVEL	0.998737	0.9987808		0.41961	0.9368390
ESTADO	0.4660670	0.4660670		0.14912	0.9370514
ESCO	0.9615138	0.9615138		2.9723	0.9348394
ESQUELA	0.9887249	0.9884040		0.93548	0.9364342

AT STEP 3, ESCO WAS INCLUDED IN THE ANALYSIS.

WILKS' LAMBDA	EQUIVALENT F	DEGREES OF FREEDOM	SIGNIFICANCE	BETWEEN GROUPS
0.9348394	27.71832	3 1	1195.0	
		3	1193.0	0.0000

----- VARIABLES IN THE ANALYSIS AFTER STEP 3 -----

VARIABLE	TOLERANCE	F TO REMOVE	WILKS' LAMBDA
SEXO	0.9740856	3.3534	0.9374672
EDAD	0.9845744	70.507	0.9900891
ESCO	0.9615138	2.9723	0.9371685

----- VARIABLES NOT IN THE ANALYSIS AFTER STEP 3 -----

VARIABLE	TOLERANCE	MINIMUM		F TO ENTER	WILKS' LAMBDA
		TOLERANCE			
NIVEL	0.8083314	0.8069762		0.223628-02	0.9348377
ESTADO	0.4474675	0.4453382		0.181038-02	0.9348388
ESQUELA	0.9829029	0.9365388		1.2172	0.9338828

AT STEP 4, ESQUELA WAS INCLUDED IN THE ANALYSIS.

WILKS' LAMBDA	EQUIVALENT F	DEGREES OF FREEDOM	SIGNIFICANCE	BETWEEN GROUPS
0.9338828	21.09683	4 1	1195.0	
		4	1192.0	0.0000

----- VARIABLES IN THE ANALYSIS AFTER STEP 4 -----

VARIABLE	TOLERANCE	F TO REMOVE	WILKS' LAMBDA
SEXO	0.9740683	3.3265	0.9364920
EDAD	0.9718837	67.138	0.9855014
ESCO	0.9365888	3.2528	0.9364342
ESQUELA	0.9829029	1.2172	0.9348394

VARIABLES NOT IN THE ANALYSIS AFTER STEP 4

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	MILKS' LAMBDA
NIVEL	0.8379133	0.8017572	0.516630-02	0.9338818
ESTADO	0.4473591	0.4436376	0.356080-02	0.9338830

F LEVEL OR TOLERANCE OR VIN INSUFFICIENT FOR FURTHER COMPUTATION.

SUMMARY TABLE

STEP	ACTION ENTERED	VARS REMOVED	MILKS' LAMBDA IN	SIG.	LABEL
1	EDAD	1	0.940754	0.0000	
2	SEXO	2	0.937169	0.0000	
3	ESCO	3	0.934839	0.0000	
4	ESQUELA	4	0.933886	0.0000	

CANONICAL DISCRIMINANT FUNCTIONS

FUNCTION	EIGENVALUE	PERCENT VARIANCE	CUMULATIVE PERCENT	CANONICAL CORRELATION :	AFTER FUNCTION	MILKS' LAMBDA	CHI-SQUARED	D.F.	SIGNIFICANCE
1*	0.07079	100.00	100.00	0.2571268	0	0.9338830	81.603	4	0.0000

\* MARKS THE 1 CANONICAL DISCRIMINANT FUNCTION(S) TO BE USED IN THE REMAINING ANALYSIS.

STANDARDIZED CANONICAL DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1
SEXO	-0.20788
EDAD	0.91109
ESCO	0.20754
ESQUELA	-0.12529

CANONICAL DISCRIMINANT FUNCTIONS EVALUATED AT GROUP MEANS (GROUP CENTROIDS)

GROUP	FUNC 1
0	-0.18411
1	0.38388

TEST OF EQUALITY OF GROUP COVARIANCE MATRICES USING BOX'S M  
THE RANKS AND NATURAL LOGARITHMS OF DETERMINANTS PRINTED ARE THOSE OF THE GROUP COVARIANCE MATRICES.

GROUP LABEL	RANK	LOG DETERMINANT	
0 LOS DEVS	4	1.044737	
1 1ER A 3ER LUGAR	4	2.063744	
POOLED WITHIN-GROUPS COVARIANCE MATRIX	4	1.502140	
BOX'S M	APPROXIMATE F	DEGREES OF FREEDOM	SIGNIFICANCE
152.25	15.160	10,	2842065.9 0.0000

CLASSIFICATION RESULTS -

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP	
		0	1
GROUP 0	809	461	348
LOS DENAS		57.0%	43.0%
GROUP 1	388	114	274
IER A 3ER LUAR		29.4%	70.6%

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED: 61.40%

CLASSIFICATION PROCESSING SUMMARY

1197 CASES WERE PROCESSED.

1197 CASES WERE USED FOR PRINTED OUTPUT.

## 13.15)

## FACTOR

```

1
Sintaxis: FACTOR
16
VARIABLES= <lista de variables>/
TYPE= PA1 / ROTATE= VARIMAX /
      PA2          QUARTIMAX
      RAO          EQUIMAX
      ALPHA        OBLIQUE
      IMAGE        NOROTATE
      BYPASS
DIAGONAL=<lista valores>/ NFACTORS=<valor>/
MINEIGEN= <valor>/ ITERATE= <valor>/
STOPFACT= <valor>/ DELTA= <valor>/
FACSCORE= <prvm>

```

El subprograma FACTOR ejecuta una variedad de técnicas analíticas-factoriales. Acepta como entrada un archivo de casos, una matriz de correlación o una matriz factorial. Dispone de 5 diferentes métodos factoriales y 4 métodos alternativos de rotación a las soluciones factoriales, 3 ortogonales y uno oblio. Además, puede producir gráficas de relaciones entre parejas de factores rotados, con cualquier procedimiento de rotación ortogonal.

El usuario tiene total control sobre el criterio de extracción y rotación de factores, el número de factores a extraer y rotar, el mínimo valor EIGEN y los valores de la diagonal de la matriz de correlación.

La característica más distintiva del análisis factorial es su capacidad de reducción. Dados los coeficientes de correlación para un conjunto de variables, técnicas analíticas-factoriales permiten ver si existe algún patrón importante de relación para que los datos puedan ser reducidos a un conjunto más pequeño de 'factores' que puedan tomarse como variables fuente para las interrelaciones observadas en los datos.

Las aplicaciones más comunes del análisis factorial pueden ser: (1) Exploratorios- exploración y detección de patrones de

variables para descubrir nuevos conceptos y una posible reducción de datos; (2)Confirmatorios- prueba de hipótesis sobre la estructura de las variables en términos del número esperado de factores significantes y cargas factoriales; y (3)Fuente de Medición- construcción de índices para su uso como nuevas variables.

Para ejecutar el subprograma, se coloca la palabra FACTOR en el campo de control seguida por la palabra VARIABLES como primer parámetro en el campo de especificación. A continuación, una lista de hasta 62 variables a incluir en el análisis factorial.

Para la solución de factores iniciales, se selecciona el método deseado por medio de la especificación TYPE, seguida por la palabra que especifica el método de factorización a ser empleado.

PA1	Componentes principales.
PA2	Factorización principal con iteración (DEFAULT)
RAD	Factorización canónica de Rao.
ALPHA	Factorización ALPHA.
IMAGE	Factorización IMAGE.
BYPASS	Ningún factor es extraído.

Si el parámetro BYPASS es usado, FACTOR obtendrá la matriz de correlación para las variables en VARIABLES y la imprimirá, si se solicita, pero no hará más cálculos. La especificación TYPE puede omitirse, en cuyo caso se supondrá el método PA2.

Si se elige TYPE=PA1, se puede alterar la diagonal principal de la matriz de correlación por medio de la especificación DIAGONAL y a continuación una lista de estimadores de comunalidad para cada variable en el orden especificado en VARIABLES.

Cuando se desea cambiar la diagonal, comúnmente será usado: (1)el mayor valor absoluto de los coef. de corr. en cada columna, (2)la correlación múltiple cuadrada de cada variable con las demás, o (3)cualquier valor entre la correlación múltiple cuadra-

da y 1. Los valores serán siempre positivos y menores o iguales a 1. Cuando se desea dar un mismo 'valor' a 'n' diagonales adyacentes, se puede indicar de la sig. manera: <n\*valor>.

El parámetro DIAGONAL es opcional, si se omite, se supondrá 1 para todos los elementos de la diagonal de la matriz de correlación. En otros métodos diferentes a PA1, las diagonales son reemplazadas por los valores requeridos por cada método. DIAGONAL sólo debe usarse con TYPE=PA1.

El usuario controla el procedimiento de factorización por medio de los parámetros NFACTORS, MINEIGEN, ITERATE y STOPFACT. Sin embargo, no todas son aplicables a todos los métodos de factorización. NFACTORS y MINEIGEN, métodos alternativos para determinar el número de factores a extraer, deben usarse con PA1 y PA2. ITERATE y STOPFACT que controlan el proceso iterativo en PA2, RAD y ALPHA, no tienen sentido para IMAGE y PA1. Estas instrucciones son opcionales, en la mayoría de los casos se les han dado valores estándar, convencionales a los métodos de factorización. En otros casos como ITERATE, se le ha dado un valor que previene consumir inadvertidamente gran cantidad de tiempo de máquina. Se recomienda modificar estos valores default sólo cuando se tenga una buena razón para hacerlo.

Si se conoce el número de factores a extraer, se puede especificar en NFACTORS, el cual no puede exceder el número de variables y puede ser igual si se desea la solución completa de PA1. NFACTORS se requiere si se da como entrada una matriz factorial.

El número de factores extraídos es determinado normalmente por el mínimo valor EIGEN. Convencionalmente, FACTOR elimina todos los factores con un valor EIGEN menor a 1. Si se desea

alterar el criterio, el valor deseado se especifica en MINEIGEN.

Si NFACTORS es usado, tomará precedencia sobre MINEIGEN. Normalmente se utilizará MINEIGEN sólo si se tiene una buena razón para modificar el número de factores a extraer. De otro modo, el valor EIGEN default de 1 debe bastar. Debe tenerse cuidado al usar MINEIGEN con otros procedimientos PA1 o PA2.

Los métodos PA2, RAD y ALPHA emplean un método de convergencia iterativa para alcanzar las comunalidades finales. Tales procedimientos son costosos en términos de tiempo de máquina ya que pueden requerir de cientos de iteraciones para conseguir una convergencia relativamente exacta. Por otro lado, no está determinada la exactitud entre el 2o y 3er dígito. Por estas razones, se ha fijado el número máximo de iteraciones en 25. El número de iteraciones puede modificarse en ITERATE, el cual se reporta en la salida con un mensaje que indica si se ha alcanzado una convergencia adecuada, o no.

Una convergencia satisfactoria en el proceso iterativo se indica con el parámetro STOPFACT. Si se omite, la iteración termina si 2 conjuntos sucesivos de estimadores de comunalidad son diferentes entre sí pero no más de 0.001. Si se desea modificar este valor, y así el número de iteraciones, se indica con STOPFACT, no debiendo tener más de 7 dígitos significativos a la derecha del punto decimal y 1 como valor máximo.

Se puede seleccionar entre 4 métodos de rotación, o por ninguna, por medio del parámetro ROTATE. Los 3 métodos ortogonales son VARIMAX, QUARTIMAX, y EQUIMAX. El método OBLIQUE puede ser modificado por la especificación DELTA. Si no se desea rotación, se da la palabra NOROTATE, en donde la solución factorial



ortogonal inicial será aceptada como la solución final, y si se solicitan puntajes factoriales, se basarán en la solución inicial. La rotación tampoco se ejecutará si NFACTORS=1. Si se omite ROTATE, se ejecutará la rotación VARIMAX.

Se puede controlar la rotación oblicua especificando el parámetro DELTA cuyo valor default es cero. Si se desea mayor o menor rotación, se modifica el valor de DELTA. También pueden solicitarse varias rotaciones, una para cada valor en DELTA, lo que permite determinar cómo se afecta la estructura factorial para diferentes valores de correlación entre factores. Con:

DELTA= <valor inicial>,<incremento>,<valor final>/

se pueden dar valores múltiples en DELTA, pero las matrices de puntajes factoriales no serán calculadas, aunque se soliciten, ni se imprimirán los puntajes factoriales para casos individuales.

El parámetro FACSCORE especifica que se desea a impresión los puntajes factoriales, uno para cada factor en la solución final, para cada caso procesado por FACTOR. Si además se desea los puntajes de casos con algunas variables con valores missing, se indica la proporción de valores missing a ser reemplazados por la media de la variable, a continuación de FACSCORE. Si ningún valor missing se reemplazará, un slash seguirá inmediatamente a FACSCORE. La proporción deberá ser un valor entre 0 y 1. Si se omite FACSCORE, los puntajes factoriales no serán impresos. El reemplazo de los valores missing sólo se llevará a cabo, cuando el análisis factorial se ha ejecutado utilizando la opción 2, eliminación PAIRWISE de valores missing. Los puntajes factoriales no pueden calcularse si se da como entrada una matriz.

Puede solicitarse más de un análisis factorial en una tar-

jeta FACTOR. Al terminar de dar todos los parámetros necesarios para un análisis, puede definirse otro, dando otra lista VARIABLES y los parámetros necesarios para definir ese análisis. No hay límite en el número de análisis solicitados. Sin embargo, FACTOR opera diferente a los demás subprogramas de SPSS en que las variables referenciadas en la segunda y subsecuentes listas VARIABLES, deben ser un subconjunto de las variables de la primera lista. La matriz de correlación se calcula sólo una vez, para la primera lista, y para las sig. listas se seleccionarán los renglones y columnas apropiadas de la matriz de correlación.

Con excepción de VARIABLES, que debe ser el primer parámetro definido, el orden de todos los demás parámetros es opcional.

El subprograma FACTOR permite al usuario dar los datos en la forma usual como casos, o como una matriz de correlación simple, o una matriz factorial. Esto permite saltar el paso más costoso de calcular la matriz de correlación. Una vez con las correlaciones, se pueden tratar diferentes métodos de factorización y rotación sin leer los datos cada vez. Sin embargo, dar una matriz de coeficientes es completamente diferente a dar un archivo de casos, se deben observar ciertas convenciones.

FACTOR acepta matrices de correlación en el formato matricial estándar de SPSS, que pueden obtenerse de las subrutinas PEARSON CORR, NONPAR CORR, PARTIAL CORR, REGRESSION o del mismo FACTOR. SPSS supone que la matriz es cuadrada, simétrica y unos en la diagonal. Los coeficientes deben estar en formato BF10.7 y cada renglón de la matriz debe iniciar en un nuevo registro.

- Se recomienda usar una matriz de correlación producida por PEARSON CORR, PARTIAL CORR, NONPAR CORR o REGRESSION cuando el archivo contiene muchos casos, debido a que estos procedimientos se ejecutan en doble precisión mientras que FACTOR en simple precisión.

Hay dos métodos para especificar la estructura de una matriz de correlación de entrada a FACTOR. Con la opción 3, el número y orden de las variables en la matriz debe corresponder exactamente al de la lista en VARIABLES. Especificando la opción 9 en adición a la 3, la estructura de la matriz estará definida por el número y orden de las variables en la instrucción VARIABLE LIST. Método idéntico al opcional en la subrutina PARTIAL CORR. Con este método se da una matriz y ejecuta un análisis factorial en cualquier subconjunto de variables de la matriz.

Cuando el usuario desea dar algunos resultados de factorización ortogonal y desea rotar y/o graficar los resultados, esto puede ejecutarse dando la matriz factorial y su correspondiente vector de comunalidad (no acepta matrices oblicuas).

Las matrices factoriales de entrada a FACTOR deben darse en formato estándar. Cuando la matriz y vector asociado no se han obtenido de FACTOR, se debe verificar que el formato sea compatible. Todas las cargas factoriales de la primera variable se dan en formato BF10.7. El primer factor de cada variable inicia en un nuevo registro. Los estimadores de comunalidad van a continuación de la matriz, con formato BF10.7 también.

La estructura de una matriz factorial de entrada debe estar definida por VARIABLES. Sólo puede ejecutarse un análisis factorial cuando se da una matriz factorial, por lo tanto sólo debe haber una especificación VARIABLES.

Cuando se procesan subarchivos (mediante RUN SUBFILES), debe darse ya sea una matriz de correlación o una matriz factorial, para cada subarchivo. Deben estar colocadas en el archivo en el orden que serán procesadas y nada debe haber entre ellas.

Cuando se da cualquiera de las matrices, se debe informar al subprograma por medio de la tarjeta OPTIONS. La opción 3 para cuando se da una matriz de correlación, la opción 4 si una matriz factorial. Adicionalmente deben prepararse las instrucciones:

- 1.- VARIABLE LIST.
- 2.- INPUT MEDIUM.
- 3.- N OF CASES. Que se utilizará sólo para pruebas de significancia y no tendrá nada que ver con la lectura de la matriz.
- 4.- READ MATRIX. Debe seguir a OPTIONS o STATISTICS y debe preceder a la matriz si se encuentra en tarjetas. Tiene la misma finalidad que READ INPUT DATA.

Las opciones disponibles para esta subrutina son:

- 1.- Entrarán todos los casos, incluso los valores missing, en el cálculo de las correlaciones simples.
- 2.- Eliminación PAIRWISE de valores missing. Un caso se omitirá de los cálculos de un coeficiente dado, si el valor de alguna de las variables consideradas es missing. La eliminación PAIRWISE tiene la ventaja de utilizar todos los datos posibles en el cálculo de cada coeficiente. Tiene la desventaja de, en algunas circunstancias, producir correlaciones basadas en un número diferente de casos y quizás hasta en subpoblaciones completamente diferentes.

Si no se especifica la opción 1 ó 2, se tendrá una eliminación LISTWISE de valores missing. Un caso no se incluirá en el cálculo de los coeficientes si contiene un valor missing en cualquier variable dada en VARIABLES. Los casos eliminados tampoco entrarán al cálculo de un análisis subsecuente, aunque éstos no tengan valores missing en el subconjunto definido de variables. En general, la eliminación LISTWISE reduce el número de casos sobre los cuales se ejecuta el análisis factorial. Por otro lado, se asegura que el análisis se ha obtenido del mismo conjunto de casos.

- 3.- Informa a FACTOR que será leída una matriz de correlación en vez de un archivo de datos. Cuando se contemplan varios experimentos con diferentes factorizaciones y métodos de rotación o se tiene un archivo muy grande de casos, se recomienda la entrada de una matriz de correlación, ya que representa un considerable ahorro en tiempo de máquina.
- 4.- Indica la entrada de una matriz factorial y su vector de comunialidades asociado, que pueden producirse por medio de la opción 6, cuando se ha decidido en el método de factorización pero se desea experimentar con diferentes soluciones rotacionales. NFACTORS especifica el número de factores a leerse.

Para las opciones 5,6,7 y 8, ver la instrucción RAW OUTPUT UNIT.

- 5.- Ordena la impresión de la matriz de correlación simple con todas las variables que entraron al análisis factorial.
- 6.- Imprime la matriz factorial y su vector de comunalidades asociado.
- 7.- Imprime la matriz de coeficientes de puntajes factoriales.
- 8.- Esta opción ordena la impresión de medias e, iniciando en un nuevo registro, las desviaciones estándar de las variables en la primera lista VARIABLES con formato de 8F10.4. Si la opción 5 y/o 6 también es usada, el orden en el archivo será: (1)Medias, (2)Desviaciones Estándar, (3)Matriz de Correlación, y (4)Matriz factorial y comunalidades.
- 9.- El orden y número de las variables en la matriz de correlación de entrada se especifica por medio de la instrucción VARIABLE LIST. Esta opción sólo puede usarse junto con la opción 3.
- 10.- Si se especifica reemplazo de valores missing (FACSCORE=n), esta opción producirá puntajes factoriales que son productos ponderados de los datos existentes. Esta opción sólo estará en efecto junto con la opción 2.
- 11.- Si se solicita la impresión de puntajes factoriales (por medio de FACSCORE) el formato bajo el cual serán impresos es:  
col 1 a 8 la variable SEQNUM,  
col 9 a 10 el número del registro,  
col 12 a 15 la variable SUBFILE,  
col 21 a 80 seis puntajes factoriales formato 6F10.6  
(formato default 8F10.6).

Si se omite la tarjeta OPTIONS, FACTOR no aceptará entrada de matrices, los valores missing serán excluidos LISTWISE, no producirá impresión de matrices y los puntajes factoriales no serán ponderados ni secuenciados.

Las estadísticas disponibles para FACTOR son:

- 1.- Medias y desviaciones estándar de las variables en VARIABLES, junto con el número de casos válidos observados.
- 2.- La matriz de correlación simple de orden cero para todas las variables que entraron al análisis factorial.
- 3.- La inversa de la matriz de correlación y su determinante.

- 4.- Ordena la impresión de: (1) los valores EIGEN asociados con los factores iniciales rotados y no-rotados; (2) la proporción de la varianza total en base a los factores iniciales; (3) la proporción de la varianza común en base a los factores rotados; y (4) los estimadores iniciales y finales de comunalidad.
- 5.- Matriz factorial no-rotada. Esta matriz sólo contiene el número de factores significantes o especificados por NFACTORS. Si se desea la matriz completa, NFACTORS debe igualarse al número de variables.
- 6.- Matriz factorial rotada. Para rotaciones ortogonales la matriz de transformación usada para convertir la matriz inicial en la matriz rotada final también será impresa. Para la rotación OBLIQUE, también se imprimirá la matriz patrón de factores rotados, la matriz de correlación entre los factores rotados y la matriz estructural de factores rotados.
- 7.- Matriz de coeficientes de puntajes factoriales. Comúnmente usada para construir variables compuestas basadas en las soluciones factoriales rotadas. No está disponible, si se da como entrada una matriz factorial.
- 8.- Para rotaciones ortogonales, se graficará la posición relativa de cada variable en todas las posibles parejas de ejes de factores. Si se tienen muchos factores, la gráfica puede ser muy voluminosa. No está disponible para rotación oblicua.

Si se omite la instrucción STATISTICS, las estadísticas 4, 5 y 6 serán impresas, ya que usualmente contienen información valiosa. Sin embargo, si se da la tarjeta STATISTICS, las estadísticas 4, 5 y 6 serán impresas sólo si son referenciadas específicamente.

RUN NAME	EJEMPLO DE RESULTADOS DE FACTOR.
VARIABLE LIST	VAR230 TO VAR245
INPUT FORMAT	FIXED (16F1.0)
INPUT MEDIUM	DISK
N OF CASES	UNKNOWN
VAR LABELS	VAR230 CONFIANZA IGLESIA/ VAR231 CONFIANZA PARTIDOS POLITICOS/ VAR232 CONFIANZA COMUNIDAD MEDICA/ VAR233 CONFIANZA POLICIA/ VAR234 CONFIANZA GRANDES EMPRESAS/ VAR235 CONFIANZA FAMILIA/ VAR236 CONFIANZA GOBIERNO/ VAR237 CONFIANZA UNIVERSIDADES/ VAR238 CONFIANZA PRENSA/ VAR239 CONFIANZA EJERCITO/

VAR260 CONFIANZA CIENCIA Y TECNOLOGIA/  
 VAR261 CONFIANZA SINDICATOS/  
 VAR262 CONFIANZA TELEVISION/  
 VAR263 CONFIANZA MATRIMONIO/  
 VAR264 CONFIANZA PRESIDENTES DE MEXICO/  
 VAR265 CONFIANZA PUBLICIDAD/

VALUE LABELS VAR250 TO VAR265 (1)NINGUNA (2)POCA(3)SUFICIENTE(4)MUCHA/

FACTOR VARIABLES= VAR250 TO VAR265

STATISTICS 4,5,6,7,8

READ INPUT DATA

1.VARIABLE LIST

VARIABLES..	LABELS..
VAR250	CONFIANZA IGLESIA
VAR251	CONFIANZA PARTIDOS POLITICOS
VAR252	CONFIANZA COMUNIDAD MEDICA
VAR253	CONFIANZA POLICIA
VAR254	CONFIANZA GRANDES EMPRESAS
VAR255	CONFIANZA FAMILIA
VAR256	CONFIANZA GOBIERNO
VAR257	CONFIANZA UNIVERSIDADES
VAR258	CONFIANZA PRENSA
VAR259	CONFIANZA EJERCITO
VAR260	CONFIANZA CIENCIA Y TECNOLOGIA
VAR261	CONFIANZA SINDICATOS
VAR262	CONFIANZA TELEVISION
VAR263	CONFIANZA MATRIMONIO
VAR264	CONFIANZA PRESIDENTES DE MEXICO
VAR265	CONFIANZA PUBLICIDAD

DETERMINANT OF CORRELATION MATRIX = 0.0472538 (0.47253788-01)

VARIABLE	EST COMMUNITY	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
VAR250	0.12159	1	4.29698	26.9	26.9
VAR251	0.22350	2	1.53491	9.6	36.4
VAR252	0.19526	3	1.11366	7.0	43.4
VAR253	0.27418	4	1.00125	6.3	49.7
VAR254	0.21225	5	0.90613	5.7	55.3
VAR255	0.13293	6	0.85476	5.3	60.7
VAR256	0.39329	7	0.81068	5.1	65.7
VAR257	0.27743	8	0.72908	4.6	70.3
VAR258	0.28971	9	0.71021	4.4	74.7
VAR259	0.31290	10	0.68917	4.3	79.0
VAR260	0.25864	11	0.64919	4.1	83.1
VAR261	0.26968	12	0.61999	3.9	87.0
VAR262	0.20816	13	0.60384	3.8	90.7
VAR263	0.12652	14	0.56128	3.5	94.3
VAR264	0.39241	15	0.48859	3.1	97.3
VAR265	0.25413	16	0.43023	2.7	100.0

CONVERGENCE REQUIRED 21 ITERATIONS

FACTOR MATRIX USING PRINCIPAL FACTOR WITH ITERATIONS

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
VAR250	0.28810	-0.08395	0.32453	0.16619
VAR251	0.47096	-0.15597	-0.06579	0.06658
VAR252	0.40522	0.25572	0.08685	0.22037
VAR253	0.51966	-0.22696	-0.03111	0.16940
VAR254	0.47397	0.14626	-0.02226	0.01233
VAR255	0.25934	0.29821	0.24704	0.18912
VAR256	0.63559	-0.17523	-0.21517	0.15489
VAR257	0.45524	0.41890	-0.06954	0.01913
VAR258	0.55203	-0.02303	-0.00728	-0.13545
VAR259	0.57819	-0.03017	-0.03539	-0.11201
VAR260	0.39665	0.51780	-0.13871	-0.18037
VAR261	0.54306	-0.11496	-0.04800	-0.07835
VAR262	0.43899	-0.24155	0.35036	-0.25549
VAR263	0.30437	0.17945	0.14182	-0.01678
VAR264	0.61253	-0.26222	-0.13044	0.07135
VAR265	0.50805	-0.02923	0.04300	-0.17995

VARIABLE	COMMUNALITY	FACTOR	EIGENWALLE	PCT OF VAR	CUM PCT
VAR250	0.22299		3.45211	68.7	68.7
VAR251	0.25489	2	0.90778	17.1	85.8
VAR252	0.28570	3	0.41538	7.8	93.6
VAR253	0.35123	4	0.34217	6.4	100.0
VAR254	0.24649				
VAR255	0.25298				
VAR256	0.50497				
VAR257	0.38792				
VAR258	0.32367				
VAR259	0.34902				
VAR260	0.47722				
VAR261	0.31657				
VAR262	0.43909				
VAR263	0.14531				
VAR264	0.46605				
VAR265	0.29320				



VARIMAX ROTATED FACTOR MATRIX

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
VAR250	0.16788	-0.03580	0.18910	0.39720
VAR251	0.47032	0.09303	0.12272	0.09984
VAR252	0.20375	0.33751	-0.02651	0.35996
VAR253	0.35072	0.02765	0.10509	0.19005
VAR254	0.30129	0.34762	0.11037	0.15130
VAR255	0.00319	0.27762	0.02304	0.41877
VAR256	0.68327	0.16626	0.03996	0.09423
VAR257	0.17975	0.57254	0.00375	0.16670
VAR258	0.39903	0.27400	0.29302	0.05919
VAR259	0.43870	0.28109	0.27182	0.06049
VAR260	0.06981	0.68967	0.05943	-0.00661
VAR261	0.46469	0.18943	0.24992	0.04788
VAR262	0.24684	-0.00714	0.59335	0.16137
VAR263	0.08603	0.26770	0.15446	0.20589
VAR264	0.65361	0.08440	0.16003	-0.07815
VAR265	0.33848	0.24754	0.33851	0.05261

TRANSFORMATION MATRIX

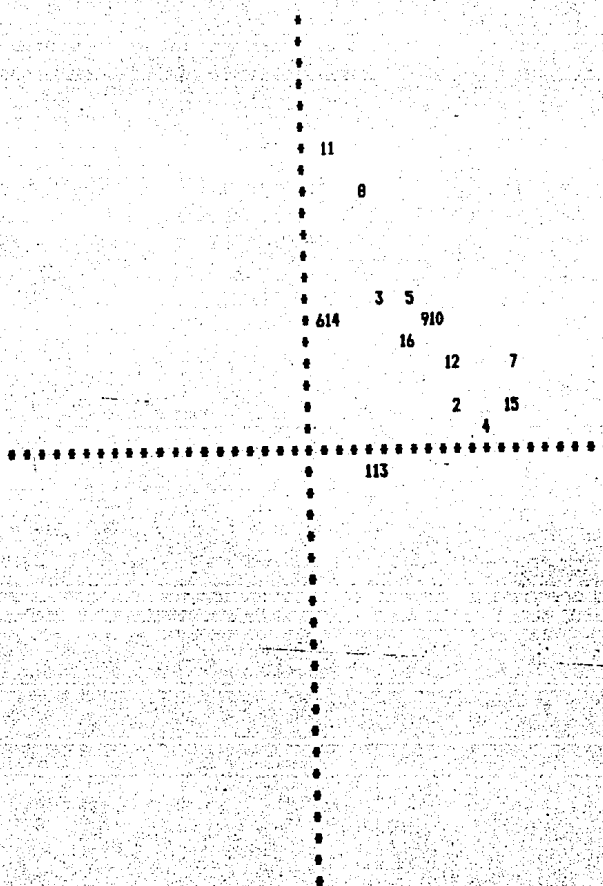
	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
FACTOR 1	0.73551	0.47417	0.33520	0.27965
FACTOR 2	-0.46807	0.82758	-0.25160	0.18887
FACTOR 3	-0.39424	-0.20447	0.56232	0.67754
FACTOR 4	0.25388	-0.22013	-0.70308	0.63444

FACTOR SCORE COEFFICIENTS

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
VAR250	-0.00480	-0.08358	0.06990	0.28402
VAR251	0.14047	-0.03037	-0.00901	0.01567
VAR252	0.01428	0.09862	-0.10642	0.24476
VAR253	0.19937	-0.08355	-0.05675	0.09728
VAR254	0.04281	0.10982	0.00144	0.04305
VAR255	-0.08477	0.08406	-0.02198	0.31056
VAR256	0.33882	-0.01566	-0.17665	-0.02406
VAR257	-0.02077	0.28598	-0.07848	0.04254
VAR258	0.07232	0.07415	0.13436	-0.05451
VAR259	0.08744	0.07033	0.11711	-0.05212
VAR260	-0.08660	0.45423	0.00858	-0.13830
VAR261	0.12139	0.02503	0.08454	-0.05661
VAR262	-0.04387	-0.09506	0.49761	0.08010
VAR263	-0.04536	0.07876	0.06030	0.10247
VAR264	0.27578	-0.06355	-0.01429	-0.03305
VAR265	0.02926	0.06954	0.17567	-0.04937

HORIZONTAL FACTOR 1    VERTICAL FACTOR 2

1 = VNR250    2 = VNR251  
3 = VNR252    4 = VNR253  
5 = VNR254    6 = VNR255  
7 = VNR256    8 = VNR257  
9 = VNR258    10 = VNR259  
11 = VNR260    12 = VNR261  
13 = VNR262    14 = VNR263  
15 = VNR264    16 = VNR265



```

1
Sintaxis:  CANCORR      16
                   VARIABLES= <lista de variables>/
                   RELATE= ((parámetros)) <conjunto variables 1>
                   WITH <conjunto variables 2>/
                   CANVAR= <prvm>/

```

El subprograma CANCORR ejecuta un análisis de correlación canónica en los datos. La entrada a este programa puede ser un archivo de datos o una matriz de correlación. El subprograma automáticamente imprime: (1) las correlaciones canónicas, (2) pruebas de significancia estadística para las correlaciones, y (3) los coeficientes de variación canónicos.

Pueden darse parámetros opcionales que controlan el número de coeficientes de variación canónicos a extraer y el nivel de significancia de rechazo para correlaciones canónicas.

El análisis de correlación canónico es un miembro de la familia de la estadística lineal multivariada, por lo que comparte muchas de sus propiedades: supone que las variables son intervalares, que las relaciones entre las variables es lineal y tener tanta varianza en ciertas variables como sea posible.

El análisis de correlación canónico toma como entrada básica dos conjuntos de variables, cada uno de los cuales tienen un significado técnico como conjunto. La estrategia básica es obtener una combinación lineal de cada uno de los conjuntos de variables de tal manera que la correlación entre las dos sea maximizada. Se pueden obtener muchas de esas parejas de combinaciones lineales. Estos coef. de variación canónicos, como se denominan, son esencialmente equivalentes a los componentes principales, con la excepción de que el criterio para su selección ha sido altera-

do. Mientras que ambas técnicas producen combinaciones lineales de las variables originales, el análisis de correlación canónico no lo hace con el objeto de acumular tanta varianza como sea posible dentro de un conjunto de variables, sino en acumular la máxima relación entre dos conjuntos de variables.

La tarjeta de procedimiento CANCERR consiste en 3 segmentos básicos denotados por las palabras VARIABLES, RELATE y CANVAR.

Mediante VARIABLES se especifican todas las variables a entrar en el análisis de correlación canónica. Le sigue RELATE, que informa qué conjuntos de variables serán correlacionados, separándolos por la palabra WITH. El orden en que se encuentren las variables en cada conjunto no afecta los resultados.

La especificación básica de RELATE puede completarse con 3 parámetros opcionales que pueden darse entre paréntesis inmediatamente después de RELATE. El formato general de la lista de parámetros es '(n,F,t)', donde 'n' es el número de parejas de coeficientes canónicos a calcularse, 'F' es el nivel de significancia mínimo para seguir calculando coeficientes canónicos, y 't' es el nivel de tolerancia usado para la detección de problemas en la matriz de correlación como singularidad y definición-no-positiva. La especificación default es (5, .05, .00001).

El tercero y último segmento de la tarjeta es opcional y se denota por la palabra CANVAR, que indica el cálculo e impresión en un archivo definido por la instrucción RAW OUTPUT UNIT, de los puntajes de coeficientes canónicos para cada caso en el análisis.

Se puede definir más de un análisis en una tarjeta CANCERR. Es necesario simplemente que a los primeros parámetros VARIABLES, RELATE y CANVAR, le sigan tantas especificaciones RELATE y CANVAR

como se requieran.

La especificación CANVAR funciona igual que la especificación FACSCORE del subprograma FACTOR. Puede darse únicamente la palabra CANVAR o, además un número entre 0 y 1 que indique la proporción de reemplazo de valores missing (prvm), es decir, los puntajes de coeficientes canónicos se obtienen reemplazando los valores missing por la media de la variable (opción default) o por un producto ponderado de los datos existentes (opción 10).

Así como en el análisis factorial o regresión múltiple, el primer paso en el análisis de correlación canónica es crear una matriz de correlación de las variables dentro del análisis. Puede ahorrarse considerable tiempo de máquina dando directamente esa matriz a CANGCORR, especialmente si se espera ejecutar varios análisis en los mismos datos. La matriz deberá seguir el formato BF10.7, iniciando cada renglón de la matriz en un nuevo registro.

Para indicar que se dará como entrada una matriz de correlación se debe especificar la opción 4. Hay dos métodos para definir la estructura de la matriz. Bajo el método default, el número y orden de las variables en la matriz debe seguir el número y orden en la lista VARIABLES de CANGCORR. Alternativamente, especificando la opción 9, la definición se hace por el número y orden de las variables en la instrucción VARIABLE LIST. Independientemente al método elegido, al dar una matriz como entrada deben prepararse adicionalmente las sig. instrucciones:

1.- VARIABLE LIST.

2.- INPUT MEDIUM.

3.- N OF CASES. Que se utilizará sólo para pruebas de significancia y no tendrá nada que ver con la lectura de la matriz.

4.- READ MATRIX. Debe seguir a OPTIONS o STATISTICS y debe preceder a la matriz si se encuentra en tarjetas. Tiene la misma finalidad que READ INPUT DATA.

El parámetro CANVAR no puede usarse al dar de entrada una matriz ya que los puntajes sólo pueden calcularse al leer el archivo de casos.

Las opciones disponibles para esta subrutina son:

- 1.- Entrarán al análisis todos los casos, incluso los missing.
- 2.- Eliminación PAIRWISE de valores missing. Un caso se omitirá de los cálculos de un coeficiente dado, si el valor de alguna de las variables consideradas es missing. La eliminación PAIRWISE tiene la ventaja de utilizar todos los datos posibles en el cálculo de cada coeficiente. Tiene la desventaja de, en algunas circunstancias, producir correlaciones basadas en un número diferente de casos y quizás hasta en subpoblaciones completamente diferentes.

Si no se especifica la opción 1 ó 2, se tendrá una eliminación LISTWISE de valores missing. Un caso no se incluirá en el cálculo de los coeficientes si contiene un valor missing en cualquier variable dada en VARIABLES. En general, la eliminación LISTWISE reduce el número de casos sobre los cuales se ejecuta el análisis factorial. Por otro lado, se asegura que el análisis se ha obtenido del mismo conjunto de casos.

- 3.- Mediante esta opción se obtiene una matriz de covarianza en lugar de una matriz de correlación cuando todas las variables en el análisis tienen varianzas = 1, ya que una matriz de covarianza es más fácil de calcular que una matriz de correlación, y además es equivalente a ella.
- 4.- Informa a CANCERR que será leída una matriz de correlación en vez de un archivo de datos.
- 5.- Indica que los puntajes de variables de casos individuales no se ajustarán por las medias de las variables en el cálculo de la matriz de correlación.
- 6.- Si cualquier conjunto de variables es linealmente dependiente, continuarán los cálculos. Un conjunto de variables es considerado lin. dep. si cualquier variable en este puede expresarse como una comb. lineal de alguna o todas las demás. Como consecuencia de ello, la matriz de correlación no puede invertirse y el análisis canónico es imposible. Como la var. dep. es redundante, se puede excluir de la matriz de correlación. Esta opción fuerza los coeficientes de esa variable a cero. La diferencia entre esto y excluir la variable es que no se reducen los grados de libertad para la Ji-cuadrada.

- 7.- Continúa los cálculos aun y cuando la matriz no sea definida positiva. Las matrices calculadas usando la opción 2, pueden tener esta condición, lo que es una indicación de que se ha basado en casos diferentes e inconsistentes. La opción default detiene el proceso si se detecta esta condición. Esta opción fuerza producir una solución al problema, el cual sólo es una aproximación al planteado por el investigador.
- 8.- Imprime la matriz de correlación de todas las variables en CANCERR en el dispositivo indicado por el usuario. Ver la instrucción RAW OUTPUT UNIT.
- 9.- El orden y número de las variables en la matriz de correlación de entrada la especifica la instrucción VARIABLE LIST. Esta opción sólo puede usarse junto con la opción 4.
- 10.- Si se especifica reemplazo de valores missing en el cálculo de los puntajes de los coeficientes canónicos (CANVAR=n), esta opción hace que el reemplazo sea por puntajes ponderados en la proporción del número total de variables en el coeficiente de variación por el número de variables no-missing. Por default, los valores missing son estimados usando la media de la variable. Esta opción sólo es aplicable junto con la opción 2.

Las estadísticas disponibles para CANCERR son:

- 1.- Medias y desviaciones estándar de las variables en VARIABLES, junto con el número de casos válidos observados.
- 2.- Matriz de correlación para las variables en el análisis.
- 3.- CANCERR imprimirá la matriz de correlación sólo si contiene coeficientes no calculables, que indica como .99. Si se solicitan las estadísticas 2 y 3, la 2 tendrá precedencia.
- 4.- Si fueran solicitados los puntajes de variación canónicos para cada caso procesado por CANCERR, serán impresos en el dispositivo indicado por el usuario. Ver la instrucción RAW OUTPUT UNIT.

COMMENT EJEMPLO DE RESULTADOS DE CANCORR.

VAR LABELS VAR017 TIENE REGADERA/  
 VAR018 TIENE CALENTADOR AGUA-LEÑA/  
 VAR019 TIENE CALENTADOR AGUA-GAS/  
 VAR020 TIENE ESTUFA DE GAS/  
 VAR021 TIENE ESTUFA DE PETROLEO/  
 VAR022 TIENE CONEDOR DE FORMICA O MUEBLERIA/  
 VAR023 TIENE CONEDOR DE HECHIZO-RUSTICO/  
 VAR024 TIENE JUEGO DE SALA/  
 VAR025 TIENE CONAL-ANAFRE-BRACERO/  
 VAR026 TIENE DRENAJE/ VAR027 TIENE FOSA SEPTICA/  
 VAR028 TIENE TONA DE AGUA/ VAR029 TIENE POZO-MORTA/  
 VAR030 TIENE RIO/  
 VAR031 TIENE LUZ ELECTRICA EN CASA/  
 VAR032 TIENE REFRIGERADOR/  
 VAR033 TIENE T.V. BLANCO-NEGRO CUANTAS/  
 VAR034 TIENE T.V. COLOR CUANTAS/  
 VAR035 TIENE LICUADORA/  
 VAR036 TIENE PLANCHA ELECTRICA/  
 VAR037 TIENE VENTILADOR/ VAR038 TIENE RADIO/  
 VAR039 TIENE TOCADISCOS/ VAR040 TIENE GRABADORA/  
 VAR041 TIENE VIDEOCASSETERA/  
 VAR042 TIENE AUTO-CAMIONETA CUANTAS/  
 VAR043 TIENE BICICLETA CUANTAS/  
 VAR044 TIENE MOTOCICLETAS CUANTAS/

VALUE LABELS VAR017 TO VAR044 (1) SI (0) NO/  
 SEX0 (0)HOMBRES (1)MUJERES/  
 ESC0 (1)SIN INSTRUCCION (2)PRIM INCOMP  
 (3)PRIM COMP (4)SEGUND O MAS/  
 EMB0 (1)15-19 (2)20-24 (3)25-29 (4)30-34  
 (5)35-39 (6)40-44 (7)45 O MAS/  
 EROCVIL (1)ACT UNIDO (0)NO UNIDO/  
 HIJOS (1) SI (0)NO/

CANCORR VARIABLES=SEX0 ESC0 EMB0 EROCVIL HIJOS VAR017 TO VAR044/  
 RELATE= SEX0 TO HIJOS WITH VAR017 TO VAR044

----- CANONICAL CORRELATION ----- RELATE LIST 1

NUMBER	EIGENVALUE	CANONICAL CORRELATION	WILK'S LAMBDA	CHI-SQUARE	D.F.	SIGNIFICANCE
1	0.21938	0.46838	0.70956	958.98608	140	0.000
2	0.03907	0.19765	0.90877	266.75171	108	0.000
3	0.02840	0.16249	0.94093	155.37222	78	0.000
4	0.01697	0.13028	0.97158	80.58906	50	0.004
5	0.01165	0.10791	0.98835	32.73882	24	0.110



COEFFICIENTS FOR CANONICAL VARIABLES OF THE SECOND SET

	CANVAR 1	CANVAR 2	CANVAR 3	CANVAR 4
VAR017	0.15680	-0.15914	-0.16577	-0.26124
VAR018	-0.03949	0.10443	-0.15742	0.02410
VAR019	0.00630	0.24518	0.21598	-0.11554
VAR020	0.17736	0.23666	0.14337	-0.05500
VAR021	-0.02591	-0.18725	-0.39906	0.28502
VAR022	0.06723	0.23611	0.17900	0.34255
VAR023	-0.01793	0.01362	-0.12345	-0.34497
VAR024	0.13185	0.14239	-0.13176	-0.35067
VAR025	-0.04408	-0.04053	-0.28750	-0.26103
VAR026	0.11174	0.01880	-0.23849	0.03315
VAR027	0.08213	0.16530	-0.11169	0.09659
VAR028	0.11639	-0.03943	-0.10166	-0.04525
VAR029	0.02032	-0.02259	-0.00747	0.13739
VAR030	-0.02672	-0.08740	0.10325	-0.17360
VAR031	-0.00621	-0.14484	0.03399	-0.15945
VAR032	0.09130	0.31774	-0.35020	-0.10539
VAR033	-0.00373	0.05796	-0.05102	-0.07049
VAR034	0.08859	-0.01964	0.00333	0.16073
VAR035	0.24548	-0.18359	0.08108	0.12044
VAR036	0.20036	0.15382	-0.28218	0.27162
VAR037	-0.06388	-0.09433	0.52116	0.01183
VAR038	0.08487	-0.26331	0.46697	-0.19795
VAR039	0.09082	-0.13931	0.04037	0.02902
VAR040	0.18146	-0.62619	-0.08402	0.22135
VAR041	-0.00795	-0.05322	-0.13947	-0.22036
VAR042	0.02347	0.11258	-0.02541	-0.18780
VAR043	0.03598	-0.22340	-0.11935	-0.01688
VAR044	0.02903	-0.21771	0.15305	-0.23382

COEFFICIENTS FOR CANONICAL VARIABLES OF THE FIRST SET

	CANVAR 1	CANVAR 2	CANVAR 3	CANVAR 4
SEED	-0.01752	0.56371	-0.67883	0.48561
ESD	1.04110	0.38077	0.10860	-0.11904
EMD	0.20828	0.45073	-0.51943	-0.92357
EDCIVIL	0.08328	0.47573	0.17864	0.44367
HTJG	-0.15459	0.12428	0.82904	-0.11832

## 13.17)

## GUTTMAN SCALE.

1 16  
 Sintaxis: GUTTMAN SCALE <nombre de la escala> =  
 <variable> (<puntos de división>) ...  
 <variable> (<puntos de división>) /  
 <nombre de la escala> ...

La escala de Guttman conocida también como Escalograma tiene como objetivo definir qué es lo que está midiendo la escala (como problema unidimensional). Por el tipo especial de tratamiento al que se somete la escala, se busca la eliminación de factores extraños a la dimensión que se pretende medir.

La escala también debe ser acumulativa, propiedad que diferencia a la escala de Guttman de casi todas las demás escalas e índices. Operacionalmente, una escala acumulativa implica que los enunciados que la componen puedan ordenarse por grado de dificultad y que la respuesta positiva a un enunciado o ítem difícil supone respuestas también positivas a ítems menos difíciles.

Cada ítem en la escala de Guttman debe ser ordinal al grado de tener la capacidad de estar dividido en algún punto en dos porciones -pasar o no pasar. Las observaciones con un valor mayor o igual que el punto de corte seleccionado se consideran haber 'pasado' un ítem; los valores menores que el punto de corte se consideran 'no haber pasado'. Pasar o aceptar un ítem se indicará como 1, mientras que reprobar o rechazar se indicará como 0.

Los enunciados deben poder ordenarse del más difícil al menos difícil. Usualmente, el ordenamiento se obtiene al sortear los enunciados en orden descendente de acuerdo a la proporción de respondientes que pasan o no. Alternativamente, el investigador puede personalmente decidir el orden.

La lógica general del análisis de Guttman, espera que las respuestas de los enunciados (por ejemplo 5), sigan el patrón ilustrado en la siguiente tabla:

TIPO DE ESCALA	ITEMS				
	A	B	C	D	E
5	1	1	1	1	1
4	0	1	1	1	1
3	0	0	1	1	1
2	0	0	0	1	1
1	0	0	0	0	1
0	0	0	0	0	0

Si los enunciados bajo análisis forman una escala perfecta, formarán un patrón como el anterior. Pero los datos raramente se ajustan así a las expectativas del investigador, debido principalmente a las actitudes de los entrevistados. Cada desviación del patrón esperado se cuenta como un error. Los errores son acumulados y un número de coeficientes estandarizados se obtienen a partir de ellos, los que ayudan a determinar si los enunciados realmente forman una escala unidimensional y acumulativa.

El subprograma GUTTMAN SCALE permite seleccionar entre 5 estadísticas, las cuales ayudan a evaluar la escalabilidad de los enunciados. Los Coeficientes de Correlación entre enunciados. El Coeficiente de Reproducibilidad, medida de la coherencia en la pauta de respuesta del individuo. Varía de 0 a 1, y debe ser mayor a .90, para considerar una escala válida. La Reproductividad Marginal Mínima, constituye el coeficiente de reproductibilidad mínimo ocurrido para la escala. La diferencia entre las dos estadísticas anteriores es el Porcentaje de Perfeccionamiento. La última medida se obtiene dividiendo la estadística anterior entre 1 menos la estadística 3, y se denomina Coeficiente de Escalabi-

lidad, que también varía entre 0 y 1, y debe ser mayor que .60, si la escala es verdaderamente unidimensional y acumulativa.

Una vez que los datos han sido procesados por GUTTMAN SCALE, el siguiente paso será construir la escala. A cada individuo se le puede asignar un puntaje basado en el número de enunciados que ha 'pasado'. Esto se hace creando una variable nueva por medio de las instrucciones COMPUTE e IF de SPSS. Por ejemplo:

```
COMPUTE   GUTSCAL = 0
IF        (ITEM1 GE 2) GUTSCAL = 1
IF        (ITEM2 GE 5) GUTSCAL = GUTSCAL + 1
IF        (ITEM3 GE 3) GUTSCAL = GUTSCAL + 1
```

La constante es el punto de división. El resultado de estas instrucciones será un valor entre 0 y 3. Si no 'pasa' ninguno de los 3 ítems, tendrá un puntaje de 0. Si los ítems tuvieran valores de 0 y 1, se obtendría el mismo resultado sumando simplemente los datos por medio de:

```
COMPUTE   GUTSCAL = ITEM1+ITEM2+ITEM3
```

Se activa la subrutina colocando GUTTMAN SCALE en la zona de control de una tarjeta, seguida por el nombre dado a la primera escala. El nombre de la escala sólo sirve para identificarla en la impresión. Puede seleccionarse cualquier nombre que no exceda 8 caracteres e inicie con letra. A continuación hasta 12 nombres de variables y sus puntos de división asociados. A cada variable en la lista le sigue uno o hasta 3 puntos de división, separados por comas y entre paréntesis.

Pueden definirse hasta 50 escalas individuales, separadas únicamente por un slash (/).

Si se indica que el subprograma ordene las variables (opción default) el orden al dar las variables no es importante. Sin ea-

bargo, si se selecciona la opción 3 (el usuario elige el orden de los enunciados), el orden de las variables debe ser el deseado. La primera variable a la izquierda será considerada la más difícil, y la más a la derecha la menos difícil.

El punto de división o de corte es el que determina si un respondente pasa o no, un ítem. Pueden darse números enteros o decimales como puntos de división. Cuando uno o más enunciados tienen puntos múltiples de corte, se producirá una escala para cada posible combinación de puntos, pero se cuenta sólo como una escala en términos de la limitación de 50 escalas.

Las opciones disponibles para esta subrutina son:

- 1.- Todos los datos, incluso los missing, entrarán a todas las escalas y cálculos.
- 2.- En la impresión, se omitirán las etiquetas de las variables.
- 3.- El orden de los ítems se tomará de la tarjeta GUTTMAN SCALE.

Si se omite la tarjeta OPTIONS, los valores missing serán excluidos, las etiquetas de las variables se imprimirán y los enunciados serán ordenados según su grado de dificultad de acuerdo a la proporción de respondentes que pasan o no, un ítem.

Las estadísticas disponibles para GUTTMAN SCALE son:

- 1.- Los coeficientes de correlación entre enunciado y de cada enunciado contra el resto de la escala.
- 2.- El coeficiente de reproductibilidad.
- 3.- La reproductibilidad marginal mínima.
- 4.- El porcentaje de perfeccionamiento.
- 5.- El coeficiente de escalabilidad.

RUN NAME EJEMPLO DE RESULTADOS DE GUTTMAN SCALE.  
 VAR LABELS V17 TIENE REBADEÑA/  
 V18 TIENE CALENTADOR/  
 V19 TIENE ESTUFA/  
 V20 TIENE COPEADOR/  
 V21 TIENE JUEGO DE SALA/  
 V22 TIENE LUZ ELECTRICA EN CASA/  
 V23 TIENE REFRIGERADOR/  
 V24 TIENE T.V./  
 V25 TIENE RADIO-TOCADISCOS-GRAMADISCOS/  
 V26 TIENE VIDEOCASSETERA/  
 V27 TIENE AUTO-BICI-MOTO/  
 VALUE LABELS V17 TO V27 (1) = SI (0) = NO/

GUTTMAN SCALE ESCALA= V17(1) V18(1) V23(1) V24(1) V21(1)  
 STATISTICS ALL

\*\*\*\*\* GUTTMAN SCALE (ESCALA ) USING \*\*\*\*\*

V17 TIENE REBADEÑA DIVISION POINT = 1.00  
 V18 TIENE CALENTADOR DIVISION POINT = 1.00  
 V23 TIENE REFRIGERADOR DIVISION POINT = 1.00  
 V24 TIENE T.V. DIVISION POINT = 1.00  
 V21 TIENE JUEGO DE SALA DIVISION POINT = 1.00

\*\*\*\*\* RESP = 1 FOR VALUES EQUAL TO DIVISION POINT AND ABOVE \*\*\*\*\*

ITER.	V18	V17	V21	V23	V24	TOTAL
RESP..	0 1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1
	I-ERR	I-ERR	I-ERR	I-ERR	I-ERR	I
E	1	1	1	1	1	1
S	5 1 0	5 6 1 0	5 6 1 0	5 6 1 0	5 6 1 0	5 6 1 0
C	1	ERR	1	1	1	1
A	1	1	1	1	1	1
L	4 1 29	30 1 5	5 4 1 20	39 1 5	5 4 1 0	59 1 39
A	1	1	ERR	1	1	1
	1	1	1	1	1	1
3	1 92	25 1 81	3 6 1 3 6	8 1 1 2 9	9 3 1 1	11 6 1 117
	1	1	1	ERR	1	1
	1	1	1	1	1	1
2	1 25 4	2 4 1 25 4	2 4 1 22 3	5 5 1 8 9	10 9 1 4	2 6 4 1 27 6
	1	1	1	1	ERR	1
	1	1	1	1	1	1
1	1 4 3 9	9 1 4 4 4	4 1 4 1 9	2 9 1 4 1 3	3 5 1 7 7	3 7 1 1 4 4 8
	1	1	1	1	1	ERR
	1	1	1	1	1	1
0	1 5 0 3	0 1 5 0 3	0 1 5 0 3	0 1 5 0 3	0 1 5 0 3	0 1 5 0 3
	1	1	1	1	1	1
SUMS	1317	144 1287	174 1201	260 1034	427 395 866	1461
PCTS	90	10 88	12 82	18 71	29 41 59	
ERRORS	0	88 5	64 56	84 118	35 92 0	542

1461 CASES WERE PROCESSED  
 0 (OR 0.0 PCT) WERE MISSING

STATISTICS..

COEFFICIENT OF REPRODUCIBILITY = 0.9258  
 MINIMUM MARGINAL REPRODUCIBILITY = 0.7810  
 PERCENT IMPROVEMENT = 0.1448  
 COEFFICIENT OF SCALABILITY = 0.6613

YULE'S Q..

	V17	V18	V23	V24	V21
V17	1.0000	0.9535	0.7596	0.8358	0.7935
V18	0.9535	1.0000	0.6186	0.7599	0.6975
V23	0.7596	0.6186	1.0000	0.8219	0.7481
V24	0.8358	0.7599	0.8219	1.0000	0.6601
V21	0.7935	0.6975	0.7481	0.6601	1.0000

BISERIAL CORR

SCALE-ITEM	0.8559	0.7344	0.4489	0.3066	0.6411
------------	--------	--------	--------	--------	--------

#### XIV. VERSIONES 7 A 9 DE SPSS.

La versión de SPSS anteriormente documentada es la 6, y desde entonces, SPSS se ha seguido desarrollando y refinando. En las versiones 7 y 8 se han agregado 4 nuevos procedimientos estadísticos y un generador de reportes. La versión 9 contiene 3 nuevos procedimientos estadísticos y una opción para gráficas.

Algunas opciones ya existentes han sido completamente o en parte reescritas, y la mayoría han sido modificadas para ofrecer nuevas facilidades y mayor eficiencia.

Los procedimientos agregados en las versiones 7 y 8 son:

REPORT - Programa que cuenta con un formato flexible y la opción de etiquetamiento de listados de casos y estadísticas.

SURVIVAL - Tablas de vida, gráficas de funciones de supervivencia y comparación de funciones entre grupos.

NPAR TESTS - 14 pruebas no paramétricas.

MULT RESPONSE - Distribución de frecuencias y tablas cruzadas de variables de respuesta múltiple.

RELIABILITY - Análisis de escalas aditivas así como medidas repetitivas, factorial ANOVA de dos vías.

Procedimientos agregados en la versión 9:

MANOVA - Procedimientos de modelos lineales en general.

BOX JENKINS - Identificación, estimación y predicción para series de tiempo univariadas.

NEW REGRESSION - Análisis de regresión múltiple con amplias facilidades para análisis de residuales.



#### 14.1)

#### REPORT.

REPORT es un flexible formateador de información que puede ser usado para generar una amplia variedad de reportes. Calcula estadísticas univariadas disponibles en CONDESCRIPTIVE, frecuencias absolutas y relativas disponibles en FREQUENCIES, medias poblacionales y estadísticas disponibles en AGGREGATE y BREAKDOWN, así como estadísticas no directamente disponibles en algún otro procedimiento de SPSS. También puede ser usado para listar los contenidos de casos individuales como en LIST CASES, pero con un control sobre su formato. La impresión se realiza en un formato que el usuario puede controlar: dimensiones, ancho de columnas, espaciamento, títulos, pies de página, etiquetas, etc.

#### 14.2)

#### SURVIVAL.

El procedimiento SURVIVAL de SPSS produce tablas de vida, gráficas de funciones SURVIVAL y comparación del procedimiento entre varios subgrupos de la muestra.

El análisis SURVIVAL calcula el tiempo de intervalo entre dos eventos, un evento inicial y uno terminal. Comúnmente es usado para caracterizar los tiempos de supervivencia de pacientes con enfermedades graves estudiando los efectos de diferentes tratamientos. También puede aplicarse a otras áreas de investigación donde los eventos inicial y terminal pueden definirse y donde el intervalo de tiempo entre estos eventos es de interés.

La selección de los eventos inicial y terminal es una parte importante del análisis. Las únicas restricciones para su definición son:

- a) El evento inicial debe ocurrir.
- b) El evento terminal no debe ocurrir antes del inicial.
- c) Debe definirse el evento terminal de manera que sea imposible que ocurra más de una vez.
- d) No es necesario que el evento terminal suceda para todos los casos.

### 14.3) NPAR TESTS.

Muchos procedimientos estadísticos requieren supuestos acerca de las distribuciones de las variables. Por ejemplo, la prueba-T, que es usada para probar si dos muestras tienen la misma media, supone que ambas muestras provienen de una población con distribución normal y con la misma covarianza. Los procedimientos que hacen suposiciones mínimas acerca de las distribuciones son llamadas 'no-paramétricas'. Las pruebas disponibles con NPAR TESTS y sus aplicaciones, son las siguientes:

TIPO DE DATOS	ESCALA NOMINAL	ESCALA ORDINAL
1 muestra	Ji-cuadrada Corridas	Kolmogorov-Smirnov
2 muestras relacionadas	Mc Nemar	Signo Wilcoxon
k muestras relacionadas	Cochran Q	Friedman Anova de dos vías.
2 muestras independientes	(Ji-cuadrada) (Fisher)	Mediana Mann-Whitney Kolmogorov-Smirnov Wald-Wolfowitz Moses
k muestras independientes	(Ji-cuadrada)	Mediana Kruskal-Wallis

NOTA: Las pruebas entre paréntesis no son parte de NPAR TESTS pero están disponibles en otros subprogramas de SPSS.

Provee de un mecanismo para el análisis de enunciados de respuesta múltiple. Típicamente, un ítem de respuesta múltiple es una pregunta a la cual el respondente puede dar más de una respuesta, y son usualmente codificadas en una de dos formas (ignorando la posibilidad de múltiples perforaciones en una misma columna, práctica no apoyada en SPSS):

- a) Dado un número moderado de respuestas posibles, por ejemplo 10, el usuario puede crear un número igual de variables dicotómicas, donde cada una represente una de las posibles respuestas.
- b) Si el rango de respuestas posibles es grande, la opción de crear variables dicotómicas llega a ser impracticable. Así, el usuario almacena cierto número de variables a las primeras 'n' respuestas dadas por el respondente. En este caso, cada variable es codificada desde 1 hasta 'm', donde 'm' es el número de respuestas únicas. Claro que habrá valores missing también codificados, particularmente para cubrir el caso de menos respuestas que las permitidas.

Cada una de estas opciones tiene sus ventajas y limitaciones. El uso de variables dicotómicas, las cuales llamaremos 'múltiples dicotómicas', permiten su uso como variables de control para tablas cruzadas y variables de selección de casos. El uso de la segunda opción que llamaremos 'respuesta múltiple', preserva el orden de las respuestas y permite ciertos análisis para la respuesta más importante.

El procedimiento MULT RESPONSE de SPSS, proporciona un mecanismo, mediante el cual, los enunciados del tipo anteriormente descrito pueden ser analizados individualmente o en combinación con otro. El procedimiento da dos tipos de resultados: tablas de frecuencia similares a las de FREQUENCIES y tablas similares a las producidas por CROSSTABS.

#### 14.5)

#### RELIABILITY.

Provee de un medio para evaluar ítems múltiples de escalas aditivas a través del cálculo de coeficientes ampliamente reconocidos de confiabilidad. Tiene dos modelos: SPLIT HALF y ALFA. Además, RELIABILITY proporciona estadísticas básicas como medias y desviaciones estándar para cada ítem, matrices de covarianza y correlación entre enunciados, medias de escala y correlaciones entre escalas. También puede ejecutar análisis de varianza, factorial ANOVA de dos vías con una observación por celda, prueba de Tukey para aditividad, prueba T-cuadrada de Hotelling para igualdad de medias y análisis de varianza para rangos de Friedman.

RELIABILITY puede usar como archivo de datos, una matriz de correlación o una de covarianza. Si la entrada consiste de un archivo de datos, puede obtenerse una serie de estadísticas que no están disponibles si la entrada es de matrices. Por otro lado, puede ser considerable el ahorro en tiempo de proceso al trabajar con matrices.

#### 14.6)

#### MANOVA.

Programa que ejecuta análisis multivariados de varianza y covarianza, los cuales, calculan estimaciones lineales univariadas y multivariadas, así como pruebas de hipótesis para cualquier cruce y/o diseño de anidamiento con o sin variables covariantes. El usuario tiene completo control sobre las especificaciones del modelo. También cuenta con iteración entre factores y variables covariantes.

#### 14.7)

#### BOX-JENKINS.

Procedimiento usado para ajustar y predecir datos sobre series de tiempo según algunos modelos estadísticos. BOX-JENKINS modelará una variable con observaciones equi-espaciadas en tiempo y sin valores missing. Una observación en un tiempo dado es modelada como una función de sus valores pasados y/o valores presentes y pasados tomando en cuenta los impactos aleatorios ambos en fluctuaciones no estacionales y estacionales.

El modelamiento de datos de series de tiempo normalmente se hace en 3 pasos. Primero, se identifica un modelo tentativo para las series, segundo, se obtiene estimación de parámetros, y tercero, se analizan conjuntos de estadísticas de diagnóstico y gráficas. Si el modelo no es considerado aceptable, se prueba con otros modelos hasta que uno sea aceptable.

El procedimiento está diseñado para definir modelos de identificación fáciles y flexibles, estimación y predicción. Pueden incluirse estimadores paramétricos, predicción y gráficas de autocorrelación parcial.

#### 14.8)

#### NEW REGRESSION.

Incorpora desarrollos metodológicos recientes. Hay 5 métodos de construcción de ecuaciones: entrada FORWARD, eliminación BACKWARD, selección STEPWISE, entrada forzada y salida forzada. Estos métodos pueden usarse juntos para explorar varios modelos de un conjunto de variables para obtener un modelo final. Los tratamientos disponibles para valores missing son: eliminación LISTWISE, sustitución de medias e inclusión de los valores

missing. Produce además estadísticas descriptivas univariadas como la media y desviación estándar, estadísticas bivariadas como correlaciones y matrices de varianza y covarianza.

La parte del programa sobre el manejo de residuales produce gráficas CASEWISE, gráficas de probabilidades normales, histogramas y gráficas lineales (todos útiles para verificar si se cumplen los supuestos del modelo de regresión). La impresión de residuales ha mejorado agregando estadísticas que ayudan a detectar puntos que influyen, y violaciones en los supuestos del modelo. Estas estadísticas incluyen los residuales 'studentizados', la eliminación Student de residuales, la D de Cook y la distancia 'Mahalanobis'. Otras opciones incluyen generación y grabado de variables relacionadas con la regresión, matrices de entrada, selección interna para validación cruzada y regresión a través del origen.

#### 14.9) GRAFICAS.

Esta opción es muy útil en aquellos casos cuando una gráfica proporciona mayor información que una tabla de números. Consiste en tres procedimientos: PIECHART, BARCHART y LINECHART. Estas subrutinas representan la facilidad de no tener que transportar los datos a otro sistema para su graficación, ni aprender otro lenguaje, ni requerir de un dispositivo especial de graficación ya que se ejecuta en casi cualquier graficadora del mercado.

Las gráficas de pay, barras o líneas que se pueden producir permiten etiquetarse, títulos en mayúsculas y minúsculas, pies de página y comentarios. Se pueden producir en dos modos, según se quieran para pruebas y revisión o para su presentación final.

Se pueden elegir colores y/o patrones de acuerdo al tipo de dispositivos con que cuente el usuario. Múltiples gráficas de pay, barras o líneas pueden graficarse en la misma hoja bajo un título común. Utiliza del SPSS el etiquetamiento de variables y valores, declaración de valores missing, así como todas las facilidades de transformación de datos.

#### 14. 10) SELECCION Y MODIFICACION DE DATOS.

La selección y modificación de datos de SPSS ha sido mejorada de la siguiente manera:

##### 14. 10. 1) SAMPLE.

Un investigador que cuenta con un número grande de casos, puede ahorrar tiempo y costos usando una muestra aleatoria de sus datos para algunos análisis específicos.

Ahora los comandos SAMPLE y \*SAMPLE, permiten la selección de muestras aleatorias de tamaño exacto\*, si es conocido el tamaño exacto del archivo, con la siguiente sintaxis:

```
1      16  
SAMPLE (Número de casos de la muestra) FROM (Número de casos del archivo)
```

Si alguna instrucción SELECT IF precede al comando SAMPLE, el número de casos del archivo serán aquellos que hayan sido

\* Anteriormente, se daba un número 'n' que indicaba el porcentaje del archivo que iba a componer la submuestra. El SPSS tomaba cada caso y lo consideraba con el n% de probabilidad de ser seleccionado, lo que originaba que el tamaño de la submuestra no fuera necesariamente el n% del tamaño de la muestra original.

seleccionados. Instrucciones SELECT IF que se encuentren después de SAMPLE, no tienen ningún efecto sobre el tamaño del archivo.

La sintaxis de SAMPLE descrita en la sección 12.7 sigue siendo válida y sus resultados no han cambiado.

#### 14.10.2) SEED.

Ordinariamente, ni las muestras ni las funciones aleatorias son reproducibles: cierto conjunto de comandos y cierto conjunto de datos producirán diferentes muestras y diferentes valores de funciones aleatorias para cada ejecución, ya que el generador de números aleatorios, utilizado por SPSS, toma la hora de la ejecución como base.

El comando SEED permite reproducir estos procesos aleatorios. Su sintaxis es la siguiente:

```
1      16  
SEED  (un número entero grande)
```

Si se ejecuta un conjunto de comandos y datos varias veces usando el comando SEED con el mismo número entero, se producirán los mismos resultados.

Este comando generalmente se coloca antes de la instrucción SAMPLE, \*SAMPLE o de la función aleatoria utilizada. Si son usadas muestras temporales, debe colocarse la instrucción SEED especificando el mismo valor, antes de cada instrucción \*SAMPLE para que la misma muestra sea seleccionada en cada caso.

El valor especificado en el comando SEED, es usado durante la ejecución y si es usada la instrucción SAVE FILE, este valor no será grabado.



### 14.10.3) FUNCIONES: NORMAL, UNIFORM E YRMODA.

Dos nuevas funciones están permitidas con el fin de facilitar la generación de pruebas de datos. La función NORMAL generará una variable distribuida normalmente con media de cero. La desviación estándar debe ser requerida como un argumento de la función. Por ejemplo, la siguiente instrucción generará una variable 'X', la cual es una muestra aleatoria de una población con media 0 y desviación estándar de .5:

```
COMPUTE      X = NORMAL(.5)
```

La función UNIFORM creará una variable distribuida uniformemente. Requiere como argumento, el límite superior. El límite inferior será 0. Por ejemplo, el siguiente COMPUTE generará la variable 'Y', donde  $0 < Y < 100$ :

```
COMPUTE      Y = UNIFORM(100)
```

La función UNIFORM puede ser usada para obtener muestras estratificadas.

Para calcular el intervalo de tiempo entre dos fechas de calendario, la función YRMODA convierte cada fecha a un número de día. Cuenta los días a partir del primer día del calendario Gregoriano: 15 de octubre de 1582. Los siguientes días son numerados consecutivamente y usado en los cálculos, por ejemplo:

```
YRMODA (1900,1,1) = 115,860
```

La fecha se da en el sig. orden: AÑO, MES y DÍA.

## 14.11) LECTURA DE DATOS.

En la versión 8 se han realizado varios cambios en la forma que SPSS maneja la entrada de datos de la siguiente manera:

### 14.11.1) N OF CASES.

La tarjeta N OF CASES es ahora opcional. Si se omite, se supone UNKNOWN. Sólo es necesario utilizarla bajo las siguientes condiciones:

- \* Cuando la N (tamaño de archivo) es usada para calcular pruebas de significancia para procedimientos que usan lectura de matrices, las que no requieren conteo de grupos.
- \* Cuando se quiere restringir a SPSS leer los primeros N casos del archivo.

### 14.11.2) END INPUT DATA.

Es ahora posible especificar N OF CASES UNKNOWN cuando se leen los datos de un archivo de tarjetas perforadas, los cuales, se encuentran entre las tarjetas de control de SPSS. El inicio de los datos se indica con la tarjeta READ INPUT DATA, como se hacía en versiones anteriores, y el final se indica con la tarjeta END INPUT DATA colocada al final del archivo de datos y antes de la siguiente tarjeta de control de SPSS.

La tarjeta END INPUT DATA sólo debe usarse cuando se especifica INPUT MEDIUM CARD y N OF CASES UNKNOWN, en cuyo caso, se puede omitir N OF CASES.

### 14.11.3) READ INPUT DATA Y READ MATRIX.

Las tarjetas READ INPUT DATA y READ MATRIX, que se requerían para indicar al sistema dónde iniciar la lectura, son ahora requeridas sólo si INPUT MEDIUM CARD; si los datos se encuentran en disco o cinta, éstos comandos se pueden omitir.

### 14.12) SORT CASES.

La instrucción SORT CASES ha sido completamente reescrita para la versión 8, las especificaciones descritas en la sección 13.3.1 todavía permanecen con los siguientes cambios:

- \* Ya no necesita estar al final del archivo de instrucciones SPSS, antes de SAVE FILE. Ahora los archivos pueden ser sorteados antes de cualquier procedimiento.
- \* Pueden sortearse variables alfanuméricas.
- \* Hasta 64 variables pueden nombrarse como llaves de ordenamiento
- \* Puede agrupar casos y definir automáticamente subarchivos.
- \* La secuencia original de casos puede preservarse dentro de los grupos sorteados.

Ahora el comando puede aparecer donde un comando de procedimiento pueda aparecer y el archivo será sorteado en ese momento en la ejecución. Puede ser usado más de una vez durante una ejecución.

SORT CASES es una modificación permanente al archivo de datos y SAVE FILE grabará el archivo sorteado.

1	16
Sintaxis: SORT CASES	<variables de ordenamiento> (A/D) <variables de ordenamiento> ... /SUBFILES ( <variables de agrupamiento> ) = <lista de nombres de subarchivo>

`SORT CASES` puede definir nuevos subarchivos según los valores de las variables del archivo. Esas variables deben especificarse primero en  `SORT CASES` y la lista de variables de ordenamiento será seguida por  `SUBFILES`.

El nombre de los subarchivos puede omitirse, en cuyo caso, el primero será  `S1`, el segundo  `S2`, etc. Si se dan los nombres, serán aplicados en el orden dado.

Los casos pueden ordenarse dentro de cada subarchivo. Las variables de ordenamiento usadas para definir los subarchivos se listan en paréntesis después de  `SUBFILES`. Las variables usadas para especificar subarchivos deben aparecer primero en la lista de ordenamiento.

Si hay más de 100 combinaciones diferentes de valores de las variables usadas para definir los subarchivos, se definirán los primeros 99, y el resto formará el subarchivo 100. Si se omite la lista de variables que definen los subarchivos, todas las variables de ordenamiento serán usadas como variables de agrupamiento. Si el parámetro  `SUBFILE` se omite, el archivo sorteado formará un único subarchivo.

Para conservar el orden original de los casos dentro de los grupos formados por el ordenamiento, se debe agregar  `SEQNUM` al final de la lista de variables de ordenamiento, la cual, no debe ser usada para definir subarchivos. Si  `SUBFILES` se omite,  `SEQNUM` no será usada como variable de agrupamiento.

No pueden ser usadas más de 8 variables de agrupamiento; si se omite la lista, serán utilizadas las primeras 8 variables.

Ejemplo:

`SORT CASES SEXO /SUBFILES`

`SORT CASES` ordenará el archivo según los valores de la variable `SEXO`. Si sus valores son (1) `HOMBRES` y (2) `MUJERES`, formará 2 subarchivos, uno con todos los casos que tengan valor 1, denominándolo `S1`, y otro con el valor 2 que denominará `S2`. Si se quiere asignar nombres a los subarchivos se puede dar:

`SORT CASES SEXO /SUBFILES = HOMBRE MUJER`

Si a la palabra `SUBFILE` le sigue una lista de nombres, serán aplicados a los subarchivos según se vayan obteniendo. Si no se dan nombres para todos los subarchivos generados, `SORT CASES` usará sus propios nombres para nombrar a los que faltaron.

`SORT CASES NIVEL SEXO EDAD /SUBFILES (NIVEL)`

Si los valores de nivel son (1) `ALTO`, (2) `MEDIO`, (3) `BAJO`, se formarán 3 subarchivos en los cuales los casos estarán ordenados por `SEXO`, y cada grupo de `SEXO`, por `EDAD`. Se puede nombrar a los subarchivos de la siguiente manera:

`SORT CASES NIVEL SEXO EDAD /SUBFILES (NIVEL) = ALTO MEDIO BAJO`

`SORT CASES CIUDAD SEGUN /SUBFILES`

Formará un subarchivo con todos los casos que tomen cada valor de `CIUDAD`, pero dentro de cada subarchivo se conservará el orden original de casos.

## XV. VERSION X DE SPSS.

SPSSx se creó con base en 17 años de desarrollo y en la experiencia de 9 versiones anteriores de SPSS.

Los comandos de SPSS, han cambiado lo menos posible en SPSSx con el fin de conservar la compatibilidad entre trabajos previamente elaborados, los cuales, para poderse ejecutar en SPSSx, necesitarán pequeñas modificaciones. Un resumen de cambios de los principales comandos se muestra en la siguiente tabla:

COMANDO SPSS	STATUS	COMANDO SPSSx	COMANDO SPSS	STATUS	COMANDO SPSSx
COMMENT	igual		PRINT BACK	reemplazado	SET
COMPUTE	igual		PRINT FORMATS	modificado	
COUNT	igual		RAW OUTPUT UNIT	reemplazado	PROCEDURE OUTPUT
DATA LIST	modificado		READ INPUT DATA	reemplazado	REBIN DATA
DO REPEAT	igual		READ MATRIX	reemplazado	INPUT MATRIX
DOCUMENT	igual		RECORD	igual	
EDIT	igual		RUN NAME	reemplazado	TITLE
END INPUT DATA	reemplazado	END DATA	RUN SUBFILES	reemplazado	SPLIT FILE
END REPEAT	igual		SAMPLE	igual	
FILE NAME	obsoleto		SAVE FILE	reemplazado	SAVE
FINISH	igual		SEED	reemplazado	SET
GET FILE	reemplazado	SET	SELECT IF	modificado	
IF	modificado		SORT CASES	igual	
INPUT FORMAT	reemplazado	DATA LIST	STATISTICS	igual	
INPUT METHOD	reemplazado	FILE HANDLE	SUBFILE LIST	obsoleto	
LIST CASES	reemplazado	LIST	TASK NAME	reemplazado	SUBTITLE
LIST FILEINFO	reemplazado	DISPLAY	VALUE LABELS	igual	
MISSING VALUES	igual		VAR LABELS	igual	
N. OF CASES	modificado		VARIABLE LIST	reemplazado	DATA LIST
NUMBERED	igual		WEIGHT	igual	
OPTIONS	igual		WRITE CASES..	reemplazado	WRITE o EXPORT
PAGESIZE	reemplazado	SET	WRITE FILEINFO	reemplazado	EXPORT

Si un comando es obsoleto, significa que el comando ya no existe y tendrá que ser eliminado del programa, si ha sido modificado, se debe revisar la nueva sintaxis, si permanece igual, trabajará en SPSSx como lo hacía en SPSS.

Los procedimientos también han sido mejorados en SPSSx aunque esencialmente son los mismos que en SPSS:

COMANDO SPSS	STATUS	COMANDO SPSSx	COMANDO SPSS	STATUS	COMANDO SPSSx
AGGREGATE	reescrito		NEW REGRESSION	renombrado	REGRESSION
ANOVA	igual		NONPAR CORR	igual	
BARCHART	igual		NPAR TESTS	igual	
BOX-JENKINS	igual		ONEWAY	igual	
BREAKDOWN	igual		PARTIAL CORR	igual	
CANCORR	obsoleto		PEARSON CORR	igual	
CONDSCRIPTIVE	igual		PIECHART	igual	
CROSSTABS	modificado		REGRESSION	reemplazado	REGRESSION
DISCRIMINANT	modificado		RELIABILITY	igual	
FACTOR	reescrito		REPORT	modificado	
FREQUENCIES	reescrito		SCATTERGRAM	igual	
GUTTMAN SCALE	obsoleto		SURVIVAL	igual	
LINECHART	igual		T-TEST	igual	
MANOVA	igual		TRANSFORM	obsoleto	
MULT RESPONSE	igual				

Si algún procedimiento ha sido reescrito o reemplazado, se debe revisar su descripción completa en el manual de SPSSx, si un procedimiento es obsoleto, no está disponible en SPSSx, si permanece igual, trabajará en SPSSx como lo hacía en SPSS. Algunas estadísticas y opciones para procedimientos han sido alteradas, agregadas o borradas. Los nuevos procedimientos no usan los comandos STATISTICS y OPTIONS.

A continuación se mencionan los principales cambios de SPSSx con respecto a SPSS:

- \* La lista de errores que en SPSS aparecía al final de la ejecución, ha sido cambiada, y ahora los WARNINGS y mensajes de error aparecen en SPSSx cuando el error es encontrado.
- \* Los archivos creados por SPSS se pueden leer por SPSSx, si fueron creados y se van a leer en el mismo tipo de computadora. Archivos creados por SPSSx no pueden ser leídos por SPSS.
- \* En SPSS todas las variables eran impresas como numéricas enteras a menos que fuera usado el comando PRINT FORMATS para especificar si era alfanumérica o el número de decimales. SPSSx ya no requiere de este comando ya que el formato definido para leer los datos es automáticamente asignado como formato de impresión.

- \* Las transformaciones permanentes de datos en SPSSx pueden definirse en cualquier parte del programa incluso después del primer procedimiento. No hay comandos de transformación con asterisco en SPSSx. Las transformaciones temporales serán aquellas que sigan al comando TEMPORARY.
- \* Todos los comandos de SPSSx empiezan en la columna 1 de un registro y continúan por todos los que se requieran. Las líneas de continuación **deben** indentarse al menos una columna y tantas como se deseen.

Se tienen algunas nuevas funciones estadísticas y lógicas:

MOD(arg1,arg2)	Módulo (residuo) de la división del argumento 1 entre el argumento 2.
ASIN(arg)	Arcoseno (argumento en radianes).
SUM (lista arg)	La suma de los valores de las variables en la lista de argumentos.
MEAN(lista arg)	Media de la lista de argumentos.
SD(lista arg)	Desviación estándar de los argumentos.
VARIANCE(lista arg)	Varianza de la lista de argumentos.
CFVAR(lista arg)	Coficiente de variación de la lista.
MIN(lista arg)	Valor mínimo de la lista de argumentos.
MAX(lista arg)	Valor máximo de la lista de argumentos.
RANGE(arg,lista)	Tomará el valor 1 si el valor del argumento se encuentra en los rangos definidos por la lista de argumentos que se dará en pares, en caso contrario valdrá cero.
ANY(arg,lista)	Tomará el valor 1 si el argumento es igual a alguno de los argumentos en la lista, si no, valdrá cero.

Aunque SPSSx no requiere que se idente a la columna 16 como se hacía en SPSS, se puede hacer y no tendrá ningún efecto en la ejecución. Los mensajes de error están diseñados para ayudar a la traducción de SPSS a SPSSx, pero es recomendable antes de empezar, leer las secciones más importantes del manual para así aprovechar al máximo las nuevas características y no simplemente traducir un programa.



## 15. 1) NUEVAS RUTINAS EN PROXIMAS VERSIONES DE SPSS.

Constantemente se están desarrollando nuevos procedimientos.

En las versiones subsecuentes se podrá contar con los siguientes procedimientos:

- a) **TABLE BUILDER** dará a los usuarios de CROSSTABS el poder que **REPORT** le da a los usuarios de **BREAKDOWN**. **REPORT BUILDER** produce reportes con el listado de casos de un archivo sorteado, además de dar capacidades adicionales a los usuarios de **REPORT**. Ambos reportes son adecuados para la publicación de reportes.
- b) **CLEAN** y **UPDATE** permiten la verificación y actualización de datos. **CLEAN** puede especificar los valores permitidos para todas las variables y su relación entre otras. Con **UPDATE** se pueden reemplazar valores para variables y casos específicos.
- c) **USERPROC** permite al usuario escribir su propio procedimiento estadístico e incorporarlo fácilmente a SPSSx.
- d) **PROBIT** es usado para estimar los efectos de una o mas variables independientes en una variable dependiente dicotómica. El procedimiento está optimizado para análisis de dosis (medicina) y modelos relacionados; además, **PROBIT** puede estimar modelos logísticos de regresión. Los modelos estimados tienen una gran variedad de aplicaciones, por ejemplo, se puede:
  - 1) Modelar respuestas de sujetos para variar la dosis en la preparación de varias drogas.
  - 2) Analizar los índices de mortalidad de hierbas con varios niveles de diferentes soluciones de herbicidas.
  - 3) Relacionar los índices de compra de un producto para probar precios de venta.
  - 4) Predecir la participación de la fuerza laboral de un conjunto de variables demográficas.
- e) **MATRIX** provee un lenguaje completo de matrices para ejecutar una amplia variedad de manipulación de matrices y cálculos.
- f) **LOGLINEAR** hace ajustes de modelos, pruebas de hipótesis y estimación de parámetros para cualquier modelo que tenga variables categóricas como sus principales componentes. Modela tablas multivariadas de contingencia usando el modelo de respuesta multinomial. Puede ajustarse a modelos JERARQUICOS, no JERARQUICOS, CUASI-INDEPENDIENTES o con ceros estructurales, modelos LOGIT, regresión logística, etc. Trabaja con contrastes predefinidos por el usuario, produce probabilidades observadas y esperadas, residuales estandarizados y no estandarizados, residuales ajustados y una gráfica de probabilidad normal de residuales ajustados.

- g) **CLUSTER** obtiene conglomerados JERARQUICOS ya sean de pocos a un número moderado de casos. También se puede conglomerar variables usando como entrada una matriz de similaridad o distancia. CLUSTER provee 7 métodos para formar conglomerados más generales. Se pueden elegir varias medidas de disimilitud, las cuales incorporan una elección de tratamientos de valores missing.
- h) **HILOGLINEAR** estima parámetros de modelos JERARQUICOS LOGARITMICOS-LINEALES para tablas de frecuencia. Debido a que usa un algoritmo de ajuste iterativo y proporcional, HILOGLINEAR es mucho más eficiente para dichos modelos que LOGLINEAR. HILOGLINEAR también puede ejecutar automáticamente una eliminación BACKWARD de términos de estos modelos.
- i) **PLOT** produce gráficas de líneas de dos dimensiones. Se pueden solicitar gráficas simples, gráficas con una variable control y estadísticas de regresión, gráficas de entorno y gráficas sobrepuestas. PLOT cuenta con opciones de formato para controlar el tamaño de los ejes, la escala y símbolos de graficamiento. También provee gran versatilidad en gráficas para presentación y una mayor variedad de tipos de gráficas que las disponibles con SCATTERGRAM.
- j) **QUICK CLUSTER** se puede usar para agrupar eficientemente un número grande de casos dentro de un determinado número de grupos. Se necesita sólo nombrar las variables que serán usadas para agrupar los casos. Se puede elegir entre varios tratamientos para valores missing. Después de agrupar los casos, QUICK CLUSTER imprime los centros de los conglomerados y el número de casos en cada uno.
- k) **LISREL** analiza una variedad de modelos de relaciones estructuralmente lineales. Incluye el cálculo de estimadores iniciales, impresión de etiquetas para variables LATENT, optimización de tiempos de máquina y nuevos métodos para obtener estimadores de parámetros. Investigadores en Sociología, Psicología, Mercadotecnia, Econometría, Educación y otros campos usan modelos estructurales para representar relaciones entre variables observadas. Se puede usar LISREL para:
- \* Estimar modelos econométricos de ecuaciones simultáneas, teniendo variables medidas con o sin error.
  - \* Estudiar relaciones dependientes del tiempo en datos longitudinales.
  - \* Ejecutar análisis factoriales confirmatorios y análisis estructurales de covarianza.
  - \* Estima componentes de varianza en datos estructurales no balanceados o irregulares en familias u otros grupos sociales.
  - \* Prueba la igualdad de matrices de covarianza o de correlación, patrones de Regresión o Factoriales para 2 o más grupos.
  - \* Ejecuta técnicas multivariadas estándar tales como componentes principales y análisis de correlación canónica.

## XVI. VERSION MICRO DE SPSS.

SPSS/PC para la IBM/PC es un nuevo producto de SPSS INC., que provee poderosos sistemas de análisis estadísticos de información para una amplia selección de computadoras personales. Como una versión del actual software SPSS, SPSS/PC se beneficia de la confiabilidad y refinación de 10 versiones del sistema SPSS.

SPSS/PC comparte una sintaxis casi idéntica a SPSSx, formando ambos paquetes un gran equipo.

SPSS/PC es un programa diseñado para correr en una IBM/PC con al menos 320 k de memoria RAM. Si el usuario tiene instalados otros programas bajo el sistema operativo DOS, 320 k podrían ser insuficientes. Si no hay suficiente memoria disponible, SPSS/PC abortará con un mensaje. Suponiendo que hay suficiente memoria disponible, el mínimo espacio de trabajo (memoria en donde los comandos son almacenados durante el proceso y resultados intermedios) disponible debe ser de 5,000 bytes. La cantidad de espacio de trabajo requerido para ejecutar una tarea determinada varía según la tarea y generalmente no depende del número de casos del archivo. Si no hay suficiente memoria para ejecutar determinada tarea, SPSS/PC manda un mensaje de error y salta la tarea. Si se tiene 384 k de RAM, SPSS/PC incrementa el tamaño del área de trabajo a un máximo de 64,000 bytes. Memoria adicional a ésta no incrementa el área de trabajo ni disminuye el tiempo de proceso.

SPSS/PC es un programa que requiere de archivos de diez diskettes que deben ser copiados al disco duro antes de operar el sistema. SPSS/PC es un sistema iterativo basado en comandos consistentes en palabras llave y especificaciones. Para dar

correctamente un comando se debe seguir un conjunto de reglas del lenguaje SPSS/PC. Hay 3 tipos principales de comandos:

- a) Comandos de operación
- b) Comandos de definición y manipulación de datos
- c) Comandos de procedimientos

Los comandos de operación proporcionan información acerca de la forma que el sistema SPSS/PC opera. SPSS/PC ejecuta un comando de operación tan pronto como se dé. Los comandos de manipulación y definición de datos dicen cómo y dónde leer los datos, cómo calcular variables nuevas, cómo cambiar valores de variables existentes, cómo identificar valores missing, cuáles casos usar y cómo etiquetar. Tan pronto se dé el comando, SPSS/PC verifica que siga las reglas de sintaxis para el comando y no lo ejecuta hasta que se dé un comando de procedimiento.

El conjunto de comandos que SPSS/PC requiere para ejecutar un análisis o producir un reporte, puede algunas veces ser muy complejo. Al trabajar iterativamente, sólo algunos comandos son visibles a un tiempo. Se puede olvidar lo ya hecho y no se puede regresar a ver comandos ya dados. Puede que se ejecute un análisis periódicamente, en este caso, tal vez no se quiera perder tiempo en dar el mismo conjunto de comandos cada vez.

En ambas situaciones puede ser preferible crear un archivo de comandos usando un editor y dar a ejecutar el archivo a SPSS/PC usando el comando INCLUDE. SPSS/PC entonces, ejecuta los comandos en el archivo como si fueran entrando del teclado.

Los comandos en el archivo aparecerán en la pantalla como vayan siendo leídos. Una vez que SPSS/PC ha leído todos los comandos en el archivo, quedará en espera de un nuevo comando, a menos que el archivo contenga el comando FINISH.

SPSS/PC escribe 5 tipos de archivos:

- 1) LISTING FILES. Contiene la salida del monitor, puede enviarse a impresora y/o pantalla y puede grabarse en disco. El archivo de disco default para salida de monitor es SPSS.LIS.
- 2) LOG FILES. Forma un archivo con los comandos que se dieron a SPSS/PC. El archivo default es SPSS.LOG.
- 3) SYSTEM FILE. Graba los datos y un diccionario para sesiones posteriores de SPSS/PC. El archivo SYSTEM default es SPSS.SYS.
- 4) PORTABLE FILE. Es usado para transportar los datos y un diccionario a otras máquinas. No hay archivo PORTABLE default.
- 5) RESULTS FILE. Incluye los resultados de WRITE, CORRELATION, CLUSTER, FACTOR, REGRESSION y ONEWAY. El default es SPSS.PRC.

Al principio de cada sesión, los archivos SPSS.LIS, SPSS.LOG y SPSS.PRC son inicializados. Esto significa que si existían estos archivos, creados en una sesión previa, se borran, a menos que hayan sido direccionados a otro archivo diferente al default mediante el comando SET o renombrado usando el comando RENAME de DOS. SPSS.SYS se inicializa sólo si se usa el comando SAVE, en cuyo caso el archivo SPSS.SYS anterior se pierde, a menos, de nuevo, que haya sido renombrado o redireccionado.

Los procedimientos disponibles en SPSS/PC son:

Reporte de datos	- LIST, PLOT y REPORT.
Estadísticas descriptivas	- DESCRIPTIVES y FREQUENCIES.
Estadísticas categóricas	- CROSSTABS y HILOGLINEAR
Comparación de grupos	- T-TEST, ONEWAY, MEANS y ANOVA.
Estadística multivariada	- CORRELATION, CLUSTER, REGRESSION y FACTOR.
Estadística no paramétrica	- NPAR TESTS.
Utilerías	- WRITE, SORT CASES, EXPORT y SAVE.

Para estudiar en más detalle cada comando SPSS/PC, se recomienda consultar el manual para una descripción completa. En casi todos los casos, se debe pensar en los comandos y características de SPSS/PC como un subconjunto de SPSSx.

La mayoría de los comandos de SPSS/PC operan de la misma

manera que los comandos de SPSSx y tienen la misma sintaxis. Con sólo algunos cambios, se puede aplicar lo aprendido de SPSSx en SPSS/PC. A continuación se resumen las principales diferencias en la sintaxis y en la operación entre SPSS/PC y SPSSx.

La principal diferencia entre SPSS/PC y SPSSx es la diferencia básica entre un sistema iterativo y un sistema 'batch': en SPSS/PC se puede ejecutar un comando de forma inmediata, ver los resultados y dar el siguiente comando. Por esta razón, cada comando SPSS/PC debe terminar con un punto ('.').

Las ventajas de un sistema iterativo son que se pueden hacer ajustes para errores, de forma inmediata, y continuar el proceso de comandos. SPSS/PC también opera en modo 'batch' procesando un archivo de comandos como SPSSx. Por otro lado, SPSSx tiene muchas menos restricciones y más posibilidades.

COMANDO SPSS/PC	STATUS	COMANDO SPSSx	COMANDO SPSS/PC	STATUS	COMANDO SPSSx
ANDWA	igual		MEANS	modificada	BREAKDOWN
BEGIN DATA-END DATA	modificada		MISSING VALUES	modificada	
* (comentario)	modificada	COMMENT	N	modificada	N OF CASES
CLUSTER	modificada		NPAR TESTS	igual	
COMPUTE	modificada		ONEWAY	modificada	
CORRELATION	modificada	PEARSON CORR	PLOT	igual	
COUNT	igual		PROCESS IF	nueva	ninguna
CROSSTABS	modificada		RECDEBE	modificada	
DATA LIST	modificada		REGRESSION	modificada	
DESCRIPTIVES	modificada	CONDESCRIPTIVE	REPORT	igual	
DISPLAY	modificada		SMPLF	modificada	
EXPORT	igual		SAVE	modificada	
FACTOR	modificada		SELECT IF	modificada	
FINISH	igual		SET	modificada	
FORMAT	modificada		SHOW	modificada	
FREQUENCIES	modificada		SORT CASES	modificada	
GET	modificada		SUBTITLE	modificada	
HELP	nueva	ninguna	TITLE	modificada	
MILOBLINEAR	igual		T-TEST	modificada	
IF	modificada		VALUE LABELS	modificada	
IMPORT	igual		VARIABLE LABELS	modificada	
INCLUDE	nueva	ninguna	WEIGHT	modificada	
LIST	modificada		WRITE	modificada	

La tabla no lista comandos de SPSSx que no están disponibles en SPSS/PC.

En términos generales de sintaxis y operación SPSS/PC difiere de SPSSx en los siguientes puntos:

- 1) Un comando y sus especificaciones deben terminar con un punto.
- 2) Cada comando, incluso el primero, puede truncarse a los primeros tres caracteres.
- 3) Los comandos no necesitan iniciar en la columna 1 (excepto END DATA que debe empezar en la columna 1, incluir un espacio y estar escrito completo).
- 4) Las líneas de continuación de los comandos pueden iniciar en la columna 1.
- 5) Variables SHORT STRING, son de 8 caracteres o menos.
- 6) Las especificaciones para variables STRING, deben ser de la misma longitud; SPSS/PC no trunca para hacer comparaciones o asignaciones.
- 7) Se han omitido funciones STRING.
- 8) Las medidas hechas para aceptar la sintaxis de SPSS versión 9 en SPSSx, no siempre se aplica en SPSS/PC.
- 9) Pueden definirse hasta 200 variables en SPSS/PC. Las variables creadas durante la sesión, cuentan dentro del límite.
- 10) SPSS/PC nombra a los archivos en el comando apropiado más que durante el manejo del archivo.
- 11) SPSS/PC usa un diccionario de datos por sesión. Una vez que el diccionario ha sido creado, los subsecuentes comandos DATA LIST, GET o IMPORT son error.
- 12) Cuando se incluya los datos con los comandos, el primer procedimiento se coloca después de los datos. Los datos se leen y las transformaciones iniciales se ejecutan antes de que el primer procedimiento sea ejecutado.
- 13) SPSS/PC no incluye en forma completa el uso de variables lógicas. Cuando un valor diferente a 0, 1 ó missing es encontrado, no manda WARNING. En lugar de usar variables lógicas, se debe dar la expresión lógica completa (VARA EQ 1).
- 14) OPTIONS y STATISTICS en SPSS/PC son subcomandos para los procedimientos apropiados.

## CONCLUSION

- \* El sentido de "Análisis de Datos" es diferente para cada quien, dependiendo del nivel de conocimientos estadísticos. Las técnicas usadas en el análisis de datos varían desde una descripción simple de los datos en un histograma o gráfica y el cálculo de estadísticas (como la media y la desviación estándar), hasta avanzados métodos de análisis multivariado. Para algunos, el análisis de datos involucra una gráfica o un conjunto de cálculos; para otros, envuelve una secuencia de pasos, detectando la presencia de sesgos e incoherencias, asegurando las suposiciones necesarias para que el análisis se resuelva. Cada paso puede sugerir análisis posteriores. En algunos casos lo que se necesita son las llamadas Tablas de Decisión, que permiten al usuario conocer cuáles son las estadísticas más convenientes y adecuadas. A continuación se dan dos de ejemplo:

Nivel de Medición de una variable.	Procedimientos de una sola variable.	Procedimientos de dos variables (bivariados). Nivel de medición de la segunda variable.			
		Dicotómicas	Nominales	Ordinales	Intervalares
Dicotómica	Proporciones y Porcentajes.	Diferencia de proporciones, Ji-cuadrada, prueba de Fisher.			
Nominal	Proporciones y Porcentajes.	Ji-cuadrada.	Ji-cuadrada		
Ordinal	Medianas, cuartiles, deciles y desviación de cuartiles.	Mann-Whitney, prueba de corridas, Smirnov, prueba del signo.	Análisis de Varianza con rangos.	Correlaciones, tau de Kendall, gamma.	
Intervalar	Medias, medianas, desviación estándar.	Diferencia de medias	Análisis de Varianza, correlación.		Correlación y regresión.



Pruebas no paramétricas de significancia estadística.

Organización de los datos	Escala de los datos	
	Nominal	Ordinal
1 muestra	Ji-cuadrada Corridas	Kolmogorov-Smirnov
2 muestras relacionadas	Mc Nemar	Signo Wilcoxon
k muestras relacionadas	Cochran Q	Friedman Anova de dos vías
2 muestras independientes	Ji-cuadrada	Mediana Mann-Whitney Kolmogorov-Smirnov Wald-Wolfowitz Moses
k muestras independientes	Ji-cuadrada	Mediana Kruskal-Wallis

- \* El fácil manejo de los complejos procedimientos estadísticos del SPSS, permite su uso sin siquiera conocer las suposiciones de los métodos ni sus bases estadísticas o matemáticas lo que conduce, muchas veces, a un abuso; se desarrollan procedimientos, que la capacidad del SPSS hacen posibles, y el desconocimiento de métodos o supuestos estadísticos conducen a un aprovechamiento nulo de los resultados, por lo que se recomienda al usuario no usar un procedimiento estadístico a menos que se comprenda el tratamiento apropiado para el tipo de datos y el significado de las estadísticas producidas. Los errores más comunes son:

- 1) Pedir medias de variables nominales como: RELIGION, CIUDAD, RAZA, etc.
- 2) Dar la instrucción STATISTICS ALL sin seleccionar sólo las estadísticas adecuadas al nivel de medición que en ese momento se está manejando.

- 3) Dar variables nominales en las subrutinas CANCORR, FACTOR o DISCRIMINANT.
- 4) Solicitar los procedimientos BREAKDOWN, ANOVA o ONEWAY en 3 o más niveles sin verificar el tamaño de los grupos que quedan. Podría suceder que algún grupo quedara con muy pocos casos lo que invalidaría automáticamente la prueba de significación.
- 5) Algunas subrutinas requieren que las variables estén completas, es decir, que no tengan valores missing. La técnica de substituir esos valores por la media de la escala es muy discutida por algunos autores. FACTOR y REGRESSION no consideran en sus cálculos a todo el caso si alguna de las variables en el modelo tiene valor missing, y podría suceder que se iniciara el proceso con 400 casos y quedar al final con 10, lo que invalida automáticamente los resultados del procedimiento.
- 6) Verificar las correlaciones de las variables antes de entrar a REGRESSION, ya que si dos variables correlacionan muy alto entre sí, sesgan todos los parámetros estimados.
- 7) Obtener resultados redundantes al pedir todo contra todo, dado que en el SPSS es tan sencillo, en vez de generar las instrucciones que produzcan únicamente lo que se desea. Por ejemplo:

V1 TO V10 WITH V1 TO V10

que produciría 100 tablas, de las cuales sólo 45 interesan, en vez de dar:

```

V1 BY V2 TO V10/
V2 BY V3 TO V10/
V3 BY V4 TO V10/
V4 BY V5 TO V10/
V5 BY V6 TO V10/
V6 BY V7 TO V10/
V7 BY V8 TO V10/
V8 BY V9 TO V10/
V9 BY V10

```

- 8) Tener siempre en cuenta que al final la experiencia y sentido común del analista es quien dice si los resultados sirven y no la autoridad del paquete.

- \* Varios tipos de errores pueden causar que los programas se ejecuten erróneamente y puedan ocasionar que terminen; estos pueden ser errores en las tarjetas del sistema, en las tarjetas de SPSS, en el tipo de datos o, posiblemente, en el programa. Un mensaje de error es usualmente impreso para localizar su ocurrencia. Como un error puede acarrear otros, siempre es recomendable revisar el primer mensaje y corregir el error indicado. Cuando el error es en la sintaxis, se debe verificar la instrucción donde se indica y la instrucción anterior (si es necesario) para ver si se ha especificado correctamente la instrucción. Por último, recuerde las ventajas que ofrece usar la instrucción EDIT.

¿ Ha ud.?

¿ Incluido una diagonal ( '/' ) entre los párrafos ?

¿ Cerrado el mismo número de paréntesis que abrió ?

¿ Colocado comas o blancos entre listas ?

¿ Colocado apóstrofes alrededor de valores de variables alfanuméricas ?

¿ Evitado usar un nombre reservado ?

Recuerde que palabras mal escritas pueden ocasionar errores.

¿ Requiere su trabajo mayor espacio de memoria ?

¿ Está correcta la forma de leer sus datos ?

¿ Trata de leer más variables de las que tiene ?

¿ Lee caracteres ilegales (alfabéticos en lugar de numéricos) ?

¿ Está correcto el argumento de la función ?

¿ Trata de obtener el logaritmo de cero ?

¿ Trata de obtener la raíz cuadrada de un número negativo ?

¿ Trata de dividir entre cero ?

¿ Es el exponente muy grande ?

- \* Existe una versión del SPSS Conversacional, el cual no se había mencionado hasta ahora debido a que el sentimiento general de los que han trabajado con él es: que es muy lento, no cuenta con todas las ventajas y opciones de que dispone SPSS en su forma Batch (descrito en este manual), y si el usuario no tiene bien claro lo que desea obtener del paquete ni conoce las técnicas estadísticas de SPSS, le será igual trabajar en una como en otra versión de SPSS.
  
- \* Cuando uno como usuario conoce y trabaja en una versión del SPSS y desea trabajar en otra, el paso de una versión a otra es realmente sencillo, no involucra mucho tiempo de adaptación ni laboriosas traducciones de programas. Todos los manuales de SPSS incluyen un capítulo de ayuda a usuarios de versiones anteriores. Además, una característica del SPSS que vale la pena destacar, es que cuenta con mensajes sumamente claros de error, que auxilian enormemente al usuario en su diario trato con el paquete SPSS.

**APENDICE I.  
PALABRAS RESERVADAS DE SPSS.**

ABS	FACTOR	N OF CASES	SAMPLE
AGGREGATE	FILE NAME	NE	SAVE FILE
AGGSTATS	FIN	NEW REGRESSION	SCATTERGRAM
ALL	FINISH	NFACTORS	SCHEFFE
ALPHA	FIXED	NO	SD
ANALYSIS	FORMAT	NOEJECT.	SEED
AND	FOUT	NONPAR CORR	SELECT IF
ANOVA	FREEFIELD	NORMAL	SIN
ATAM	FREQUENCIES	NOROTATE	SIZE
BARCHART	FUNCTIONS	NOT	SKEW
BINARY	GE	NPAR TESTS	SNK
BLANK	GENERAL	NUMBERED	SORT CASES
BOX JENKINS	GET FILE	OBLIQUE	SORTVARS
BREAKDOWN	GROUPS	ONEWAY	SORT
BY	GROUPS	OPTIONS	STATISTICS
BYPASS	GROUPVARS	OR	STOPFACT
CANCORR	GT	OTHER	SUBDIRECTORY
CANVAR	GUTTMAN SCALE	PA1	SUBFILE LIST
CARD	HIGHEST	PA2	SUBFILES
CASES	IF	PAGESIZE	SUM
COMMENT	IMAGE	PAIRS	SURVIVAL
COMPLETE	INPUT FORMAT	PARTIAL CORR	T-TEST
COMPUTE	INPUT MEDIUM	PCTBTN	TABLES
CONDESCRIPTIVE	INTEGER	PCTGT	TAPE
CONTRAST	ITERATE	PCTLT	TASK NAME
CONTROL	KURT	PEARSON CORR	THRU
CONVERT	LABELS	PIECHART	TO
COS	LE	PIN	TOLERANCE
COUNT	LG10	POLINOMIAL	TRUNC
CROSSBREAK	LINECHART	POUT	TUKEY
CROSSTABS	LIST CASES	PRINT BACK	TUKEYB
DATA LIST	LIST FILEINFO	PRINT FORMAT	TYPE
DELTA	LN	PRIORS	UNIFORM
DIAGONAL	LOWEST	QUARTIMAX	UNKNOWN
DIRECT	LSD	RANGES	VALIDN
DISCRIMINANT	LSDMOD	RAO	VALUE LABELS
DISK	LT	RAW OUTPUT UNIT	VAR LABELS
DO REPEAT	NAHAL	READ INPUT DATA	VARIABLE LIST
DOCUMENT	MANOVA	READ MATRIX	VARIABLES
DOCUMENTS	MAX	RECODE	VARIMAX
DUNCAN	MAXINF	REGRESSION	VARINFO
EACH	MAXSTEPS	RELATE	VARLIST
EDIT	MEAN	RELIABILITY	VIN
ELSE	METHOD	REPORT	WEIGHT
END INPUT DATA	MIN	RESID	WILKS
END REPEAT	MINEIGEN	RMISS	WITH
EQ	MINRESID	RND	WRITE CASES
EQUAL	MISSING VALUES	ROTATE	WRITE FILEINFO
EQUIMAX	MOD10	RUN NAME	YES
EXP	MULT RESPONSE	RUN SUBFILES	YRMDA
FACSCORE			

## APENDICE II. OTROS PAQUETES ESTADISTICOS.

Este apéndice tiene la finalidad de mostrar a los lectores de este manual, que el SPSS no es el único paquete en el mercado, ni siquiera el único paquete estadístico. Cuando se habla de paquetes estadísticos, se pueden mencionar muchos, los hay desde los muy completos a los muy específicos, pero los más conocidos son, además del SPSS: SAS, BMDP y OSIRIS, los que se describen brevemente a continuación:

### SAS

SAS es un sistema de cómputo para el análisis de datos cuya primera versión data de 1966. Sus siglas significan "Statistical Analysis System" (Sistema de Análisis Estadístico). SAS sólo ejecuta en computadoras IBM 360/370 (y compatibles).

Es un sistema fácil de usar ya que provee todas las herramientas necesarias para el análisis de datos:

- . Almacenamiento y acceso a la información.
- . Modificación de datos y programación.
- . Impresión de reportes.
- . Análisis Estadístico.
- . Manejo de archivos.

SAS lee datos de tarjetas, disco o cinta; que luego organiza en archivos. Los datos pueden combinarse con otros archivos de SAS para analizarse estadísticamente y producir reportes de su contenido. Los archivos de SAS se documentan a sí mismos, ya que contienen tanto los datos como sus descripciones. La estructura especial de la base de datos de SAS minimiza su mantenimiento.

Cuando los valores de los datos necesitan ser modificados,

tiene disponible un conjunto completo de instrucciones y funciones. Algunas de ellas ejecutan operaciones estándar como crear variables nuevas, acumulación de totales, verificación de errores; otras son herramientas poderosas de programación como IF/THEN/ELSE y DO/END.

Los procedimientos de análisis estadístico de SAS tienen desde estadísticas descriptivas simples hasta técnicas multivariadas complejas. El usuario de SAS puede diseñar e imprimir reportes en cualquier forma. Una característica especial de SAS que vale la pena destacar es el procedimiento MATRIX que da al usuario la habilidad de manejar cualquier problema que pueda expresarse en la notación tradicional de matrices.

Las características de manejo de datos de SAS son tan valiosas que SAS es usado por muchos como un sistema de manejo de base de datos. SAS también cuenta con procedimientos de utilería para copiar archivos, revisar contenidos de cinta, mover programas de librería, etc.

Los procedimientos con los que cuenta se pueden clasificar de la siguiente manera:

- a) Estadísticas descriptivas.
- b) Análisis de diseño de experimentos.
- c) Análisis de regresión.
- d) Análisis Multivariado.
- e) Econometría y Series de tiempo.
- f) Gráficas.
- g) Especiales (MATRIX).
- h) Interfaces (BMDP).
- i) Utilerías SAS.
- j) Utilerías OS.

Para mayor información comunicarse a:

SAS Institute Inc.  
Box 8000. Cary, North Carolina 27511.  
U.S.A.

PROCEDIMIENTOS SAS:	SPSS TIENE ALGUN PROCEDIMIENTO COMPARABLE ?	PROCEDIMIENTOS SAS:	SPSS TIENE ALGUN PROCEDIMIENTO COMPARABLE ?
ANOVA	SI	ANOVA	NO
AUTOREG	NO	NEIGHBOR	SI
CANCORR	SI	NESTED	NO
CHVRT	SI	NLIN	NO
CLUSTER	SI	NPAR1WAY	SI
CORR	SI	PLAN	NO
DISCRIM	SI	PLOT	SI
DUNCAN	SI	PROBIT	SI
FACTOR	SI	RANK	SI
FREQ	SI	RSQUARE	NO
FLNCAT	NO	SCORE	NO
GLM	SI	SPECTRA	NO
GUTTMAN	SI	STEPWISE	SI
MATRIX	NO	SUMMARY	NO
MEANS	SI	SYSPRINT	NO
		TTEST	SI
		UNIVARIATE	SI
		VARCOMP	NO
			NO
			MMANOVA
			NPAR TESTS
			GRAPHICS
			PROBIT (SPSSx)
			SORT
			REBREGSSION
			T-TEST
			FREQUENCIES

## BMDP.

La primera versión de BMDP apareció en 1961. En 1968 se empezó a desarrollar el "English based Control Language" (Lenguaje de Control basado en el inglés), usado con los actuales programas BMDP. Este método de instrucciones específicas es más flexible que el formato fijo usado en los anteriores programas BMD.

Las siglas de BMDP significan "Biomedical Computer Programs" (Programas Biomédicos de Computadora). Los programas de BMDP están diseñados para auxiliar al análisis de datos al proveer métodos que van desde gráficas y la descripción simple de datos, hasta avanzadas técnicas estadísticas.

Los programas BMDP están diseñados para manejar todos los pasos en un análisis, desde lo simple hasta lo sofisticado. Están organizados según el problema a ser analizado, las variables a ser usadas en el análisis, y la estructura de los datos, que



ser usadas en el análisis, y la estructura de los datos, que deben estar especificados de una manera uniforme para todos los programas. Esto permite realizar diferentes análisis de los mismos datos con cambios mínimos en las instrucciones.

Los programas están clasificados en las siguientes series:

D: Descripción de datos.  
 F: Tablas de frecuencia.  
 R: Análisis de regresión.  
 V: Análisis de varianza.  
 M: Análisis multivariado.  
 L: Tablas de vida y análisis de supervivencia.  
 S: Especiales.  
 T: Series de Tiempo

Para mayor información comunicarse a:

BMDP Statistical Software.  
 1964 Westwood Blvd., Suite 202.  
 Los Angeles, California. 90025.  
 U.S.A

**PROCEDIMIENTOS  
 BMDP:**

**SPSS TIENE ALGUN  
 PROCEDIMIENTO COMPARABLE ?**

1D Descripción simple de datos	SI	LIST CASES
2D Descripción detallada de datos	SI	FREQUENCIES
3D Comparación de dos grupos con prueba T	SI	T-TEST
4D Frecuencias simples	SI	FREQUENCIES
5D Histogramas y gráficas univariadas	SI	FREQUENCIES
6D Gráficas bivariadas	SI	SCATTERGRAM
7D Descripción de grupos y an. de varianza	SI	BREAKDOWN
8D Correlaciones	SI	PEARSON CORR
9D Descripción de grupos multivariados	SI	MULT RESPONSE
4F Tablas de frecuencia multivariadas	SI	CROSS TABS
1L Tablas de vida y funciones de supervivencia	SI	SURVIVAL
2L Análisis de supervivencia con covariables	SI	SURVIVAL
1M Análisis cluster de variables	NO	
2M Análisis cluster de casos	NO	
3M Análisis cluster de variables y casos	NO	
4M Análisis factorial	SI	FACTOR
6M Análisis de correlación canónica	SI	CAN CORR
7M Análisis discriminante Stepwise	SI	DISCRIMINANT
8M Análisis factorial booleano	NO	
9M Puntajes lineales	NO	
AM Descripción y estimación de valores missing	NO	
KM Análisis cluster para casos	SI	QUICK CLUSTER (SPSSx)

2R	Regresión Stepwise	SI	REGRESSION
3R	Regresión no lineal	NO	
4R	Regresión en componentes principales	NO	
5R	Regresión polinomial	NO	
6R	Corr. parcial y regresión multivariada	SI	REGRESSION
9R	Regresión en todos los conjuntos posibles	NO	
AR	Regresión no lineal de derivadas libres	NO	
LR	Regresión logística Stepwise	NO	
1S	Transformaciones multipass	NO	
3S	Estadísticas no paramétricas	SI	NPAR TESTS
1T	An. espectral univariado y bivariado	NO	
2T	An. de series de tiempo Box-Jenkins	SI	BOX-JENKINS (SPSS::)
1V	An. de varianza y covarianza de una vía	SI	ONEWAY
2V	An. de varianza con repeticiones	SI	MANOVA
3V	Modelo general mixto de an. de varianza	SI	ANOVA
4V	An. de varianza univariada y multivariada	SI	ANOVA
8V	An. de varianza general con celdas iguales	SI	ANOVA

## OSIRIS

Osiris es un sistema de Análisis Estadístico y manejo de datos, compuesto de subrutinas diseñadas a ser usadas para manipulación de datos, conversión de datos, impresión de matrices, ordenamiento de arreglos, entrada / salida, etc.

Osiris IV es la versión actual de un paquete de software el cual ha sido desarrollado desde 1971, como parte del continuo desarrollo del sistema de programas Osiris. Hace uso de los conocimientos más recientes para reducir costos y mejorar capacidad y flexibilidad en las áreas de manejo de datos y análisis estadístico.

Osiris cuenta con facilidades para:

- . Agrupamiento de datos.
- . Almacenamiento y análisis de archivos.
- . Despliegue de datos.
- . Manejo de datos ponderados.
- . Edición y corrección de datos.
- . Transformación de valores de datos a través de operaciones aritméticas y lógicas.
- . Generación de distribuciones de frecuencia, univariadas y bivariadas, y estadísticas relacionadas.
- . Produce gráficas de puntos.

- . Produce gráficas de puntos.
- . Ejecuta análisis de correlación, regresión múltiple, varianza univariada y multivariada.
- . Análisis conductual con predictores múltiples, donde los predictores son nominales u ordinales.
- . Calcula errores de muestreo para diseños complejos no aleatorios.
- . Análisis factoriales.
- . Análisis cluster.

Para mayor información comunicarse a:

Statistical Analysis and Data Management Software System.  
 Institute for Social Research.  
 Computer Support Group, P.O.  
 Box 1248. University of Michigan. 48106.  
 U.S.A

PROCEDIMIENTOS  
 OSIRIS:

SPSS TIENE ALGUN  
 PROCEDIMIENTO COMPARABLE ?

PROCEDIMIENTOS OSIRIS:	DESCRIPCIÓN	SPSS TIENE ALGUN PROCEDIMIENTO COMPARABLE ?	PROCEDIMIENTO COMPARABLE
&SCAT	Diagramas bivariados	SI	SCATTERGRAM
&TABLES	Tablas bivariadas y estadísticas	SI	CROSSTABS
&USTATS	Estadísticas univariadas	SI	CONDESCRIPTIVE
&DREG	Análisis de regresión dicotómica	NO	
&MDC	Análisis de correlación de Pearson	SI	PEARSON
&PARTIALS	Análisis de correlación parcial	SI	PARTIAL CORR
&REGRESSN	Análisis de regresión múltiple	SI	REGRESSION
&ANOVA	Análisis de varianza de una vía	SI	ONEWAY
&MANOVA	Análisis de varianza multivariado	SI	MANOVA
&MCA	Análisis de clasificación múltiple	SI	ANOVA
&MNA	Análisis multivariado nominal	NO	
&SEARCH	SEARCH	NO	
&CAP	Análisis en una configuración espacial	NO	
&COMPARE	Compara matrices de configuración	NO	
&FACTAN	Análisis factorial	SI	FACTOR
&MINISSA	Escalamiento multidimensional	NO	
&CLUSTER	Análisis cluster jerárquico.	SI	CLUSTER (SPSSx)
&PSALMS	Análisis de error de muestreo	NO	
&REPERR	Error de muestreo con repeticiones	NO	
&PSRATIO	Medias estimadas de subpoblaciones	NO	
&PSTABLE	Proporciones estimadas de subpob.	NO	
&PSTOTAL	Totales estimados de subpoblaciones	NO	

## BIBLIOGRAFIA.

Bialock, Hubert M. :  
Social Statistics.  
2a ed. New York. Ed. Mc Graw-Hill. 1972.

BMDP Statistical Software.  
1983 Revised Printing. Printing with Additions.  
University of California Press.

Hull, C. Hadlai y Norman H. Nie :  
SPSS. Update 7-9.  
New Procedures and Facilities for Releases 7-9.  
Ed. Mc Graw-Hill. 1981.

Nie, Norman H., C. Hadlai Hull, Jean G. Jenkins,  
Karin Steinbrenner y Dale H. Bent. :  
SPSS. Statistical Package for the Social Sciences.  
Second Edition. Ed. Mc Graw-Hill. 1975.

Norusis, Marija J. :  
SPSS/PC for the IBM PC/XT.  
SPSS Inc. 1984.

OSIRIS IV Subroutine Manual.  
4a ed. Survey Research Center. 1986.

SAS USER'S GUIDE.  
SAS Institute Statistical Analysis System. 1979.

SPSSx. USER'S GUIDE.  
SPSS Inc. 1983.

## REFERENCIAS BIBLIOGRAFICAS.

Para profundizar en los temas estadísticos tratados en cada una de las subrutinas se recomienda consultar los siguientes libros:

### CONDESCRIPTIVE y FREQUENCIES

Blalock, Hubert M. :  
Social Statistics.  
2a ed. New York. Ed. Mc Graw-Hill. 1972.

Mueller, John H., Karl F. Schuessler, y Herbert L. Costner :  
Statistical Reasoning in Sociology.  
2a ed. Boston. Ed. Houghton Mifflin. 1970.

### CROSTABS

Blalock, Hubert M. :  
Social Statistics.  
2a ed. New York. Ed. Mc Graw-Hill. 1972.

Davis, James A. :  
Elementary Survey Analysis.  
Englewood, Cliffs. Ed. Prentice-Hall. 1971.

\*Rosenberg, Morris :  
The Logic of Survey Analysis.  
New York. Ed. Basic Books. 1968.

### BREAKDOWN y T-TEST

Blalock, Hubert M. :  
Social Statistics.  
2a ed. New York. Ed. Mc Graw-Hill. 1972.

Hays, William G. :  
Statistics for Psychologists.  
New York. Ed. Holt, Rinehart and Winston. 1963.

### PEARSON CORR, NONPAR CORR, SCATTERGRAM, PARTIAL CORR y REGRESSION

Blalock, Hubert M. :  
Social Statistics.  
2a ed. Ed. New York. Ed. Mc Graw-Hill. 1972.

Blalock, Hubert M. :  
Causal Inference in Non-experimental Research.  
Chapel, Hill. Ed. University of North Carolina Press. 1964.

\*Draper, N. R. y H. Smith :  
Applied Regression Analysis.  
New York. Ed. Wiley. 1966.

Johnston, J. :  
Econometric Methods.  
2a ed. New York. Ed. Mc Graw-Hill. 1972.

Kerlinger, F. N. y E. Pedhazzer :  
Multiple Regression in Behavioral Research.  
New York. Ed. Holt Rinehart and Winston. 1973.

#### **ANOVA y ONEWAY**

Cochran, W. G. y G. M. Cox :  
Experimental Designs.  
2a ed. New York. Ed. Wiley. 1957.

Kirk, R. :  
Experimental Design: Procedures for the Behavioral Sciences.  
Belmont, Calif. Ed. Brooks. 1968.

\*Scheffe, H. A. :  
The Analysis of Variance.  
New York. Ed. Wiley. 1959.

Winer, B. J. :  
Statistical Principles in Experimental design.  
2a ed. New York. Ed. Mc Graw Hill. 1971.

#### **DISCRIMINANT**

Cooley, William W. y Paul R. Lohnes :  
Multivariate Data Analysis.  
New York. Ed. Wiley. 1971.

Tatsuoka, Maurice M. :  
Multivariate Analysis.  
New York. Ed. Wiley. 1971.

Van de Gear, John P. :  
Introduction to Multivariate analysis for the Social Sciences.  
San Francisco. Ed. W.H. Freeman. 1971.

#### **FACTOR**

Harman, Harry H. :  
Modern Factor Analysis.  
Chicago. The University of Chicago Press. 1967.

\*Stephenson, W. :  
The study of Behavior.  
Chicago. The University of Chicago Press. 1953.

## CANCORR

- Anderson, T. W. :  
Introduction to Multivariate Statistical Analysis.  
New York. Ed. Wiley. 1958. chap. 12.
- Cooley, W. W. y P. R. Lohnes :  
Multivariate Procedures for the Behavioral Sciences.  
New York. Ed. Wiley. 1962. chap. 3.
- Van der Geer, J. :  
Introduction to Multivariate Analysis for the Social Sciences.  
San Francisco. Ed. Freeman. 1971. chap. 14.

## GUTTMAN SCALE

- Dawes, Robyn M. :  
Fundamentos y Técnicas de Medición de Actitudes.  
México. Ed. Limusa. 1975.
- Summers, Gene F. :  
Medición de Actitudes.  
México. Ed. Trillas. 1984.

## RELIABILITY

- Dawes, Robyn M. :  
Fundamentos y Técnicas de Medición de Actitudes.  
México. Ed. Limusa. 1975.
- Guilford, J.P. y Benjamin Fruchter :  
Estadística Aplicada a la Psicología y la Educación.  
México. Ed. Mc Graw-Hill. 1984.
- Summers, Gene F. :  
Medición de Actitudes.  
México. Ed. Trillas. 1984.

## MANOVA

- Bock, R. D. :  
Multivariate Statistical Methods in Behavioral Research.  
New York. Ed. Mc Graw-Hill. 1975.
- Cooley, W. W. y P. R. Lohnes :  
Multivariate Data Analysis.  
New York. Ed. Wiley. 1971.
- Morrison, D. F. :  
Multivariate Statistical Methods.  
2a ed. New York. Ed. Mc Graw-Hill. 1976.

## **BOX-JENKINS**

Anderson, O. D.,  
Time Series Analysis and Forecasting - The Box-Jenkins Approach.  
Boston. Ed. Butterworth. 1976.

\*Box, G. E. P. y G. M. Jenkins :  
Time Series Analysis: Forecasting and Control.  
San Francisco. Ed. Holden-Day. 1976.

Fuller, Wayne A. :  
Introduction to Statistical Time Series.  
New York. Ed. Wiley. 1976.

Nelson, Charles R. :  
Applied Time Series Analysis for Managerial Forecasting.  
San Francisco. Ed. Holden-Day. 1973.

## **SURVIVAL**

Gehan, E. :  
Statistical Methods for Survival Time Studies, in Cancer Therapy:  
Prognostic Factors and Criteria of Response.  
New York. Ed. Raven Press. 1975.

Gross, A. J. y V. Clark :  
Survival Distributions: Reliability Applications in the  
Biomedical Sciences.  
New York. Ed. Wiley. 1975.

## **NONPAR TESTS**

Conover, W. J. :  
Practical Nonparametric Statistics.  
New York. Ed. Wiley. 1971.

Hollander, M. y D. A. Wolfe :  
Nonparametric Statistical Methods.  
New York. Ed. Wiley. 1973.

\*Siegel, S. :  
Nonparametric Statistics for the Behavioral Sciences.  
New York. Ed. Mc Graw-Hill. 1956.

\* Libros muy reconocidos en el tema, considerados por algunos autores como clásicos.



Desde 1986 SPSS Inc. es el nuevo distribuidor de la documentación de su completa línea de software. Los manuales que se ordenaban a Mc Graw-Hill Book Company deben ahora ordenarse directamente a:

SPSS Inc.  
Publication Sales  
444 North Michigan Avenue  
Chicago, Il 60611

312/ 329-3600

Los manuales y reportes técnicos disponibles son:

SPSS Combined Edition. 1981.

SPSS Primer. 1975.

SPSS Introductory Guide: Basic Statistics and Operations. 1982.

SCSS: A guide to the SCSS Conversational System. 1980.

SCSS Short Guide: An introduction to the SCSS. 1981.

SPSSx Users Guide. Second Edition. 1985.

SPSSx Basics. 1983.

SPSSx Introductory Statistics Guide. 1983.

SPSSx Advanced Statistics Guide. 1985.

SPSSx TABLES. 1985.

SPSS Graphics. 1985.

SPSS/PC. 1985.

SPSS/PC + Advanced Statistics. 1986.

SPSS/PC + TABLES. 1986.

SPSS-11: The SPSS Batch System for the DEC PDP-11. 1982.

SPSS/Pro: SPSS for the DEC Professional 350. 1984.

SPSSx Statistical Algorithms. 1983.

SPSSx Processing of U.S. Census Data. 1984.

# INDICE ALFABETICO

ABS (Función)	55	CONTRAST (ONEWAY)	139
AGGREGATE	17, 77	CONVERT (RECODE)	54
AGGSTATS	78	COS (Función)	55
Ejemplo	81	*COUNT	60
Estadísticas	80	COUNT	60
GROUPVARS	77	CROSSBREAK (BREAKDOWN)	96
Opciones	80	Ejemplo	97
RMISS	78	Estadísticas	96
AGGSTATS (AGGREGATE)	78	Opciones	96
ALPHA (TYPE)	160	CROSSTABS	17, 83
ANALYSIS (DISCRIMINANT)	147	Ejemplo	87
AND (Op. Lógico)	57	Estadísticas	86
ANOVA	18, 132	Opciones	86
Ejemplo	137		
Estadísticas	135	DATA LIST	27
Opciones	134	DELTA (FACTOR)	163
ANY (Función)	202	DIAGONAL (FACTOR)	160
ASIN (Función)	202	DIRECT (METHOD)	144
ATAM (Función)	55	DISCRIMINANT	19, 144
		ANALYSIS	147
BARChart (Gráficas)	192	Ejemplo	154
BLANK (RECODE)	54	Estadísticas	153
BOX-JENKINS	191	FIN	148
BREAKDOWN	17, 90	FOUT	149
Ejemplo	94	FUNCTIONS	150
Estadísticas	92	GROUPS	146
Opciones	93	MAXSTEPS	148
BYPASS (TYPE)	160	METHOD	144
Backward (REGRESSION)	123	Opciones	151
		PIN	149
		POUT	149
CANCORR	19, 173	PRIORS	150
CANVAR	174	TOLERANCE	148
Ejemplo	178	VIN	149
Estadísticas	177	DO REPEAT	66
Matriz de entrada	175	DOCUMENT	44
Opciones	176	DUNCAN (ONEWAY)	140
RELATE	174		
CANVAR (CANCORR)	174	EDIT	21
CFVAR (Función)	202	ELSE	53
CLEAN (SPSSx)	203	END INPUT DATA	196
CLUSTER (SPSSx)	204	END REPEAT	66
COMMENT	21	EQ (Op.Relacional)	57
*COMPUTE	55	EQUAL (PRIORS)	150
COMPUTE	55	EQUIMAX (ROTATE)	162
CONDESCRIPTIVE	17, 69	EXP (Función)	55
Ejemplo	71		
Estadísticas	69		
Opciones	70		

FACSCORE (FACTOR)	163	KURT (AGGSTATS)	78
FACTOR	19, 159		
DELTA	163		
DIAGONAL	160	LE (Op.Relacional)	57
Ejemplo	168	LG10 (Función)	55
Estadísticas	167	LINECHART (Gráficas)	192
FACSCORE	163	LISREL (SPSSx)	204
ITERATE	161	LIST CASES	37
MINEIGEN	161	LIST FILE INFO	45
Matriz de entrada	164	LN (Función)	55
NFACTORS	161	LOGLINEAR (SPSSx)	203
Opciones	166	LOWEST	53, 60, 110
ROTATE	162	LSD (ONEWAY)	140
STOPFACT	161	LSDMOD (ONEWAY)	140
TYPE	160	LT (Op.Relacional)	57
FILE NAME	42		
FIN (DISCRIMINANT)	148	MAHAL (METHOD)	145
FINISH	34	MANOVA	190
FOUJ (DISCRIMINANT)	149	MATRIX (SPSSx)	203
FRECUENCIES	17, 72	MAX (AGGSTATS)	78
Ejemplo	75	MAX (Función)	202
Estadísticas	73	MAXINF (METHOD)	145
Opciones	73	MAXSTEPS (DISCRIMINANT)	148
FUNCTIONS (DISCRIMINANT)	150	MEAN (AGGSTATS)	78
Formato Fortran	26	MEAN (Función)	202
Forward (REGRESSION)	123	METHOD (DISCRIMINANT)	144
Funciones	55, 195, 202	MIN (AGGSTATS)	78
		MIN (Función)	202
GE (Op.Relacional)	57	MINEIGEN (FACTOR)	161
GENERAL (FRECUENCIES)	72	MINRESID (METHOD)	145
GET FILE	44	MISSING VALUES	40
GROUPS (DISCRIMINANT)	146	MOD (Función)	202
GROUPS (T-TEST)	98	MOD10 (Función)	55
GROUPVARS (AGGREGATE)	77	MULT RESPONSE	189
GT (Op.Relacional)	57		
GUTTMAN SCALE	19, 180		
Ejemplo	184	N OF CASES	30, 196
Estadísticas	183	NE (Op.Relacional)	57
Opciones	183	NEW REGRESSION	191
Gráficas	192	NFACTORS (FACTOR)	161
		NONPAR CORR	18, 106
HIGHEST	53, 60, 110	Ejemplo	108
HILOGLINEAR (SPSSx)	204	Estadísticas	107
		Opciones	106
*IF	57	NORMAL (Función)	195
IF	57	NRotate (ROTATE)	162
IMAGE (TYPE)	160	NOT (Op. Lógico)	57
INPUT FORMAT	25	NPAR TESTS	188
INPUT MEDIUM	29	NUMBERED	23
INTEGER (FRECUENCIES)	72		
ITERATE (FACTOR)	161		

OBLIQUE (ROTATE)	162	REGRESSION	18, 121
ONEWAY	19, 138	Backward	123
CONTRAST	139	Ejemplo	128
Ejemplo	141	Estadísticas	128
Estadísticas	141	Forward	123
Opciones	140	Matriz de entrada	125
POLYNOMIAL	138	Opciones	126
RANGES	139	Parametros	124
OPTIONS	32	REGRESSION	122
OR (Op. Lógico)	57	RESID	125
Op. Lógicos	57	Stepwise	123
Op. Relacionales	57	RELATE (CANCORR)	174
		RELIABILITY	190
		REPORT	187
PA1 (TYPE)	160	RMISS (AGGREGATE)	78
PA2 (TYPE)	160	RND (Función)	55
PAGESIZE	36	ROTATE (FACTOR)	162
PAIRS (T-TEST)	99	RUN NAME	23
PARTIAL CORR	18, 113	RUN SUBFILES	51
Ejemplo	119		
Estadísticas	118	*SAMPLE	66
Matriz de entrada	115	SAMPLE	66, 193
Opciones	117	SAVE FILE	43
PCTBTN (AGGSTATS)	78	SCATTERGRAM	18, 109
PCTGT (AGGSTATS)	78	Ejemplo	111
PCTLY (AGGSTATS)	78	Estadísticas	111
PEARSON CORR	18, 102	Opciones	110
Ejemplo	105	SCHEFFE (ONEWAY)	140
Estadísticas	104	SD (AGGSTATS)	78
Opciones	103	SD (Función)	202
PIECHART (Gráficas)	192	SEED	194
PIN (DISCRIMINANT)	149	*SELECT IF	62
PLOT (SPSSx)	204	SELECT IF	62
POLYNOMIAL (ONEWAY)	138	SIN (Función)	55
POUT (DISCRIMINANT)	149	SIZE (PRIORS)	150
PRINT BACK	36	SKEW (AGGSTATS)	78
PRINT FORMAT	35	SNK (ONEWAY)	140
PRIORS (DISCRIMINANT)	150	SORT CASES	82, 197
PROBIT (SPSSx)	203	SORT (Función)	55
		STATISTICS	31
QUARTIMAX (ROTATE)	162	STOPFACT (FACTOR)	161
QUICK CLUSTER (SPSSx)	204	SUBFILE LIST	50
		SUM (AGGSTATS)	78
RANGE (Función)	202	SUM (Función)	202
RANGES (ONEWAY)	139	SURVIVAL	187
RAD (METHOD)	145	Stepwise (REGRESSION)	123
RAD (TYPE)	160		
RAW OUTPUT UNIT	33	T-TEST	17, 98
READ INPUT DATA	33, 197	Ejemplo	101
READ MATRIX	197	Estadísticas	99
*RECODE	53	GROUPS	98
RECODE	53	Opciones	99
		PAIRS	99

TABLES (SPSSx)	203	VALIDN (AGGSTATS)	78
TASK NAME	31	VALUE LABELS	39
THRU	53, 60	VAR LABELS	38
TOLERANCE (DISCRIMINANT)	148	VARIABLE LIST	24
TRUNC (Función)	55	VARIANCE (Función)	202
TUKEY (ONEWAY)	140	VARIMAX (ROTATE)	162
TUKEYB (ONEWAY)	140	VIN (DISCRIMINANT)	149
TYPE (FACTOR)	160		
		*WEIGHT	63
UNIFORM (Función)	195	WEIGHT	63
UPDATE (SPSSx)	203	WILKS (METHOD)	145
USERPROC (SPSSx)	203	WRITE CASES	48
		WRITE FILEINFO	49
		YRMODE (Función)	195

## REFERENCIA RAPIDA.

AGGREGATE	77	FILE NAME	42
ANOVA	132	FINISH	34
BREAKDOWN	90	GET FILE	44
CANCORR	173	IF	57
CONDESCRIPTIVE	69	INPUT FORMAT	25
CROSSTABS	83	INPUT MEDIUM	29
DISCRIMINANT	144	LIST FILEINFO	45
FACTOR	159	MISSING VALUES	40
FREQUENCIES	72	N OF CASES	30
GUTTMAN SCALE	180	NUMBERED	23
LIST CASES	37	OPTIONS	32
NONPAR CORR	106	PAGESIZE	36
ONEWAY	138	PRINT BACK	36
PARTIAL CORR	113	PRINT FORMAT	35
PEARSON CORR	102	RAW OUTPUT UNIT	33
REGRESSION	121	READ INPUT DATA	33
SCATTERGRAM	109	RECODE	53
SORT CASES	82	RUN NAME	23
T-TEST	98	RUN SUBFILES	51
WRITE CASES	48	SAMPLE	66
		SAVE FILE	43
		SELECT IF	62
		STATISTICS	31
COMMENT	21	SUBFILE LIST	50
COMPUTE	55	TASK NAME	31
COUNT	60	VALUE LABELS	39
DATA LIST	27	VAR LABELS	38
DO REPEAT - END REPEAT	66	VARIABLE LIST	24
DOCUMENT	44	WEIGHT	63
EDIT	21	WRITE FILEINFO	45