

EL SPSS*: REALIDAD Y MITO

* (STATISTICS
PACKAGE
FOR SOCIAL
SCIENCES)

Alejandro Calatayud A¹

103

I. Introducción

1. ¿Qué es el SPSS?

En el propio manual¹ del SPSS, se define como: "Un sistema integrado por programas de computadora diseñado para las Ciencias Sociales".

Para los usuarios experimentados, podríamos decir que es perfectamente entendible dicha definición. Pero no para un gran número de usuarios potenciales que se acercan a él buscando la solución de sus problemas, sin ningún conocimiento tanto de computación como de estadística.

** Investigador de tiempo completo del IISUNAM y profesor de la FCPyS, UNAM.

¹ Norman, H.N.E., et al. *SPSS. Statistical Package for the Social Sciences* Mc Graw Hill, Second edition, USA. 1975.

Para hacer la definición más entendible o accesible a los interesados, trataremos de explicarla.

Por un sistema integrado de programas puede entenderse, una colección de programas interrelacionados o integrados y almacenados en conjunto. En otras palabras, un grupo de programas en el que cada uno de ellos realiza una tarea específica, que puede repetirse cada vez que se quiera, por lo que también reciben el nombre de rutinas y procedimientos.

En conjunto, con todos estos programas podemos realizar un análisis estadístico simple (distribuciones de frecuencia y/o calcular medidas de resumen, etc.) o uno muy sofisticado (análisis factorial, discriminante, etc.).

El SPSS cuenta con rutinas para crear archivos, para producir con éstos otros datos, y con procedimientos para llevar a cabo cálculos estadísticos e imprimir los resultados de una forma uniforme e igual cada vez que se requiera.

El SPSS no sólo puede usarse en las Ciencias Sociales, sino en todas las áreas del saber, como de hecho se hace.

Pero, ¿Qué es un programa? y ¿Cuál su necesidad?

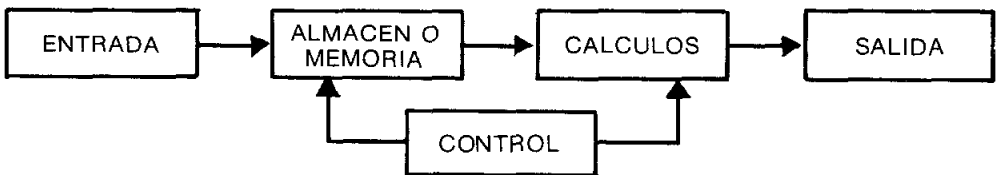
Un programa es un conjunto explícito de pasos a seguir para lograr un fin determinado, o bien, un conjunto de instrucciones (algoritmo) con las cuales

una computadora trabaja. Es decir, con ellas se le instruye para que ejecute ciertas operaciones, las cuales realiza en forma mecánica.

Para comprender mejor estas ideas veamos, aunque sea someramente, qué es una computadora.

Primero diremos que una calculadora de escritorio, como las que todos conocemos y en ocasiones hemos usado, es una máquina que realiza tareas u operaciones, como sumar y restar, en forma mecánica, apretando simplemente una tecla. Una computadora además de contar con una área de este tipo, para efectuar operaciones, cuenta con otra que, por medio de instrucciones (de transferencia, de control, de almacén, de especificación, etc.), puede controlar el orden en que la sucesión de operaciones debe realizarse.

Haciendo un poco de historia podría considerarse a la propuesta de Charles Babbage (1791-1871)² como el diseño de la primera com-



² Levine, G. Guillermo. *Introducción a la computación y la estructura programada*. Mc Graw-Hill. México, 1984.

putadora electrónica, la cual, de hecho, es la base o modelo de las computadoras modernas.

Como lo muestra el esquema, una computadora está formada por una Unidad de Entrada, que recibe tanto la información como las instrucciones para su procesamiento, almacenamiento e impresión (programa). La Unidad de Memoria, en donde se almacena la información (datos y programas). La de Procesamientos o Cálculos sobre la información almacenada. La Unidad de Control o Unidad de Procesamiento Central, ampliamente conocida como CPU, que dirige a todas las demás unidades; determinando paso a paso, de acuerdo al programa, cuándo debe leer, dónde almacenar, y los cálculos a realizar. Finalmente, cuenta con una Unidad de Salida, que nos da el reporte de la información procesada, de acuerdo también a las instrucciones recibidas.

El SPSS es mucho más de lo que se ha dicho de él, pero no es la panacea para resolver los problemas de las Ciencias Sociales.

La idea de cómo trabaja la computadora, fue propuesta mucho tiempo después por el ingeniero y matemático John Von Neuman (1946) y consiste en almacenar las instrucciones del programa en la memoria de la computadora logrando así que siga los pasos definidos por él.

En resumen, el elemento fundamental de una computadora es el programa. El concepto de programa almacenado permite al computador escoger uno o varios cursos de acción basados sobre datos de entrada o sobre resultados de un proceso previo, los cuales son pasados a través de un grupo de instrucciones ordenadas en un algoritmo.

Para continuar la explicación de la definición del SPSS, nos faltaría discutir el "... diseñado para el análisis de las Ciencias Sociales", para lo cual me permito hacer las siguientes reflexiones.

Primero, podría pensarse que el SPSS es una propuesta "nueva" respecto a las ya existentes, pero no. Vayamos por partes. Las siglas del paquete lo definen como un Paquete Estadístico para las Ciencias Sociales, lo cual concuerda con lo de "... análisis de datos" entendiendo por esto el campo de acción de las técnicas estadísticas; es decir, describir y poner a prueba hipótesis de investigación. Pero además, el SPSS, cuenta con procedimientos que permiten el procesamiento de datos, lo cual implica la transformación y el manejo de datos en términos de agregar y/o construir otros datos.

Segundo, hablar de análisis de datos para las Ciencias Sociales nos induce a pensar que se trata de un sistema aplicable a cualquier tipo de perspectiva teórica, ya que los datos son registros de hechos observados de la realidad, y a ésta nos acercamos armados con una perspectiva teórica determinada. Es necesario recordar que el análisis estadístico y las técnicas que lo permiten, fueron desarrolladas para las Ciencias Nomotéticas, es decir, aquellas cuyo interés primordial es la elaboración de leyes que permitan explicar los fenómenos. En las Ciencias Sociales, éstas son referidas como investigaciones empíricas. Por lo tanto puede decirse que el SPSS, no sólo puede usarse en las Ciencias Sociales bajo esta perspectiva, sino en todas las áreas del saber, como de hecho se da. Más adelante retomaremos esta discusión.

En resumen, el SPSS, es un conjunto de programas de rutinas y procedimientos estadísticos para la construcción y manejo de archi-

vos, para la manipulación, transformación y creación de nuevas variables, y para efectuar el análisis estadístico de los datos recabados en investigaciones del tipo del Survey (encuestas), Panels, etc. Estos programas fueron diseñados para que realizaran con rapidez y gran eficacia el procesamiento de los datos que, de otra manera, no sólo llevaría muchísimo tiempo, sino que sería muy difícil resolver.

En suma, podemos decir, que la definición del SPSS, no es muy justa. El paquete es mucho más de lo que se ha dicho de él y no es, tampoco, la panacea para resolver los problemas planteados por las Ciencias Sociales.

Los invitamos a conocerlo con más detalle.

II. La realidad

1. El desarrollo del paquete SPSS

Los diseñadores del SPSS tardaron cerca de una década en elaborar el sistema. Su desarrollo se inicia en el año de 1965 en la Universidad de Stanford y posteriormente el proyecto es trasladado al National Opinion Research Center de la Universidad de Chicago.

Según los autores del paquete, el SPSS surgió por la necesidad de responder a una generalizada frustración de investigadores en Ciencia Política de la propia Universidad de Stanford. Lo cual coincide con el desarrollo en la UNAM de Centros de Investigación Social, cuando la computación es incorporada.

La incorporación paulatina a las universidades mexicanas de la Ciencia de la Computación, aparece como la solución mágica de todos los problemas sobre el manejo y análisis estadístico de grandes volúmenes de información. En efecto, la computación abrió, a las Ciencias en general, enormes posibilidades, no sólo en el manejo y precisión en el cálculo, sino en la posibilidad de aplicar técnicas más sofisticadas y complejas, modelos matemáticos, simulación, etc.

Con la instalación de computadoras se olvidó que éstas eran máquinas y no cerebros. Estas máquinas requerían para su manejo dos condiciones fundamentales, que parece fueron soslayadas: saber operarlas y saber programarlas. Consideramos que este problema fue más significativo en las Ciencias Sociales debido a la formación



básica de sus profesionales, que no son instruidos ni preparados en el campo de la computación. La gran expansión de las computadoras y el deseo de incorporarse de lleno al avance tecnológico, hizo que se improvisaran tanto operadores como programadores para responder así al avance tecnológico. Los resultados fueron efectivamente los señalados por los autores del SPSS, frustración, pues los procesamientos de las investigaciones, como las encuestas de opinión, insumieron demasiado tiempo, en ocasiones sin los resultados esperados, y en algunos casos esta frustración provocó incluso el abandono de los proyectos.

Si bien los investigadores no podían “dedicarse” a prepararse en las ciencias de la computación y en el dominio de la estadística; los programadores no tenían tampoco tiempo, ni para conocer las técnicas de investigación, ni mucho menos las técnicas estadísticas; pues el rápido desarrollo de la computación siempre los dejaba atrás.

Con el tiempo se prepararon y formaron verdaderos profesionales de la computación. Sin embargo, a éstos se les siguió dando toda la responsabilidad que los investigadores les habían dado desde el inicio: la incorporación de la computación a las Ciencias Sociales.

Esta realidad influyó, quizás, en los autores del SPSS para conformar un equipo multidisciplinario de trabajo, en el que no sólo estuvieron profesionales de las Ciencias Sociales y de la computación, sino también expertos en Estadística Matemática. Así, el SPSS vino a resolver el problema, haciendo mucho más fácil el procesamiento de información y del cálculo de las estadísticas. Sin embargo, recuérdese que un equipo de expertos con las condiciones adecuadas, o al menos mejores que las nuestras, tardaron cerca de una década en desarrollarlo.

El equipo inicial de la Universidad de Stanford, publicó en 1970 la primera edición del SPSS. Ésta fue mejorando poco a poco, con las observaciones de sus primeros usuarios y las del mismo equipo que la elaboró, dando como resultado las llamadas “Versiones del SPSS”. Así, por ejemplo, en el año de 1972, el Instituto de Investigaciones Sociales de la UNAM, adquiere para su uso exclusivo la versión 4 del SPSS. El paquete fue adaptado a la Burroughs B-5500 —con la que operaba el Centro de Cálculo Electrónico de la propia Universidad— por el Ing. Romualdo Vitela García, Jefe del Departamento de Programación de dicho Instituto.

Durante los años siguientes el SPSS siguió mejorándose y es en 1975 cuando se publica la segunda edición, que contiene la versión 6. Esta publicación ya fue hecha en la Universidad de Chicago y considero que es la más conocida y la más utilizada de las versiones de SPSS.

Posteriormente, en el año de 1979, se publica el SPSS-UPDATE,³ con las versiones 7 y 8, y que actualmente son las que se manejan en la UNAM. Las mejoras más importantes de estas versiones son, además de incluir lo que contiene la versión 6:

—La inclusión de 14 procedimientos estadísticos no paramétricos, entre los más conocidos: la prueba de Runs, la de McNe-

³ Hull C., Hadlai and Nie H., Norman. *SPSS-UPDATE, New Procedures and Facilities. For release 7 and 8.* Mc Graw-Hill Book Company, USA, 1979.

mar, la del Signo, la de Kolmogorof-Smirnov y la de Kruskal-Wallis en el análisis de varianza simple.

—El manejo estadístico para el tratamiento de la respuesta múltiple.

—El de la construcción de Tablas de Supervivencia para Estudios Demográficos.

—Y otras mejoras de carácter de programación, como los reportes o salidas de los procedimientos.

En suma, creemos que las nuevas versiones, así como la infinidad de lenguajes y paquetes desarrollados, tienden a hacer más fácil la tarea en el uso y programación de las computadoras modernas.

Por último, en la actualidad no sólo contamos con SPSS diseñado para las grandes computadoras, sino que hay versiones para minis y para micros, como las denominadas PC (Personal Computer).

Queremos aclarar que no sólo existe el SPSS sino una gran cantidad de paquetes y facilidades para el manejo y análisis de información, algunos de ellos desarrollados con anterioridad al SPSS como el DATA TEX de la Universidad de Harvard, y otros que fueron desarrollados en forma paralela, como son el SAS⁴ (Statistical Analysis System) y el BMD⁵ series-p (Biomedical Computer Programs).

Ahora bien, ¿cuál es la llave o palabra mágica que nos permitirá usar el SPSS? En otras palabras, ¿cómo entramos al SPSS?

En general, se podría decir que existe un procedimiento distinto para cada sistema, pero en esencia es el mismo, y consiste en lo que llaman tarjetas de control de acceso a la computadora, mejor dicho al sistema.

Los ejemplos dados abajo, son para procesos en lotes (batch), éstos son mudos para el usuario, el cual ve los resultados hasta el final de su corrida.

Así por ejemplo, para la Burroughs B-7800⁶ de la UNAM las tarjetas de control son:

```
?JOB XXX USER XXXX/YY; CLASS=N : BEGIN
RUN *SERVICIO/SPSS:
DATA IOCR
  —PROGRAMA EN SPSS
  READ INPUT DATA
  —DATOS
  FINISH
```

Mientras que para la microcomputadora, como la de la UAM-Iztapalapa y la de Xochimilco, HP-3000⁷ el usuario requiere de las tarjetas de control siguientes:

⁴ Ban J., Anthony et al. *A user's guide to SAS>6SAS>7*. Supplemental Library User's Guide, Institute, Inc. Raleigh NC, 1976-1977.

⁵ Dixon, W.J. *Biomedical computer programs*. The University of California Press, Berkeley, 1977.

⁶ Vitela G., Romualdo. *Notas para una mejor utilización de SPSS*. Tecninota 79/21 C.S.C., UNAM. 1979.

⁷ Castillo M., Alberto y Serrano M., Alejandro. *Guía para el principiante SPSS*, UAM-Iztapalapa.

: JOB EJEMPLO, ALUMNO. 10999ZZ, GRUPO999/POR
: RUN SPSS2.PUB. SYS
Programa de SPSS con Datos
: EOJ



109

En las micros, compatibles con las IBM-PC (Printaform, Televideo, Corona, etc), es necesario que cuenten con disco duro de cuando menos 10 MG bytes. En las micros, el SPSS trabaja en modo iterativo (platicando con el usuario), su acceso es como sigue:

a) Desde la raíz teclear SPSSPC

b) Iniciar con su programa (por lo general, inicializar los dispositivos de entrada y salida); por ejemplo SET ON DISK, PRINT enviará los resultados y programa al disco y a la impresora.

—Tarjetas de definición de datos y/o datos. Por lo general los datos son cargados utilizando un procesador de textos, a un archivo diferente de aquel en el que está el programa.

—Procedimientos

—Fin

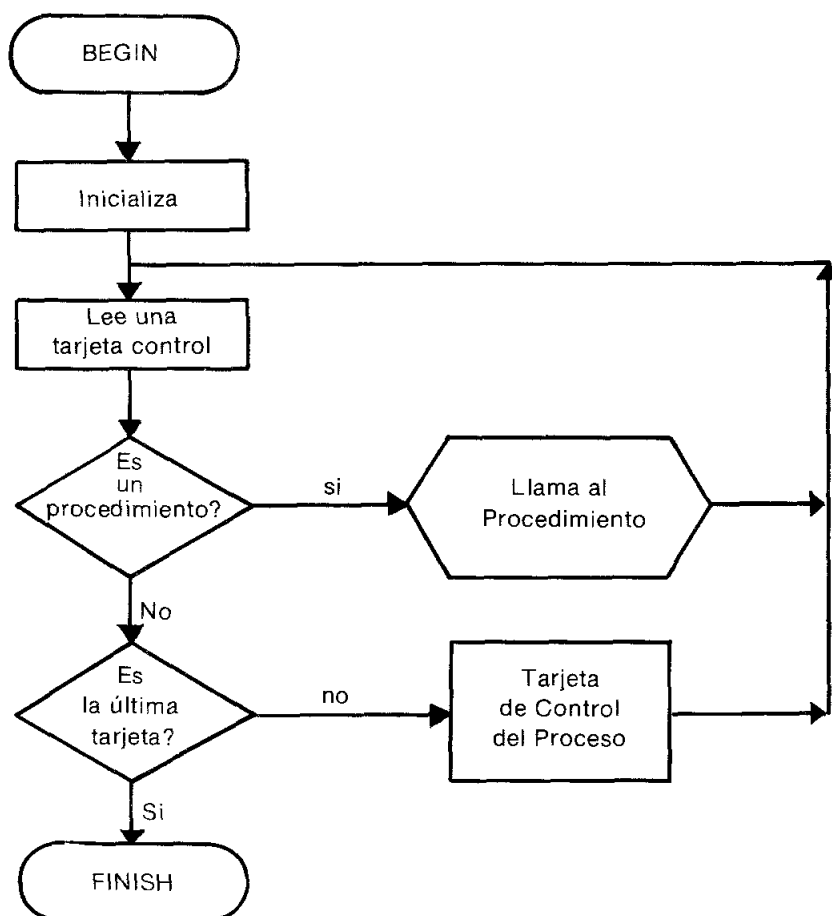
La liga entre programas y datos se establece en la instrucción de DATA LIST.

2. El SPSS un sistema integrado por programas

Podría decirse que la estructura del SPSS responde a una idea de programación modular.⁸ Todos los programas están gobernados por un Programa Central, que controla el inicio y término de una sesión del SPSS. Controla entre estas dos fases las rutinas y procedimientos estadísticos "invocados". Estas rutinas y procedimientos, una vez que son accedados o jalados, trabajan en forma independiente hasta terminar volviendo siempre al Programa Central (PC).

⁸ Anselmi, Juan Carlos. *Métodos de Programación Descripciones y Aplicaciones*. Publicación Fotocopiado de Dirección General de Servicios de Cómputo Académico, UNAM, 1986.

Diagrama de flujo del Programa Central⁹



110

Este programa muestra la lógica de cómo se lleva a cabo la supervisión de la realización de las tareas. Su función es determinar qué es lo que el usuario desea hacer.

En él se explica que el Programa Central (PC), primero ejecuta un procedimiento de inicialización, el cual, entre otras cosas, prende varios switches en su valor default —salidas de resultados de una forma fija— y determina la correspondencia entre los datos del conjunto usados en el SPSS.

Después el PC nos introduce en un proceso iterativo (LOOP) que continúa hasta que se encuentra una tarjeta FINISH (de determinación de la actividad programada). Este LOOP se inicia con la lectura de una tarjeta de control (READ), y si ésta no se refiere a un procedimiento estadístico (frecuencias, tablas de contingencia, etc.), él interpreta el campo de especificación y toma o realiza la acción que se le

⁹ Tomado del propio manual SPSS.

indica. La acción consiste en almacenar la información que contenga la tarjeta hasta que aparezca el subsecuente procedimiento de los datos al que se acopla. Así por ejemplo, una tarjeta de definición, INPUT FORMAT, que indica o señala los campos en los que se deberán leer los datos, es guardada hasta que es tiempo de leer los datos (READ INPUT DATA). Su acoplamiento permitirá la lectura correcta de las variables.

La terminación del proceso iterativo (LOOP) depende de la eficacia del propio programa para reconocer las distintas tarjetas que va encontrando. En el caso de no reconocer una tarea específica, nos manda un mensaje de ERROR, y sigue leyendo más tarjetas hasta reconocer alguna instrucción. Desde luego que un error, el cual puede ser una letra demás, o de menos, o equivocada, determina que no se procese ningún procedimiento y sólo continuará leyendo el maso de tarjetas que conforman el programa elaborado (con el que trabaja el SPSS) analizando cada instrucción en su sintaxis y señalando los errores encontrados.

En el siguiente diagrama se muestran los pasos de un procedimiento típico. La ejemplificación es sólo para mostrar su lógica. Para que sea específico le faltarían las instrucciones para llevar a cabo el procedimiento estadístico concreto, como frecuencias, medidas de resumen, etc., y la forma y áreas para el almacenamiento de los datos para procesamiento, así como la salida de los resultados.

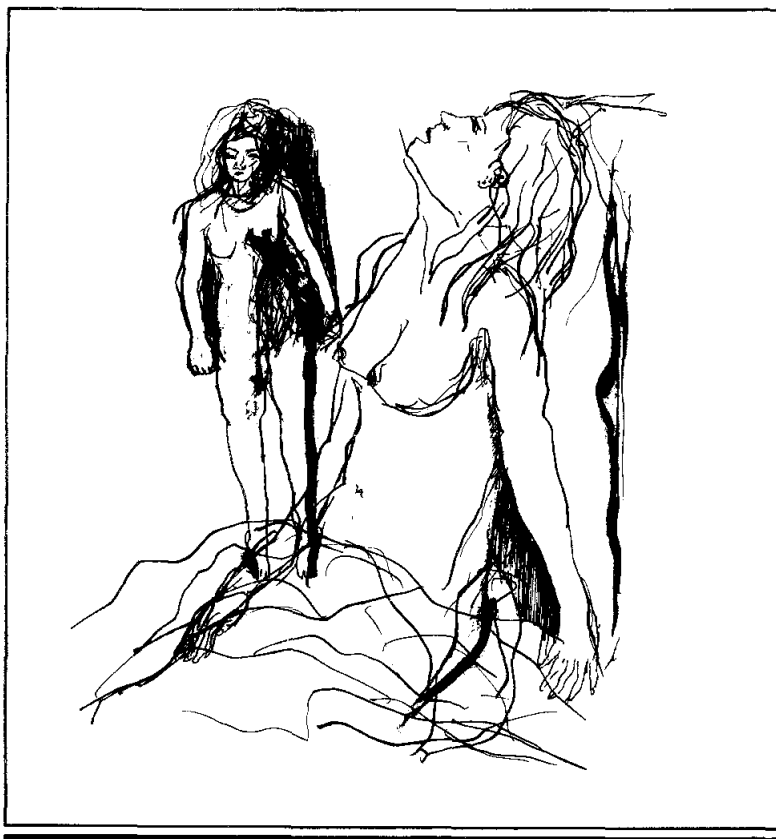
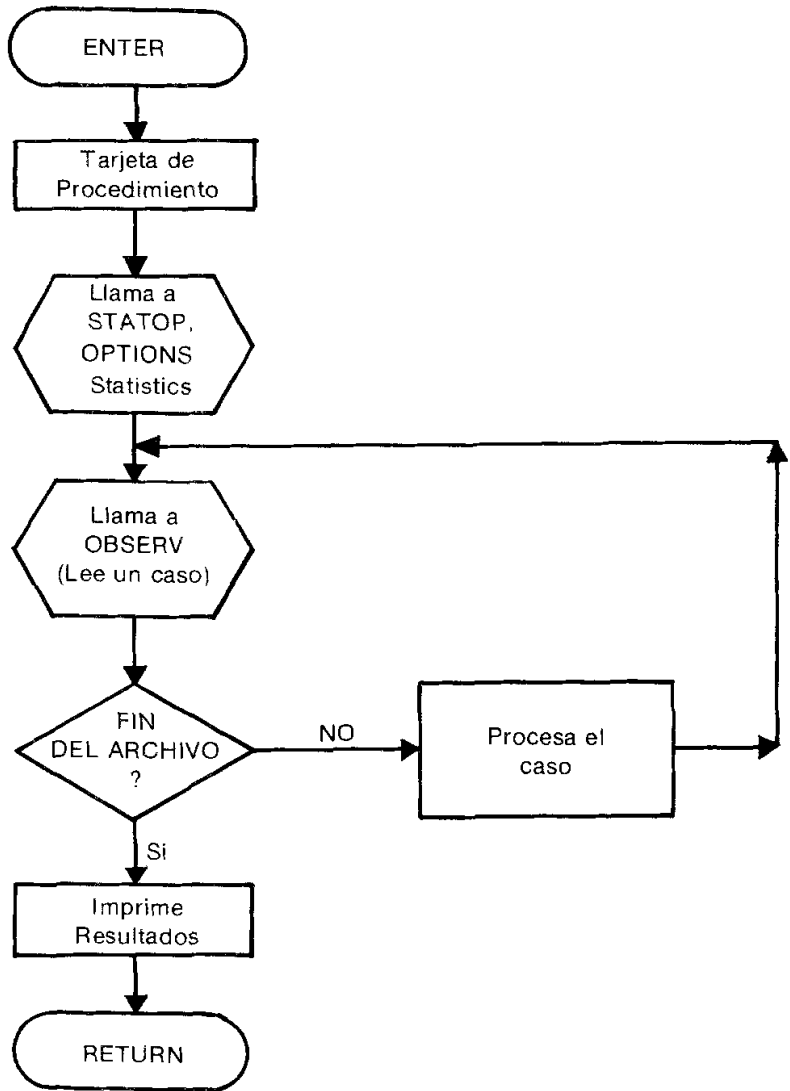


Diagrama de flujo del procedimiento típico ¹⁰



112

Con este diagrama se pretende dar la idea de las acciones que realiza cada procedimiento estadístico (subprograma). Se inicia con la información de las variables que el usuario desea que se consideren en el análisis estadístico. Se lee una tarjeta de procedimiento e inmediatamente el paquete llama un subprograma llamado STATOP, el cual accesa o jala el procedimiento estadístico que está mencionado en la tarjeta de procedimientos. Y sólo que se especifique lo contrario, dará los resultados de default previamente determinados.

Después de esto entra en un proceso iterativo (LOOP) llamando a la subrutina "OBSERV", la cual hace un enlace entre el procedimiento

¹⁰ Tomado del manual SPSS.

estadístico y el archivo de datos, proporcionando un caso a la vez, para su procesamiento, hasta complementar todos los casos. Existe una excepción a esta forma de trabajar, es con los procedimientos NOM-PAR y con el SCATTERGRAM.

Al salir del LOOP, o sea, al salir de realizar el procesamiento estadístico, se imprimen los resultados en un formato uniforme e igual para cada procedimiento en particular. Existen además algunas posibilidades de variar la forma de impresión, con tarjetas llamadas OPTIONS.

El RETURN implica el regreso al PC, el cual lee otra tarjeta de procedimientos o bien una de FINISH, con lo que termina la sesión y abandona al SPSS, quedando en el sistema de cómputo, desde el cual se está operando, para la realización de otras tareas.

3. El SPSS en el procesamiento de datos

3.1 La construcción de archivos y subarchivos

El SPSS es conducido a través de sus diversas funciones por medio de una secuencia de tarjetas de control. En el SPSS una tarjeta de control es referida como un registro de 80 caracteres reconocibles por una computadora. Es decir, una tarjeta como las IBM de 80 columnas, o la imagen de una tarjeta de registro proporcionada, vía terminal remota, u otro medio, como cinta magnética.

En computación, las unidades de análisis, determinadas en la investigación, se denominan "casos". Así que un caso puede estar referido a un individuo, o también puede representar a conglomerados de individuos referidos como ciudades, estados o naciones.

Cada caso está compuesto por todos los valores de las mediciones que se hayan hecho a las unidades de análisis y corresponden a las variables que son sujetas a procesamiento.

Cabe aclarar que el orden de las variables para distintos individuos, no sólo en SPSS, debe ser el mismo, para permitir su identificación, manejo y posterior procesamiento.

Los distintos valores de las variables deben ser codificados de tal manera que puedan ser identificados sin confusiones. Estos son alimentados a través de tarjetas perforadas u otros medios (vía terminal remota, cinta magnética, cinta de papel, etc.).

Los tipos de datos con los que se alimenta a la computadora pueden ser numéricos o alfanuméricos. Los primeros se basan en códigos o valores de la variable, que son números: enteros o con decimales. Los segundos son combinaciones de números y letras del alfabeto e incluso caracteres (*#\$%&, etc.).

La distinción es muy importante dado que su procesamiento es distinto —los datos alfanuméricos consumen un mayor tiempo, y no deben ser procesados en forma matemática. El mismo manual del SPSS recomienda no usar variables alfanuméricas, ya que al contar con pocos procedimientos para su procesamiento, se requiere de un mayor tiempo de máquina.

Al codificar los valores de las variables (ya sean numéricos o alfanuméricos —los primeros permiten su procesamiento con mayor facilidad—) las operaciones aritméticas hechas con dichos valores pueden o no tener sentido. Así que no basta con no utilizar códigos alfanuméricos para prever, entre otras cosas, cálculos sin ninguna forma de interpretación real.



Aunque una etiqueta de una variable puede ser alfanumérica (o sólo alfabética) como: sexo, ingreso, ocupación, etc., los valores de estas variables deben tener códigos numéricos para una mayor facilidad de manipulación. Nada nos limita a efectuar operaciones aritméticas con variables alfanuméricas (de hecho, si le es solicitado, la computadora busca un resultado y lo da como bueno), es decir, ni la computadora ni el paquete tienen forma de limitar el procesamiento, ni pueden distinguir si éste es o no pertinente. Queda en la ética del usuario el buen o mal uso de aquél.

Hasta ahora hemos dicho que la organización de los datos por caso deben tener el mismo orden, de igual manera los casos deben ser ordenados siguiendo una misma secuencia, para que la computadora pueda interpretar cuándo está leyendo un nuevo caso o qué variables o cuántas variables por caso debe leer.

En la definición de datos, para la creación de archivos, dos tarjetas son de gran importancia en esta tarea:

—VARIABLE LIST
—INPUT FORMAT

La primera tarjeta indica las variables para ser manejadas por el SPSS, tanto en número como en orden al modelo de su sintaxis. Es decir, instruye a la computadora para leer los datos e identifica, por medio del INPUT FORMAT, el campo para cada una de las variables en que está perforada.

Dependiendo del número de variables que se manejen, pueden ser suficiente una tarjeta o más de una. No es necesario indicar cuántas de ellas componen un caso, pues el SPSS instruye a la computadora para leer tantas variables como las que se mencionan en el VARIABLE LIST, y de acuerdo al INPUT FORMAT específico. El SPSS lleva un conteo del número de variables y por esto reconoce cuándo se completa un caso e infiere por esto que los siguientes datos que se le están proporcionando, corresponden al siguiente caso.

Esto ha permitido que, en ocasiones, se reduzca el número de tarjetas que se usan para la perforación de los datos cuando el número de variables y su formato no cubren una tarjeta en forma completa. Esto permite también perforar en una tarjeta dos o más casos. Esta situación ha creado dos tipos de formatos utilizados por SPSS, el fijo (FIXED) y el variable (FREEFIELD).

El formato fijo implica que los datos provienen de tarjetas que están perforadas en las mismas columnas (campo). Así, por ejemplo:

VARIABLES	CÓDIGO	COLUMNAS
Identificación	RFC o No. de cuestionario	1 a 10
No. de tarjeta*	1, 2, etc. (las necesarias)	11
Escolaridad	(según código)	12
Ocupación	(según código)	13
Ingreso	(según código)	14

*es necesaria sólo para control manual.

Una ventaja del formato fijo, independientemente de que se desperdicien columnas de una tarjeta, es la facilidad para checar los datos por columnas, pues de otra forma se traslapan los campos de las variables, dificultando la tarea de verificación visual de los datos.

Es recomendable entonces, siempre que no represente un gasto excesivo, utilizar un formato fijo.

En aquellas situaciones en que podemos perforar más de un caso en una misma tarjeta, se perforarán uno a continuación de otro, sin importar que algún caso especial pueda quedar en dos tarjetas consecutivas (v.g., el comienzo de un caso puede ser perforado al final de una tarjeta y el final del caso al comienzo de la siguiente). Esto no causa error, pues la computadora es instruida a través del SPSS para leer un número de variables por cada caso y el número de éstos que componen el archivo.

El único problema que podría presentarse es que, si algún dato o más de un caso particular no se perfora, la computadora tomaría el siguiente dato que encuentre sin importar a qué caso corresponda, ocasionándose un corrimiento de datos. Esto puede provocar que se mezclen los casos y que el reporte de resultados sea ininteligible o extraño, lo cual puede evidenciar errores. En otros casos podría darnos resultados que aparentemente estén correctos y que conlleven a errores, tanto de análisis como de interpretación.

Por otro lado, los datos son organizados dentro del SPSS en unidades llamadas archivos (FILES). Un archivo se compone de los datos del usuario, y se complementa con información asociada a ellos, que nos permiten describir y definir cada una de las variables, de la siguiente manera:

**Estructura de un programa para crear un
archivo SPSS
Definición de datos**

1.		80.
RUN NAME	Nombre de la corrida.	
FILE NAME	Nombre del archivo para futuras referencias.	
VARIABLE LIST	Lista de variables que conforman el archivo.	
INPUT FORMAT	Tipo de variables y su localización.	
N OF CASES	Número de casos que comprende el archivo.	
INPUT MEDIUM	Medio utilizado para introducir los datos (CARD, DISK, TAPE).	
VAR LABELS	Nombre o etiqueta de las variables.	
VALUE LABELS	Nombre o etiqueta de los valores de las variables.	
MISSING VALUES	Valores de las variables que no deseamos usar.	
PRINT FORMAT	Decimales de las variables a considerar en la impresión.	
READ INPUT DATA	Instrucción para que SPSS comience a leer datos.	

DATOS

```

XXXXXXXX XXXXXX XXXXXXXX XXXXXXXX XXXXXXXX ..... XXXX
XXXXXXXX XXXXXX XXXXXXXX XXXXXXXX XXXXXXXX ..... XXXX
.....
XXXXXXXX XXXXXX XXXXXXXX XXXXXXXX XXXXXXXX .....
    
```

SAVE FILE	Condición para hacer permanente un archivo.
FINISH	Indicación para terminar una corrida SPSS.

Un programa SPSS está estructurado por diferentes tipos de tarjetas:



1) *De control*, como son RUN NAME, READ INPUT DATA, SAVE FILE, FINISH, que tienen las siguientes funciones:

RUN NAME.- Sirve para identificar la corrida, por ejemplo, mencionar que se refiere el programa a la creación del archivo de una determinada investigación, del usuario o a una práctica. Esta tarjeta es opcional, es decir, puede incorporarse al programa o no.

Su aparición hará que el título, nombre del usuario o número de la corrida, aparezca siempre en la parte superior de cada una de las hojas de computación en las que se imprima el reporte o tarea solicitada al SPSS.

READ INPUT DATA.- Instruye al SPSS para comenzar a leer los datos. Los cuales son leídos de un dato en un dato, y al mismo tiempo procesados, de acuerdo al procedimiento estadístico requerido en el programa SPSS.

SAVE FILE.- Ordena al paquete SPSS conservar el archivo en forma permanente, pero no inmutable. Y así poderlo usar tantas veces como sea necesario. En caso de que esta instrucción no aparezca, el archivo construido desaparecerá tan pronto como termine el último procedimiento que se haya programado.

FINISH.- Informa al paquete que se han terminado las tareas encomendadas y debe transferir los controles al sistema operativo de la computadora, de la cual se genera el proceso.

2) *Definición de datos*, por medio de las cuales se construye el archivo con el que va a trabajar: FILE NAME, VARIABLE LIST, INPUT FORMAT, N OF CASES, que son obligatorias, las restantes, VAR LABELS, VALUE LABELS, MISSING VALUES y PRINT FORMATS son opcionales, es decir pueden aparecer o no.

3) *Definición de procedimientos*, que se refieren al procedimiento estadístico que demande el usuario y que se encuentran disponibles en el paquete, como: FREQUENCIES, CROSSTABS, PEARSON CORR, etc. Cada una de estas tareas tiene una especificación especial y en ocasiones diversas formas de operar, que hay que conocer para su mejor utilización. Por lo general los procedimientos estadísticos se acompañan de dos instrucciones más que son opcionales: OPTIONS y STATISTICS.

OPTIONS.- Opciones que al llamarse pueden generar dos tipos de tareas: la primera, eliminar títulos, nombres, etiquetas o hacer caso omiso de los MISSING VALUES; la segunda, indica la selección de un determinado método de cálculo y/o la inclusión de un tipo especial de reporte final.

STATISTICS.- Con la que se puede solicitar la impresión de algunas estadísticas complementarias al procedimiento estadístico que se esté procesando, o incluso algunas tareas intermedias que el mismo paquete esté utilizando.

Cabe aclarar que una vez construido un archivo, éste puede ser usado tantas veces como sea necesario —incluso permanecer “indefinidamente”—, y que las instrucciones requeridas para crear y procesar los archivos de datos, o efectuar algún procesamiento estadístico a partir de tarjetas, cinta magnética o disco, son iguales en género y número.

Con frecuencia, en la investigación social se recaban dos o más muestras con el objetivo de ser analizadas en forma independiente y, posteriormente, comparar sus resultados y comprobar o no sus diferencias estadísticamente.

En el SPSS estos grupos así formados pueden ser almacenados con una estructura de Subarchivos (SUBFILES). Por ejemplo, una encuesta de participación política puede ser aplicada en distintas ciudades de la república, o aplicada a individuos con distintas condiciones económicas, políticas o sociales. Otro ejemplo podría ser el Censo General de Población, que se conforma con los datos de todos los mexicanos provenientes de los diferentes estados. Estos en sí pueden conformar subarchivos. A su vez, cada estado está compuesto por municipios, y éstos por localidades que pueden conformar también una estructura en subarchivos.

El SPSS cuenta con la posibilidad de estructurar un archivo segmentado en subarchivos y ser operados, procesados y/o manipulados de múltiples maneras, permitiendo cualquier análisis estadístico de contrastes.

En la investigación empírica el problema que se plantea es cómo medir para poner a prueba las hipótesis de investigación.

Así, un archivo estructurado en subarchivos permite:

Primero.- Ser usado para examinar la distribución estadística, o para comparar sus valores promedio a través de una prueba de hipótesis de diferencia de medias. Es decir, el SPSS identifica la tarjeta de control SUBFILES como una variable de control.

Segundo.- Hacer comparaciones de la relación entre variables en los diferentes subfiles. Es decir, esta estructura permite la realización independiente de cualquier proceso, para cada uno de los subarchivos, o cualquier agrupación de subarchivos.

Tercero.- Seleccionar algunos subarchivos para ser procesados en forma individual, en combinaciones o incluso en forma conjunta, ignorando la estructura de subarchivos.

Cuarto.- Seleccionar algunos subarchivos para ser procesados en forma individual, en combinaciones o incluso en forma conjunta, ignorando la estructura de subarchivos.

Quinto.- Al final de cualquier corrida, los casos que componen un archivo pueden ser reordenados, y puede surgir una nueva estructura de subarchivos que concuerde con los nuevos requerimientos de la investigación.

Finalmente, las restricciones que pueden tener los subarchivos son:

—Los subarchivos deberán tener el mismo formato y las mismas variables.

—No podrán ser definidos más de 100 subarchivos.

—No se podrán definir más de 500 variables.

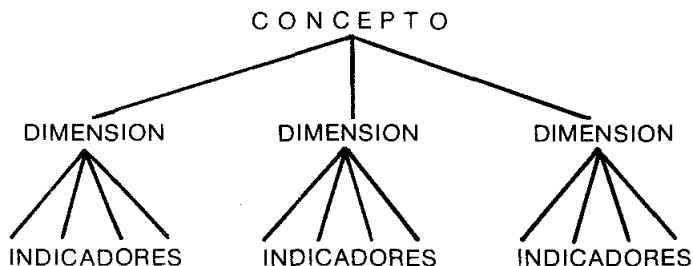
Sin embargo, el número de casos de los subarchivos pueden ser diferentes para cada uno de ellos.

Un archivo creado en el SPSS, independientemente de su estructura, podrá ser permanente o no, según intereses y requerimientos del usuario. Pero independientemente de su permanencia, éste no es inmutable, es decir, que puede ser alterado, agregando más datos o más variables, o incluso borrando datos primarios que ya no serán utilizados posteriormente, muchas de las veces restringidos por la capacidad de memoria de la computadora, o del disco que estemos utilizando.

3.2 El SPSS en el manejo de datos

La investigación empírica social, puede ser considerada como un proceso integrado por partes relacionadas, donde las técnicas — tanto de recolección de datos, como de análisis— deben responder adecuadamente a los planteamientos teóricos del problema de investigación.

En este tipo de investigación, el problema que se plantea es cómo medir para poner a prueba las hipótesis de investigación. Algunos metodólogos como Paul Lazarsfeld,¹¹ han propuesto mecanismos generales, cuyo esquema es ampliamente conocido.



Esta propuesta es interpretada como el fraccionamiento o partición de un concepto, en tantas dimensiones como se requieran. Incluso éstas pueden dividirse además en subdimensiones, para finalmente terminar en los indicadores.

Los indicadores son las preguntas que conforman nuestro instrumento de recolección de datos, cuestionario, cédula, etc.

El proceso de medición requiere además de otro proceso, inverso a éste. Es decir, de compactación o agregación de los indicadores. Este procedimiento se lleva a cabo mediante la construcción de un índice, el cual posteriormente recibirá un tratamiento de variable.

El paso de la teoría a su comprobación ha requerido de la generación de evidencias empíricas o de datos; su manejo, su almacenamiento, en suma, su procesamiento. Pero no sólo eso, sino que este hecho ha obligado al diseño y elaboración de un conjunto de técnicas que nos permitan resumir la información y/o poner a prueba la hipótesis de investigación.

¹¹ Citado por, Padua, Jorge. *Técnicas de investigación aplicada*. F.C.E., México, 1965.

El procesamiento de datos tiene como función manejar las evidencias empíricas generadas para un propósito particular, y puede decirse que comprende tanto el manejo de datos, como el cálculo de la estadística de la prueba.

El manejo de datos es la utilización lógica y ordenada de ellos; y el cálculo, se refiere a los procesos técnicos que se deben realizar, y se conocen como técnicas estadísticas.

En suma, el procesamiento de datos se refiere a su registro y manejo para convertirlos en forma útil a la investigación.

Recordemos que los datos son hechos registrados de la realidad, o evidencias empíricas; estos son la materia prima de la cual se derivará la comprobación o refutación de nuestras hipótesis de investigación. La conversión de los datos, manejo y cálculo, es función del procesamiento de datos.

El SPSS está diseñado para llevar a cabo las tareas de procesamiento de datos más usuales en la investigación empírica. Para ello realiza distintas tareas, que se ordenan a partir de otras tarjetas, como las ya mencionadas de definición de datos, y pueden realizarse las siguientes tareas: modificación de datos, selección de datos, modificación de archivos, y procesamientos estadísticos.

A. *Modificación de datos*

Las tarjetas de modificación de datos, nos permiten cambiar o reagrupar los distintos valores de las variables de la investigación y de un archivo SPSS.

El SPSS sigue, en su presentación, lo que podría decirse que es el análisis clásico de la investigación empírica.

Estas modificaciones están orientadas a resolver distintas tareas: recodificar y transformar variables en otras variables.

El primer procedimiento, RECODE, es utilizado para recodificar las variables, permitiendo reagrupar los valores

de una variable en particular, ya que en ocasiones, el número de categorías —distintos valores— de una variable, puede ser grande o no contar con el número de casos suficientes requeridos para el análisis estadístico. Las variables pueden reagruparse sumando categorías, bajo criterios conceptuales. Así por ejemplo, una de las tareas más generalizadas; es la recodificación de variables continuas en discretas (con muchos o con pocos valores o categorías), como podría ser el recodificar los ingresos declarados en una encuesta en grupos de ingresos, formando nuevas categorías de interés analítico. Por ejemplo, en grupos formados con base a los salarios mínimos. Otro ejemplo, podría ser, la recodificación de las edades de los entrevistados en grupos quinquenales, decenales, etc.; tal y como los usan los Censos Generales de Población.

B. *La transformación de datos*

Las transformaciones realizadas a través de expresiones nemónicas como: COMPUTE, IF y COUNT, nos permiten reconstruir variables, a partir de las variables que componen un archivo SPSS.

La expresión COMPUTE (calcule) ordena la realización de operaciones con variables; haciendo uso de funciones matemáticas expresadas en lenguaje de programación FORTRAN (FORmula TRANslation), como podría ser un índice sumatorio.

Como ejemplo, podríamos suponer que el costo de vida está conformado por distintos costos: de alimentación, de vivienda, de vestido y otros, considerados en una encuesta y grabados en un archivo SPSS. La generación de una nueva variable (COSTVIDA), costo de vida que los incluye sería:

1.	16.	80.
<hr/>		
COMPUTE	COSTVIDA=COSALI + COSVID + COSVES +	
	COSOTRO	

La expresión IF es un SI condicional o restricción, que se hace para que se cumplan ciertos requisitos predeterminados. La condición de valores o de categorías de variables en las que se relacionan dos cantidades algebraicamente a través de operadores relacionales se expresan en forma lógica como: Mayor que (GT, Igual a (EQ), o Menor que (LE). A su vez, estas expresiones (dos o más) pueden estar unidas por medio de operadores lógicos: AND, OR, NOT; con las que unimos, separamos o negamos las expresiones.

Supóngase que deseamos construir una variable llamada Capacidad de Ahorro (CAPAHO), la cual tendrá valores: Bajo (1), Medio (2) y Alto (3), de la siguiente manera:

1.	16.	80.
<hr/>		
IF	(INGRESO - (TAMFAM * 100000) LE 100000) CAPAHO=1	
IF	(INGRESO - (TAMFAM * 100000) GT 100000) AND	
	(INGRESO -(TAMFAM * 100000) LE 500000) CAPAHO=2	
IF	(INGRESO - (TAMFAM * 100000) GT 500000) CAPAHO=3	

La expresión COUNT (contar) nos permite la construcción de índices sumatorios, pues esta instrucción nos permite contar el número de veces en que un determinado valor de la variable aparece para cada caso, y es sumado a otra variable previamente definida. Esta expresión puede ser utilizada para sumar las respuestas a indicadores que conforman una dimensión. Supóngase que, a manera de ejemplo, contamos con la información acerca de los periódicos que consultan los entrevistados de una encuesta, y que nos interesa construir un índice que nos informe acerca de los entrevistados que recurren a más fuentes de información periodística para estar bien informados, este índice se llamaría MAXINF.

1.	16.	80.
<hr/>		
COUNT	MAXINF = JORNADA + UNO+UNO + EXCELSIOR	

C. Selección de datos

Las tarjetas de selección de datos nos permiten realizar segregaciones de un archivo SPSS, a través de la elección de una muestra aleatoria, o generar una muestra intencional, con aquellos casos que cumplan con una condición específica de interés analítico para la investigación. Pero también cuentan con un procedimiento para ponderar los casos de un archivo (cuando la investigación es llevada a cabo, a través de una muestra), y así poder hacer inferencias a toda la población.

La expresión SAMPLE (muestra) nos provee una muestra aleatoria generada del archivo SPSS en cuestión. El tamaño de ella se determi-

na en función del número de casos que se requiera trabajar y se expresa en números relativos. Este factor está comprendido entre "1.00 y 0.00", teniendo una representación porcentual. Así, podemos elegir en forma aleatoria una muestra de 10%, 15% o 20% por ejemplo, de acuerdo a nuestras necesidades, las que pueden estar determinadas por el tipo de análisis estadístico que se desea realizar. Así, por ejemplo, algunos procedimientos estadísticos, como las tablas de contingencia y el análisis de varianza, requieren de un número de casos mínimos por celda.

La expresión SELECT IF (seleccionar bajo cierta condición) nos provee una muestra intencional, en función de un criterio determinado, que se fija respecto a un valor de una variable, o grupo de variables, a través de expresiones del siguiente tipo:

```
1. _____ 16. _____ 80. _____
SELECT IF      (OCUP EQ 2 OR OCUP EQ 5)
```

Lo cual generaría un subarchivo comprendido con todos aquellos casos que cuenten con la condición de tener la ocupación 2 ó 5, según códigos correspondientes.

El procedimiento (WEIGHT) es utilizado para ponderar, o dar a cada caso su correspondiente peso relativo, y hacer inferencias a la población de la cual se extrajo la muestra. Recuérdese que los datos de la muestra no interesan en sí, sino lo que con ellos podemos inferir de toda la población; y que el uso del muestreo descansa, en términos generales, en el ahorro de tiempo y dinero, bajo el supuesto del comportamiento estadístico de los datos. A través de la instrucción WEIGHT, se le puede asignar un peso específico a cada caso de una muestra. Normalmente la ponderación sigue un procedimiento expresado a través de instrucciones IF o COMPUTE.

```
1. _____ 16. _____ 80. _____

IF      (SECTOR EQ 1) WTFACTOR=1.25
IF      (SECTOR EQ 2 AND SECTOR LE 6)
        WTFACTOR=2.75
IF      (SECTOR GT 5) WTFACTOR=3.0
WEIGHT  WTFACTOR
```

Como podemos observar, en este caso estamos ponderando en forma diferencial cada caso de la muestra, por ejemplo, según el sector de la ciudad en la que se llevó a cabo la encuesta.

Con las tarjetas de modificación de archivos se actualizan éstos agregando nuevos casos, o nuevos datos, o borrando los datos que no vamos a utilizar más, o incluso reordenándolos de acuerdo a nuevas necesidades de investigación.

LIST FILEINFO (expresión nemónica que hace referencia a listar la información relacionada con el archivo SPSS) nos provee un reporte completo de todas o de algunas definiciones de datos, utilizadas en la construcción de los archivos SPSS. Puede estar orientado a conocer qué variables contiene el archivo, qué etiquetas están asociadas a ellas, su orden, etc.



El SPSS es un conjunto de procedimientos y rutinas para realizar el procesamiento de datos, lo cual se inscribe en la metodología empirista.

Con la tarjeta LIST CASES, se solicita al sistema que liste un grupo de casos, incluso su totalidad. En otras palabras, nos permite conocer el contenido de cada caso para poder verificar, con ellos mismos, la construcción de nuevas variables y, en general, vigilar que se esté leyendo adecuadamente.

Por último, WRITE CASES y WRITE FILEINFO tienen funciones similares a las anteriores instrucciones, pero su objetivo es proporcionar información del archivo a través de otros medios: tarjetas perforadas, cinta magnética, para que los datos puedan ser utilizados ya sea con otros programas, o con otros sistemas de análisis de datos y/o computadoras.

Cabe aclarar que ambas rutinas no producirán resultados si un procedimiento estadístico no es invocado en esa corrida, pues éstos son los que permiten el acceso a los datos de un archivo SPSS.

3.2. El SPSS en el cálculo de las estadísticas

Quizás la parte más conocida del SPSS, es su tratamiento estadístico con el que se manejan los resultados de una investigación empírica. Esto es porque indudablemente la estadística es una herramienta indispensable en el manejo de datos en casi todas las áreas del saber,

como lo confirma el hecho de que la estadística, actualmente, forma parte de casi todos los programas de posgrado, y se enseña en áreas tan disímiles como medicina y sociología. Sin embargo, hasta ahora se ha cuestionado el por qué de esta importancia. No queremos desviarnos de nuestro tema principal, así que sólo mencionaremos dos de los elementos sobre los que descansa esta importancia: primero, la estadística forma parte del proceso de investigación, pues con ella se lleva a cabo el análisis de los datos con fines de comprobación teórica. Y segundo, por ser la estadística un conjunto de técnicas y métodos matemáticos que son de aplicación general, en el análisis de datos, independientemente de la ciencia en que se aplique.

El SPSS en esta parte responde al planteamiento general de la estadística. Es decir, en términos de describir e inferir sobre la base de una muestra. Su lógica de selección descansa en el nivel de medición (nominal, ordinal, intervalo, razón y absoluta), resultante de la propuesta para medir el fenómeno de interés.

El SPSS en su versión más conocida (sexta), cuenta con 16 procedimientos estadísticos (subprogramas): FREQUENCIES, CONDESCRIPTIVE, CROSSTABS, BREAKDOWN, T-TEST, PEARSON CORR, NOMPARR CORR, SCATTERGRAM, PARTIAL CORR, REGRESSION, ANOVA, ONEWAY, DISCRIMINANT, FACTOR, CANCORR Y GUTTMAN SCALE. Estos procedimientos estadísticos pueden ser utilizados en combinación con opciones (OPTIONS) y estadísticas (STATISTICS), que de hecho permiten cambiar: formas de impresión, utilización de diferentes modelos o métodos, y, proporcionan información estadística complementaria a la elaborada. Con todo esto, el SPSS cubre un conjunto de técnicas y métodos que contienen los textos especializados de la materia, como la obra *Estadística Social*, de Blalock.¹² Pero incluso cuenta con procedimientos como la cons-



¹² Blalock, Hubert Jr. *Estadística social*. F.C.E., México, 1975.

trucción de escalas, que no forma parte de la estadística, o de temas que por su complejidad o especialidad no son tratados en un texto general y que requieren, en la mayoría de los casos, un tratamiento particular, como podría ser el análisis factorial o el discriminante.

El SPSS sigue, en su presentación, lo que podría decirse que es el análisis clásico de la investigación empírica: primero, la descripción de los datos y luego los procedimientos de estadística inferencial; segundo, la lógica práctica del análisis de los datos, iniciándose en el análisis univariado (de variable por variable) y posteriormente, después de las transformaciones y recodificación de las variables, iniciar el análisis de relaciones entre variables que pueden ser bivariadas o multivariadas. Así, en toda investigación empírica como sería una encuesta, la primera tarea que podría decirse obligatoria, es la construcción de distribución de frecuencias simples, relativas y acumuladas, las cuales tienen múltiples usos: checar códigos para poder corregir posibles errores de captura; observar el comportamiento de cada una de las variables para su reconstrucción, en su eliminación o separación para estudios de pauta.

Para la realización de esta primera tarea estadística, el SPSS nos provee dos subprogramas: FREQUENCIES y CONDESCRIPTIVE.

1.	16.	80.
FREQUENCIES	GENERAL= (lista de variables o ALL)	
OPTIONS	(selección de una o varias que no se contrapongan)	
STATISTICS	(selección de una variable o incluso ALL)	

La instrucción anterior instruye al sistema para generar las distribuciones de frecuencia simples y acumuladas para cada una de las variables mencionadas o de todas, haciendo uso del término ALL.

La tarjeta OPTIONS, correspondiente al programa invocado, contiene 9 distintas opciones de salida, permitiendo diferentes tipos de impresión de resultados, eliminación de etiquetas, construcción de histogramas e incluso elaborar un diccionario que contenga, nombre de variables y página correspondiente, de acuerdo al reporte impreso. Pueden solicitarse, una o varias opciones, pero siempre que éstas no se contrapongan.

Por otra parte, una tarjeta STATISTICS, produce distintas estadísticas de resumen (media, error estándar, mediana, moda, desviación estándar, varianza, kurtosis, asimetría, rango, valor mínimo y máximo observado) pudiéndose solicitar todas ellas.

Otro programa que nos permite realizar análisis descriptivos de datos es el CONDESCRIPTIVE que, combinado con una tarjeta STATISTICS, produce un reporte compacto, solamente de las medidas de resumen seleccionadas. Queremos hacer notar que ni la computadora, ni el paquete SPSS, pueden diferenciar entre los distintos niveles de medición y producirán las estadísticas requeridas sin importar su pertinencia, esta es responsabilidad absoluta del usuario.

Otro subprograma que permite un análisis descriptivo más fino es el BREAKDOWN (romper hacia abajo). Produce algunas de las estadísticas de resumen señaladas anteriormente para variables continuas (variables criterio) que pueden calcularse para otras variables (variables control) que, incluso, tengan el nivel nominal u ordinal. Así por ejemplo, un BREAKDOWN podría calcular el ingreso promedio de individuos o el ingreso mediano, tomando en consideración las com-



binaciones producidas entre otras variables como podrían ser población, ocupación y sexo, bajo el supuesto, "en este ejemplo", de trato salarial diferencial por población, ocupación y sexo. Este subprograma puede ser utilizado en combinación con el T-TEST, con el cual se pueden poner a prueba las diferencias encontradas en el BREAK-DOWN.

Los procedimientos siguientes que contiene el SPSS no tienen una secuencia lógica necesaria, es decir, el pasar de análisis de dos dimensiones, luego de tres y a "n" variables, pues todo esto depende de los objetivos de la investigación.

El manual continúa con los procedimientos en donde se relacionan dos o más variables y dependiendo del nivel de ellas, el usuario debe seleccionar las técnicas estadísticas adecuadas. Así por ejemplo, para variables nominales se pueden utilizar las Tablas de Contingencia, para medir asociación entre variables. El procedimiento se llama CROSSTABS. En cambio, si estas variables son ordinales es posible determinar su correlación utilizando para ello cualquiera de los dos procedimientos no paramétricos: la correlación de Spearman o el Coeficiente de Kendall, contenidos en el subprograma NOMPARR. Mientras que si las variables son intervalares, o con un nivel de medición mayor, se puede utilizar el coeficiente de correlación de Pearson, o coeficiente de correlación producto momento (PEARSON CORR).



El SPSS contempla muchos otros procedimientos estadísticos, como el análisis de varianza, simple o factorial (ONEWAY y ANOVA); el análisis de regresión (REGRESSION); tópicos especiales que nos permiten analizar relaciones no lineales, la utilización de variables mudas (DUMMY) y el PATH (análisis o de sederos) para la interpretación causal de los fenómenos.

Sería muy largo seguir describiendo los demás procedimientos como del Análisis Factorial, el Discriminante, etc. y aún más si mencionamos los nuevos procedimientos con que cuentan las versiones 7 y 8, pero consideramos que es bastante claro que el SPSS es una ayuda inconmensurable para el análisis de datos.

Cabe hacer una última aclaración, que todos estos procedimientos tienen una limitación: están diseñados para el manejo de quinientas variables como máximo. Sin embargo el SPSS provee los procedimientos para el manejo de archivos mayores.

III. El Mito

Si recordamos que el SPSS fue desarrollado debido a una frustración colectiva de investigadores, podría pensarse que el paquete vino a resolver la problemática que lo originó. A decir verdad, resolvió el problema de programación de una manera muy eficiente, que es, según nosotros, parte importante del problema, pero pensamos que esto no es todo.

Hemos visto que el SPSS es un conjunto de procedimientos y rutinas para realizar el procesamiento de datos, el cual se inscribe en la metodología empirista.

El SPSS no es un sustituto de la comprensión del objetivo, ni de la lógica del proceso mismo de investigación.

Si bien se ha buscado que el SPSS contemple todos aquellos procedimientos que pudieran necesitarse en casi cualquier situación o circunstancia, el sistema requiere, para realizar cualquier tarea, que se instruya, es decir, que se

indique qué hacer. Esta palabra, indicar, es la clave de parte del mito de las computadoras, pues la selección de un determinado procedimiento requiere de la toma de decisión previa, y esta implica el conocimiento y comprensión de la lógica de la investigación empirista.

Es ampliamente conocido que la sociología norteamericana, apoya en lo general a sus investigadores fundándose en un fuerte manejo de técnicas estadísticas y, en general, en el procesamiento de datos. Éste llegó a convertirse en un elemento de "validez científica".

Por el contrario, en nuestros Centros de Educación Superior, se abandonó la investigación empírica mucho antes que se pudiera decir que se haya constituido como una práctica más o menos permanente, tanto por razones teóricas y metodológicas, como por las propias condiciones económicas y tecnológicas imperantes, al menos en las Ciencias Sociales.

En la actualidad no son un secreto los intentos para retomar nuevamente la investigación empírica, no como antaño, para sustituir a otras, sino recuperando sus alcances y reconociendo sus limitaciones.

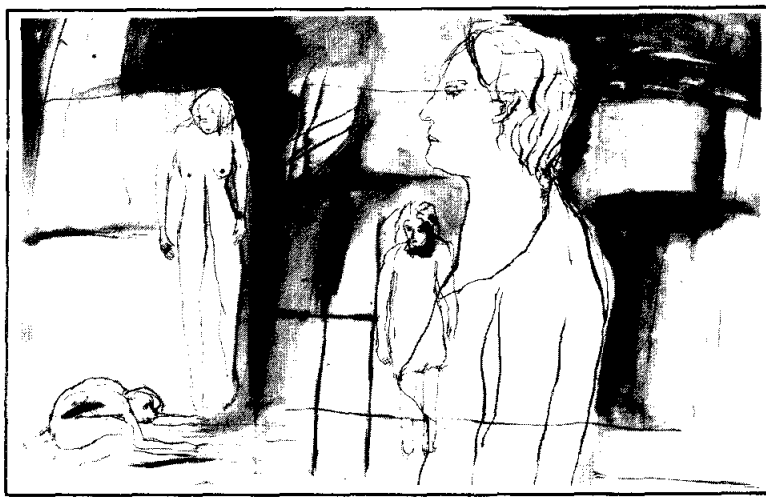
Sin embargo, una de sus limitantes académicas y de aplicación ha sido el manejo estadístico de los datos o evidencias empíricas, que como parte de la matemática ha presentado problemas de diversa índole.

La problemática del uso de la estadística arranca desde su aprendizaje. Su dificultad se ha relacionado con lo abstracto de la matemática, ya que esta Ciencia ha evidenciado serios problemas en el proceso de enseñanza-aprendizaje a nivel mundial. Y en forma lineal se ha inferido el origen del problema.

En este contexto, el SPSS viene a resolver el cálculo en estadística, liberando con ello parte de la problemática, pero esto no es todo; vayamos por partes:

Primero. Efectivamente, la enseñanza de la Estadística ha creado problemas, los que hasta ahora no han sido solucionados satisfactoriamente.

Segundo. La didáctica utilizada para su enseñanza ha sido hasta ahora similar a la de la matemática, que ha venido privilegiando la mecanización en contra de su comprensión.



Tercero. Si por un lado el SPSS nos ha liberado del cálculo, por el otro ha evidenciado la dificultad de cómo analizar los resultados.

Cuarto. El problema relacionado con el análisis de los resultados, estriba en el conocimiento y dominio de la lógica del procedimiento en cuestión.

Quinto. La selección de un procedimiento cualquiera, implica no sólo la comprensión de la técnica, sino de la lógica de ella y de la investigación empírica, que es su marco y contexto.

La no comprensión de esto, ha ocasionado el nacimiento de un nuevo mito, resultado de la simbiosis de la "magia" de la computadora y del problema aparente de la estadística, su cálculo.

Este mito se ha enmascarado en la problemática que encierra el conocer y manejar en primera instancia a la computadora y, en segunda, el aprendizaje del SPSS. Esto ha diferido su reconocimiento y nuevamente comenzará a crear frustración, si no es desmistificado a tiempo.

En otras palabras, el paquete SPSS vino a diferir el problema del aprendizaje de la Estadística por un lado, pero por otro, ha demostrado que no es suficiente contar con un sistema de programas como el SPSS, pues ¿cómo habremos de analizar los resultados de un reporte de computadora? y lo que es más importante ¿cómo seleccionaremos las técnicas estadísticas apropiadas?.

En suma, considero que el problema que se pretendió resolver, el de la programación, fue ampliamente resuelto, pero no considero que el SPSS haya sido desarrollado para resolver algo que vaya más allá de lo mecánico. Es decir, no es un sustituto de la comprensión del objetivo que se persigue con cada técnica, ni de la lógica del proceso mismo de investigación. Esto es lo único que nos permitiría comprender, y por ende analizar, los resultados complejos de los reportes SPSS.

Parece ser que el mito de las computadoras todavía está lejos de ser eliminado, pero el tiempo se encargará de diferenciar entre la gran ayuda que nos prestan, como herramienta moderna y versátil, y su capacidad para resolver por sí sola los problemas que plantea la investigación. Aún, desafortunadamente, está muy lejos la generación de computadoras que piense por nosotros.