

INTRODUCCION AL PAQUETE ESTADISTICO PARA EL ANALISIS Y CONSTRUCCION DE DATOS EN LAS CIENCIAS SOCIALES

Alejandro Calatayud Arcos[•]

EL PAQUETE ESTADISTICO para las Ciencias Sociales (SPSS)¹ fue desarrollado debido a una frustración colectiva de investigadores; podría pensarse que el paquete vino a resolver su problemática. A decir verdad, resolvió el problema de programación de una manera muy eficiente, ya que es una *parte importante del mismo, pero no abarca toda la problemática de la investigación.*

El SPSS es un conjunto de procedimientos y rutinas para realizar el procesamiento de datos, que luego el investigador debe interpretar teóricamente, tarea que se inscribe en la metodología empírica.

Si bien se ha buscado que el paquete contemple todos aquellos procedimientos que pudieran necesitarse en casi cualquier situación

[•] Investigador del Centro de Investigaciones y Servicios Educativos y profesor de la FCPyS, UNAM.

1. Statistical Package for the Social Sciences.

o circunstancia, el sistema requiere que se le instruya para realizar cualquier tarea, es decir, que se le indique qué hacer. La 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, e implica el conocimiento y comprensión de la lógica de la investigación empírica y de su para qué teórico.

Es ampliamente conocido que la Sociología norteamericana apoya a sus investigaciones, fundándose en un fuerte manejo de técnicas estadísticas y en general en el procesamiento de datos. Este 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 hubiera 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, de la cual fue víctima en su tiempo, sino recuperando sus alcances y reconociendo sus exigencias y limitaciones.

Sin embargo una de sus exigencias 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 su 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 no es toda la solución; vayamos por partes.

Primero: Efectivamente la enseñanza de la Estadística ha creado problemas no sólo en las ciencias sociales, los que hasta ahora no han sido solucionados satisfactoriamente.

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

Tercero: Si por un lado el SPSS nos ha liberado del trabajo de 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 estadístico y de sus bases matemáticas en cuestión.

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

La incompreensión 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 enmarcado en la problemática que encierra el conocer y manejar en primera instancia a la computadora y, en segundo, el aprendizaje del SPSS. Esto ha diferido su reconocimiento y nuevamente comenzará a crear frustración, si no es desmitificado a tiempo.

En otras palabras, el paquete vino a diferir el problema del aprendizaje de la Estadística por un lado, mientras que por otro, que aparecerá tarde o temprano, la cuestión de que no es suficiente contar con un sistema de programas como el SPSS, pues el problema sigue siendo: ¿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 con el SPSS fue el de la programación y quedó ampliamente resuelto; pero no considero que haya sido desarrollado el SPSS, para resolver algo que va más allá de lo mecánico, es decir, sustituir la comprensión del objetivo que se persigue con cada técnica y la lógica del proceso mismo de investigación, recursos indispensables que nos permitirán

comprender y por ende analizar los resultados complejos de los reportes del paquete.

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 la herramienta más moderna y versátil y su capacidad para resolver por sí solas las rutinas de investigación planteadas. Pero aún, desafortunadamente, está muy lejos la generación de computadoras que piensen por nosotros y resuelvan nuestros problemas de investigación por sí mismas.

El extendido uso de las computadoras para almacenar información (archivo), para efectuar clasificaciones, tabulaciones o construir otros datos por medio del cálculo con esa misma información, ha creado una imagen mítica respecto de la computación al servicio de la ciencia. Se une a este hecho la circunstancia de que, en el ejercicio del poder, la toma de decisiones basada en una información completa y confiable aumenta la probabilidad de su eficacia.

La confluencia de estas dos realidades extiende con más fuerza el mito entre quienes no conocen cuáles son las posibilidades de una máquina computadora, cuáles son los requerimientos para crear una afirmación científica que se base en información fidedigna, y cuál es la lógica de la toma de decisiones. El desarrollo de la computación ha generado facilidades (*software*) para hacerla accesible a un número cada vez mayor de usuarios, a través de los "paquetes" como el *Word Star*, *Dbase*, *Hojas de Cálculo*, etc., aplicables a infinidad de campos.

Este artículo tiene como finalidad esclarecer, para quienes construyen datos en las ciencias sociales, en qué consiste el paquete estadístico, SPSS.

Agradezco a todos aquellos compañeros universitarios, cuyos comentarios ayudaron a mejorar este artículo; muy en especial a las maestras Guadalupe Acevedo de la FCPyS, a Rosa María Rubalcava del Colegio de México y al Ing. Ernesto Arrijo de MICRODISA, por sus inapreciables sugerencias que indudablemente hicieron que la primera versión del artículo que ahora presentamos, mejorara notablemente.

DESCRIPCION Y ESTRUCTURA DEL PAQUETE ESTADISTICO PARA LAS CIENCIAS SOCIALES (SPSS)

EL PAQUETE ESTADISTICO

En el propio manual del spss, éste se define como "un sistema integrado por programas de computadora diseñado para las ciencias sociales".²

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, en busca de la solución de sus problemas, aunque con pleno desconocimiento de la materia, tanto de la computación como de la Estadística.

Para hacer la definición más entendible o accesible a una gran mayoría de interesados, trataremos de explicarla.

Tomemos la primera parte de la definición: "sistema integrado por programas de computadora...". Puede entenderse como una colección de programas interrelacionados o integrados y almacenados en conjunto. En otras palabras, un grupo de programas, donde cada uno de ellos realiza una tarea específica que puede repetirse cada vez que se quiera; por ello también recibe 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 tan sofisticado como se quiera (análisis factorial, discriminante, etc.).

El spss cuenta con rutinas para crear archivos de datos (término que indica al conjunto de casos de una investigación); puede producir con estos, otros datos; tiene procedimientos estadísticos para llevar a cabo cálculos e imprime los resultados de manera uniforme e igual cada vez que se requiera.

Pero, ¿qué es un programa? y ¿cuál es su necesidad?

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

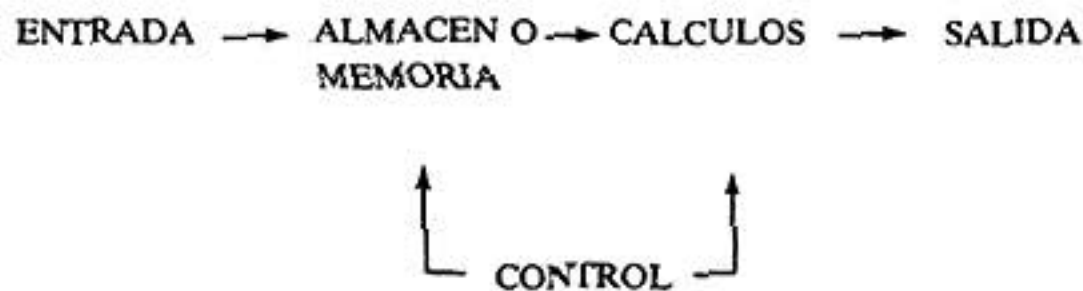
2. Norman H.N.E. et al. *SPSS. Statistical Package for the Social Sciences*, second edition, USA, Mc. Graw Hill, 1975.

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

Para comprender mejor estas ideas veamos, aunque 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 y preprogramada, apretando simplemente una tecla. Una computadora, además, de contar con una área de este tipo para efectuar operaciones (unidad aritmética), cuenta con otra que por medio de instrucciones (de transferencia, de control, de almacen, de especificación, etc.) puede controlar el orden en que la sucesión de operaciones debe realizarse (unidad lógica).

Haciendo un poco de historia podría considerarse como el diseño de la primera computadora electrónica, la propuesta por Charles Babbage (1791-1871),³ y que de hecho es la base o modelo de desarrollo de las computadoras modernas.



Así, una computadora está formada por una Unidad de Entrada (teclado, discos, modems, etc.), que recibe tanto la información a procesar como las instrucciones para su procesamiento, almacenamiento e impresión (programa). La Unidad de Memoria en donde se almacena la información (datos y programas, que se requieren al momento de una corrida). 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 (impresora, disco,

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

modems, etc.), que nos da el reporte de la información procesada, de acuerdo también, a las instrucciones recibidas.

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

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.*

Continuando con la explicación de la definición del spss, nos faltaría discutir la segunda parte de la definición que dice "... diseñado para el análisis de las ciencias sociales". Como el presente trabajo va dirigido a los estudiosos de ese campo, podría pensarse que no es necesario. Sin embargo consideramos pertinente hacer las siguientes reflexiones.

Primero. Podría pensarse que el spss es una propuesta "nueva", distinta a las ya existentes, pero no es así. Vayamos por partes. Claramente las siglas del paquete nos hablan de un Paquete Estadístico para las Ciencias Sociales, lo cual concuerda con lo de "... análisis de datos", entendiéndose por esto el campo de acción de las técnicas estadísticas, es decir describir los fenómenos y poner a prueba la representatividad estadística de las hipótesis de investigación. Pero además, el paquete cuenta con procedimientos que permiten el procesamiento de datos, lo cual implica la transformación y el manejo de datos en términos de clasificar, agregar y/o construir otros datos, como es el caso de los índices, o definiciones complejas: costo de la vida, nivel socioeconómico, nacionalismo, participación política, etc.

Segundo. Al hablar de análisis de datos para las ciencias sociales, puede afirmarse que es para cualquier tipo de perspectiva teórica, ya que los datos son registros de hechos observados de la realidad. Y a pesar que a ésta la interpretamos desde una perspectiva teórica determinada, es necesario recordar que las técnicas que permiten el análisis estadístico fueron desarrolladas para las Ciencias Nomotéticas, es decir, aquéllas cuyo objetivo primordial es organizar el mate-

rial utilizando reglas cuyo fundamento es lógico y matemático. Dispuesto así el material, se pueden detectar leyes y luego explicarlas; se retoma entonces la perspectiva teórica de la ciencia que se trate. Por tanto puede decirse que el SPSS, no sólo puede usarse en las ciencias sociales, sino en todas las áreas del saber.

En resumen, el SPSS es un conjunto de programas, de rutinas y procedimientos estadísticos para la construcción y manejo de archivos, la manipulación —transformación y creación de nuevas variables— y para efectuar el análisis estadístico de los datos recabados en las investigaciones del tipo del *Survey* (encuestas), *Panels*, etc. Estos programas fueron diseñados para realizar con rapidez y gran eficacia el procesamiento de los datos, que de otra manera, a mano, no sólo llevaría muchísimo tiempo, sino que la misma programación sería muy difícil resolver.

En suma, la definición del SPSS no es muy justa; a decir verdad, el paquete es muchísimo más de lo que sobre él se ha dicho, pero no es, por otro lado, la panacea para resolver todos los problemas planteados por las ciencias sociales.

BREVE HISTORIA DEL PAQUETE ESTADISTICO

El SPSS se llevó cerca de una década para el diseño del sistema, en su programación y su documentación. 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.

A decir de los autores del paquete, el SPSS surgió como necesidad de responder a una generalizada frustración de investigadores en Ciencia Política de la propia Universidad de Stanford. Sucedió lo mismo en nuestra Universidad, en centros de investigación social, cuando la computación fue incorporada.

La incorporación paulatina de la computación a las universidades mexicanas, aparece como la solución mágica de todos los problemas en el manejo y análisis estadístico de grandes volúmenes de información. En efecto la computación abrió enormes posibilidades a las ciencias en general, 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 adquisición de las computadoras se olvidó el problema de que eran máquinas y no los cerebros que se decía. El problema de su instalación y mantenimiento fue mayor en México, pues nosotros no desarrollamos dicha tecnología. Además estas máquinas requieren para su manejo dos condiciones fundamentales, que pareciera fueron soslayadas: saberlas operar y programar. Este problema consideramos que fue más significativo en las ciencias sociales, debido a la formación básica de sus profesionales, que no son instruidos, ni preparados en dicho campo. 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 al avance tecnológico. Los resultados fueron efectivamente los señalados por los autores del SPSS: frustración, cuando el procesamiento de las investigaciones, como encuestas de opinión, consumieron demasiado tiempo, en ocasiones sin los resultados esperados y, en otras, abandonándose incluso los proyectos.

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

Con el tiempo se prepararon y formaron verdaderos profesionales de la computación. Sin embargo, se les siguió dando todo el peso y responsabilidad que los investigadores les habían otorgado desde el inicio de la incorporación de la computación a las ciencias sociales.

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

El equipo inicial de la Universidad de Stanford publicó en 1970 la primera edición del paquete. Esta fue mejorada poco a poco por el mismo equipo de trabajo, con las observaciones de sus primeros usuarios; dándose como resultado lo que se llamó "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 cuarta versión del SPSS. El paquete fue adaptado a la Burroughs B-5500, en 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; es en 1975 cuando se publica la segunda edición que contiene la sexta versión. Esta publicación ya es hecha en la Universidad de Chicago y la considero la más conocida y utilizada de sus versiones. Cabe aclarar que este ensayo se basó en esa versión.

Posteriormente en el año de 1979 se publica el SPSS-UPDATE⁴ con las versiones siete y ocho, que son las que se manejan actualmente en la UNAM. Las mejoras más importantes de estas versiones, además de contener las de la sexta, son:

- La inclusión de catorce procedimientos estadísticos no-paramétricos, entre los más conocidos: la prueba de Rung, la de Mcnemar, la del Signo, la de Kolmogorof-Smirnov y la de Kruskal-Wall en el Análisis de Varianza Simple.
- El manejo estadístico para el tratamiento de la respuesta múltiple.
- La construcción de Tablas de Supervivencia o de esperanza de vida 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 el uso y programación de las computadoras modernas.

Por último, en la actualidad no sólo contamos con el SPSS diseñado para las grandes computadoras, sino que hay versiones para Minis y Micros, como las denominadas PC (Personal Computer); que en esencia hacen lo mismo.

4. Hull C., Hadlai y Nie H. Norman. *SPSS-UPDATE, New Procedures and Facilities, for Release 7 and 8.*, USA, Mc. Graw-Hill Book Company, 1979.

Queremos aclarar que además de éste, existen una infinidad 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 concebidos en forma paralela, como el SAS (Statistical Analysis System)⁵ y el BMD *series-p* (Biomedical Computer Programs).⁶

INSTRUCCIONES GENERALES PARA INICIAR UNA CORRIDA CON EL PAQUETE ESTADISTICO

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

En general se podría decir que existe un procedimiento distinto para trabajar en cada sistema, pero en esencia es el mismo y consiste en lo que llaman instrucciones de control de acceso a la computadora, también conocidas como tarjetas de control. La tarjeta fue el primer medio para introducir datos y/o instrucciones a una computadora, ahora es casi desconocida, pero los nuevos medios —terminal, cinta magnética, disco y diskettes— la utilizan como modelo o imagen en el registro de datos.

Las tarjetas de control o instrucciones nos permiten la entrada al sistema y contienen en primer instancia una identificación de la tarea que deseamos realice la computadora. Así, por ejemplo, JOB XXX da a la computadora la identificación que permitirá diferenciarla de otras tareas, que en ese momento esté realizando paralelamente el sistema.

En segundo lugar viene la identificación del usuario (USER), pues por lo general se requiere de autorización para la utilización de una computadora y a través de ella se lleva la contabilidad en su uso. Esta identificación está formada por dos partes: una clave que generalmente se compone de las iniciales del usuario y un número que es determinado siguiendo ciertos criterios, como ser investigador, o profesor, o estudiante. Así por ejemplo VG89, AC20, son claves de

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

6. Dixon, W.J. *Biomedical Computer Programs*, Berkeley, The University of California Press, 1977.

investigador. La otra parte corresponde al *Password*, que es una "llave" para el acceso a la computadora y por lo general se determina de acuerdo con el usuario, pues ésta debe ser secreta, para que nadie la use, excepto él. Esta llave puede estar formada por letras o números o combinaciones de ellas, v.g.: DF, LOLA,6336, etc.

En tercer lugar viene la prioridad del trabajo (*CLASS=N*), pues el costo y la mejor utilización de la computadora requiere de su optimización. Así dependiendo del tipo de trabajo que el usuario realice, se le da o asigna su prioridad, *N=50* por ejemplo. Esta se otorga de acuerdo a criterios establecidos por las autoridades que dirigen los centros de cálculo.

En cuarto lugar es necesario indicarle a la máquina, dónde comienzan las instrucciones (*BEGIN*) y dónde terminan (*FINISH*). En otros casos estas instrucciones son respectivamente *JOB* y *EOJ*, abreviaturas de fin de trabajo (*End of Job*). Cabe aclarar que las computadoras utilizan una infinidad de instrucciones para iniciar una corrida, tales como: *RUN* (correr), *EXECUTE* (ejecutar), *DO* (hacer).

En quinto lugar, comprendidos entre el *Begin* y el *Finish* vienen las instrucciones de *RUN SERVICIO/SPSS*, la cual instruye a la computadora que vamos a trabajar con dicho paquete. Existen otras instrucciones para indicar a la computadora la forma en que va a recibir la información, como el *DATAIOCR* que indica a la computadora que los datos de entrada (*Input*) y/o salida (*Output*) están en tarjetas (*Card Register*). Estas tarjetas de control se obtienen, por lo general, en la recepción de trabajos en los centros de cómputo y varían de sistema a sistema, pero en esencia son las mismas ya que se han originado bajo ideas similares, de control y optimización.

Los siguientes ejemplos son para procesos en lotes o grupos (*batch*); estos no son transparentes para el usuario, el cual sólo ve los resultados hasta el final de su corrida.

Debido a que las computadoras, macros y minis, pueden realizar al mismo tiempo varios procesos o corridas, lo que es conocido como multiproceso, estas computadoras leen todas las instrucciones desde un *JOB* o *RUN* hasta el *FINISH* o *EOJ*, de una corrida tras otras, así que hasta no tener el listado con el reporte de nuestra corrida, no podemos saber qué pasó. A esto le llamamos trabajos en lotes y por ello se dice que no son transparentes para el usuario. En cambio una PC, como su nombre lo indica, Computadora Personal, sólo atiende

un usuario a la vez, esto permite interactuar a la computadora con nosotros, es decir, aceptando o rechazando las instrucciones que le demos.

Así por ejemplo, para la Burroughs B-7800⁷ de la UNAM las instrucciones o tarjetas de control son:

```
? JOB XXX USER XXX/YY; CLASS = N : BEGIN
RUN SERVICIO/SPSS:
DATAIOCR
- PROGRAMA EN SPSS
  READ INPUT DATA
- DATOS
  FINISH
```

Mientras que para una Minicomputadora como la de la UAM Ixtapalapa, HP-3000⁸ el usuario requiere de las tarjetas de control siguientes:

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

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 (el paquete requiere de un espacio de 3 millones de caracteres para almacenarse). En las micros el SPSS trabaja en modo interactivo (platicando con el usuario), su acceso supone teclear las siguientes instrucciones:

- SPSSPC (en la pantalla aparecerá el logotipo SPSSPC)
- El programa
- Los datos
- Finish

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

8. Castillo M., Alberto y Alejandro Serrano Muñoz. *Guía para el principiante. SPSS, México, 1982, 70 p.*

Cabe advertir que por lo general se inicia una tarea prendiendo los dispositivos de entrada y salida, por ejemplo: SET *Disk On* o SET PRINT ON para enviar el programa y los resultados a disco o a la impresora, según necesidades.

Se recomienda que los datos y el programa sean cargados utilizando un procesador de textos, como el *Word Star* o *Word Perfect*.

La liga o la relación entre programas y datos se establece en la instrucción de DATA LIST. Hacemos notar que las versiones para las PC cuentan con un programa tutor que enseña a usar al paquete SPSS, de una forma amena y sencilla.

UTILIZACION DEL PAQUETE ESTADISTICO EN EL PROCESAMIENTO DE DATOS

CONSTRUCCION DE ARCHIVOS Y SUBARCHIVOS

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

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

Cada caso estará compuesto por todos los valores de las mediciones que se hayan hecho a las unidades de registro y corresponden a las variables que serán sujetas a procesamiento.

Cabe aclarar que (no sólo en el SPSS) el orden de las variables para distintos individuos, 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; son alimentados a través de distintos 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 las variables que son números: enteros o con decimales. Los segundos son combinaciones de números y letras del alfabeto e incluso caracteres (*#\$\$%&, etc.).

La distinción de procesamiento es muy importante. Los datos alfanuméricos consumen un mayor tiempo, y no pueden 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.

Si bien una etiqueta o nombre 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 manejo. Sin embargo nada existe que nos limite a efectuar operaciones aritméticas con variables alfanuméricas, de hecho, si le es solicitada, la computadora busca un resultado y lo da como bueno, es decir no puede distinguir; ni la computadora, ni el paquete tiene forma de limitar el procesamiento si éste es o no pertinente, y queda a juicio del usuario el buen o mal uso de un determinado procedimiento.

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

En la creación de archivos, dos instrucciones son de gran importancia; éstas son de definición de datos:

La primera, VARIABLE LIST, indica las variables a ser manejadas; en lo que se refiere a: número, orden y sintaxis, es decir, instruye a la computadora para leer los datos e identifica por medio de la segunda, INPUT FORMAT, en donde se especifica el formato para cada una de las variables, y que no es otra cosa que el conjunto de columnas de una tarjeta que se destinan para el registro de un dato. Desde luego que ahora ya no se usan las tarjetas, pero con todo, cualquier medio que se utilice la simula.

Dependiendo del número de variables que se maneje, puede 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 a

leer tantas variables como las que se mencionan en la VARIABLE LIST y de acuerdo al INPUT FORMAT específico. El paquete lleva un conteo del número de variables; por esto reconoce cuándo se completa un caso e infiere que los siguientes datos que se le están proporcionando, corresponden al siguiente.

Existen dos tipos de formatos que son utilizados por SPSS, el fijo (FIXED) o el variable (FREEFIELD). Ambos se originaron con base a las limitaciones de una tarjeta. El formato fijo implica que los datos se registren siempre en las mismas columnas. Así por ejemplo:

<i>variables</i>	<i>código</i>	<i>columnas</i>
identificación	RFC o N ^o . 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 en el registro de datos, es la facilidad para checarlos por columnas, pues de otra forma se traslapan los campos de las variables, dificultando la tarea de su verificación visual.

El único problema que podría ocurrir es que, si algún dato o más de un caso particular no se registra, la computadora toma el siguiente que encuentre, sin importar a qué caso corresponde, ocasionando un corrimiento de datos. Esto ocasiona que se mezclen los casos y que el reporte de resultados sea ininteligible o extraño, lo cual evidencia errores. En otras ocasiones puede darnos resultados que aparentemente sean correctos, pero que conllevarán errores tanto de análisis como de interpretación.

Los datos a procesar son organizados en unidades llamadas archivos (FILES); un archivo se compone de los datos de la investigación y se complementa con información asociada a ellos, que nos permite describir y definir cada una de las variables.

A continuación ejemplificamos un programa con el que se construye un archivo.

ESTRUCTURA DE UN PROGRAMA PARA CREAR UN ARCHIVO SPSS

(Definición de datos)

<i>1.</i>		<i>80.</i>
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 etiquetas de los valores de las variables.	
MISSING VALUES	Valores de las variables que no se van a usar o que no se reportaron.	
PRINT FORMAT	Decimales de las variables a considerar en la impresión.	
READ INPUT DATA	Instrucción para que SPSS comience a leer datos.	
D A T O S		
XXXXXX	XXXXX	XXXXXXXX XXXXXX XXXXXX XXXX
XXXXXX	XXXXX	XXXXXXXX XXXXXX XXXXXX XXXX
.....		
XXXXXX	XXXXX	XXXXXXXX XXXXXX XXXXXX.....
SAVE FILE	Condición para hacer permanente un archivo o quedar en disco.	
FINISH	Indicación para terminar una corrida.	

Como podemos observar un programa SPSS está estructurado por diferentes tipos de instrucciones, referidas en el manual como tarjetas (*Cards*). Así pues las instrucciones son:

1. *De control.* RUN NAME, READ INPUT DATA, SAVE FILE y 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 o a una práctica. Esta instruc-

ción es opcional, es decir, puede o no incorporarse al programa. 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.

READ INPUT DATA. Instruye al SPSS a iniciar la lectura de los datos, uno por uno y al mismo tiempo procesarlos de acuerdo al procedimiento estadístico requerido en el programa.

SAVE FILE. Ordena al paquete conservar el archivo en forma permanente, pero no inmutable, es decir, que se pueden hacer correcciones y/o modificaciones posteriores, para así poderlo usar tantas veces como sea necesario. En caso de que esta instrucción no se ejecute, el archivo construido desaparecerá tan pronto como termine el último procedimiento que se haya ordenado.

FINISH. Indica al paquete que se ha terminado una corrida y debe transferir los controles al sistema operativo.

2. *Definición de datos*, por medio de los cuales se construye el archivo: **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.* Se refieren a los procedimientos estadísticos que ordene el usuario y que se encuentran disponibles en el paquete, como: distribuciones de frecuencias, tablas de contingencia, correlaciones, 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

* Para las definiciones de **VARIABLE LIST** e **INPUT FORMAT** ver *supra*.

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 pueden solicitar la impresión de algunas estadísticas complementarias al procedimiento estadístico que se esté procesando o, incluso, algunos resultados intermedios que el mismo paquete esté realizando.

Cabe aclarar dos cuestiones:

Primero: Que una vez construido un archivo, éste pueda ser usado tantas veces como sea necesario, incluso permanecer casi "indefinidamente", ya que todos los medios de registro tienen un tiempo de vida o duración. Existe además en los centros de cálculo, una política de actualización y depuración, con el objeto de no disminuir la capacidad de memoria de la computadora (o los discos), por lo que periódicamente se borran los archivos viejos.

Segundo: Que son iguales en género y número las instrucciones requeridas para crear y procesar los archivos de datos o efectuar algún procesamiento estadístico, a partir de tarjetas, cinta magnética o disco.

Con frecuencia, en la investigación social, se recaban dos o más muestras con el objeto de ser analizadas en forma independiente y posteriormente ser comparados sus resultados y comprobar si hay o no diferencias estadísticamente. Estos grupos así segmentados se pueden almacenar con una estructura de subarchivos (SUBFILES). Por ejemplo, una encuesta de participación política puede ser aplicada en distintas ciudades de la República o a individuos con distintas condiciones: económicas, políticas y 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; a su vez cada estado está compuesto por municipios y ellos por localidades que pueden conformar también una estructura en subarchivos.

El SPSS cuenta con la posibilidad de estructurar un archivo segmentado en subarchivos y estos a su vez ser operados, procesados y/o manipulados de diferentes maneras, permitiendo cualquier análisis estadístico de contraste.

Así un archivo estructurado en subarchivos permite:

- 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 instrucción o comando SUBFILES, como una variable de control;
- el investigador puede estar interesado en hacer comparaciones de la relación entre variables en los diferentes subarchivos. O sea, esta estructura permite la realización independiente de cualquier proceso para cada uno de los subarchivos, o cualquier agrupación de estos;
- seleccionar algunos subarchivos para ser procesados en forma individual, en combinaciones, o incluso en forma conjunta ignorando la estructura de subarchivos; y
- al término de cualquier corrida, los casos que componen un archivo pueden ser reordenados, y una nueva estructura de subarchivos puede surgir, de modo que concuerde con nuevos requerimientos de la investigación.

Finalmente, las restricciones que los archivos y subarchivos pueden tener son:

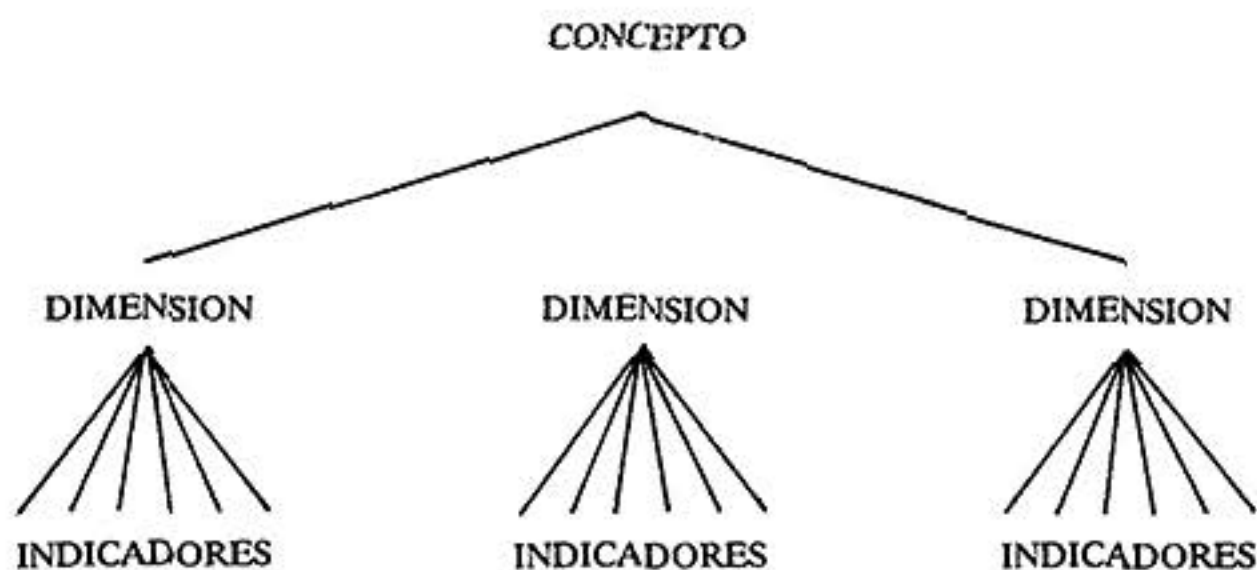
- Los subarchivos deberán tener el mismo formato y las mismas variables.
- No podrán ser definidos más de cien subarchivos.
- No se podrán definir más de quinientas variables en un archivo y consecuentemente en la totalidad de subarchivos. Sin embargo, el número de casos de los subarchivos pueden ser diferentes para cada uno de ellos.

Un archivo creado, independientemente de su estructura, podrá ser permanente o no según intereses y requerimientos del usuario. Pero más allá de su permanencia, puede modificársele, es decir, puede ser alterado agregándose más casos o más variables, o incluso borrando datos primarios que ya no serán utilizados posteriormente, para así dejar libre la capacidad de memoria de la computadora o del disco que estemos utilizando; esta capacidad tiene límites y hay que procurar no agotarla para poder trabajar.

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 y poner a prueba las hipótesis de investigación? Algunos metodólogos como Paul Lazarsfeld, han propuesto mecanismos generales, cuyo esquema es ampliamente conocido.⁹



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, o simplemente el guión para una entrevista. El proceso de medición requiere además, de otro, que es inverso al primero. Es decir, de "compactación" o agregación de los indicadores; este procedimiento es llevado a cabo mediante la construcción de un índice, el cual posteriormente recibirá un tratamiento de variable. Cabe aclarar que el mismo indicador puede ser tratado también como variable.

9. Citado por Padua, Jorge. *Técnicas de investigación aplicada*, México, FCE, 1965.

El paso de la teoría a su contrastación estadística ha requerido de la generación de información empírica 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, agrupadas en lo que conocemos como Estadística.

El procesamiento de datos tiene como función manejar las observaciones 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 utilizarlos lógicamente y ordenadamente. El cálculo se refiere al proceso que se debe realizar para resumirlos o hacer inferencias con ellos.

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 información empírica; son la materia prima de la cual se derivará la contrastación 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 el manejo de datos empíricos. Para ello realiza distintas tareas, que se ordenan a partir de instrucciones, como las ya mencionadas de definición de datos, y que se pueden utilizar para la modificación de datos, la selección de datos, la modificación o actualización de archivos, y el procesamiento estadístico de los datos.

MODIFICACION DE DATOS

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

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

RECODIFICACION DE DATOS

RECODE. Es utilizado para recodificar las variables permitiendo reagrupar los valores de una variable en particular. 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, se reagrupan entonces, sumando categorías, bajo criterios conceptuales. 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 en 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.

1.	16.	80.
RECODE	INGRESO (0 THRU 5000 = 1), (5001 THRU 10 000 = 2) (10 001 THRU HIGHEST = 3)	

LA TRANSFORMACION DE DATOS

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

COMPUTE. Literalmente calcula; ordena la realización de operaciones aritméticas con variables, haciendo uso de funciones matemáticas expresadas en lenguaje de programación FORTRAN, como podría ser un índice sumatorio.

Como ejemplo, podríamos suponer que el costo de vida (COSTVID) está conformado por los distintos costos: de alimentación (COSALI), de vivienda (COSVIV), de vestido (COSVES) y otros (COSOTROS). Considerados en una encuesta y grabados en un archivo SPSS, la generación de la nueva variable costo de vida (COSTVID) que los incluye sería:

 COMPUTE

 COSTVID = COSALI + COSVIV + COSVES + COSOTROS

IF. Es un SI condicional o restricción, que se utiliza cuando se quiere que se cumplan ciertos requisitos predeterminados. Es decir, condición de valores o categorías de variables expresadas en forma lógica, en la que se relacionan dos cantidades algebraicamente a través de operadores relacionales como: mayor que (GT), igual a (EQ), o menor que (LE). A su vez estas expresiones pueden estar unidas (dos o más), por medio de operadores lógicos: AND, OR, NOT, con los que unimos, separamos o negamos las relaciones respectivamente.

Supóngase que deseamos construir una variable llamada capacidad de ahorro (CAPAHO). Como punto inicial debemos definir, lo que para nosotros es la "Capacidad de Ahorro". Por ejemplo, el remanente o sobrante que nos queda del ingreso, después de deducir el gasto familiar o el costo de vida, para la supervivencia de una familia (TAMAÑO FAMILIA). Una vez logrado esto, habría que operacionalizar dicha definición, se supone que hemos determinado de antemano de qué tipo de ingreso hablamos y que conocemos el costo de vida *per cápita*, como el que se calculó en el ejemplo anterior (COSTVIV). Una posible definición operacional podría ser: capacidad de ahorro es la diferencia entre el ingreso del jefe de familia y el costo de vida que se requiere para la manutención de su familia, es decir, el ingreso menos el costo de vida *per cápita* multiplicado por el número de miembros de la familia en cuestión y que en este ejemplo denominamos, TAMFAM. Así la primera expresión lógica matemática se iniciará con el IF condicional, expresión que le indicará al sistema que si el ingreso de una familia x, menos el producto del tamaño de familia (3, 4, 5, miembros) por el costo de vida *per cápita* (y que en este caso hemos supuesto cien mil pesos), es menor que (LE) cien mil pesos, éste se diferenciará de otros casos por un código 1 para la variable CAPAHO que estamos construyendo. Por supuesto éste es un ejemplo hipotético y sólo nos sirve para mostrar la mecánica de funcionamiento. Finalmente la variable CAPAHO puede tener los siguientes valores: Bajo (1), Medio (2) y Alto (3), como se muestra a continuación.

IF (INGRESO - (TAMFAM x 100 000) LE 100 000) CAPAHO = 1
(INGRESO - (TAMFAM x 100 000) GT 100 000) AND
(INGRESO - (TAMFAM x 100 000) LE 500 000) CAPAHO = 2
(INGRESO - (TAMFAM x 100 000) GT 500 000) CAPAHO = 3

COUNT. Literalmente contar; nos permite la construcción de índices sumatorios, pues con esta instrucción podemos contar el número de veces en que un determinado valor de una variable aparece para cada caso, el cual puede ser previamente definido. Esta expresión puede ser utilizada para sumar las respuestas a indicadores que conforman una dimensión. Supóngase, a manera de ejemplo, que contamos con la información acerca de los periódicos que consultan los entrevistados en una encuesta, y que nos interesa construir un índice que nos informe acerca del número de fuentes de información periodística consultadas. Supongamos que se han hecho las preguntas respecto a si se consultan o no ciertos periódicos y que esta variable se ha codificado como una variable dicotómica, por ejemplo Si=1 y No=2. Este índice se podría llamar MAXINF.

COUNT MAXINF = JORNADA + UMASU + + EXCELSIOR

SELECCION DE DATOS

Las instrucciones de selección de datos nos permiten realizar segregaciones de un archivo SPSS, a través de la elección de una muestra aleatoria de casos o la generación de una muestra intencional; se forma con aquellos casos que cumplan con una condición específica de interés analítico para la investigación. Cuenta también 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 las inferencias sobre la base de unos casos para toda la población.

SAMPLE. Literalmente muestra; nos provee con una muestra aleatoria generada de un archivo SPSS. Su tamaño se determina en función del número de casos que se requieran 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 diez, quince o veinte por ciento por ejemplo, de acuerdo a nuestras necesidades, que pueden estar determinadas por el tipo de análisis estadístico que se desea realizar. Algunos procedimientos estadísticos como las tablas de contingencia y el análisis de varianza, requieren de un número de casos mínimos por celda, como restricción para su buen funcionamiento.

<i>1</i>	<i>16.</i>	<i>80.</i>
SAMPLE	0.10	

SELECT IF. Seleccionar bajo cierta condición, nos provee con una muestra intencional, en función de un criterio determinado, que se fija respecto a un valor de una variable o grupos de variables, a través de expresiones del siguiente tipo:

<i>1</i>	<i>16.</i>	<i>80.</i>
SELECT IF	(OCUP EQ 2 OR OCUP EQ 5)	

Esta selección generaría un subarchivo comprendido con todos aquellos casos que cuenten con la condición de tener la ocupación 2 o 5, según códigos correspondientes.

WEIGHT. Es utilizado para ponderar o dar a cada caso su correspondiente peso relativo, que nos permite hacer inferencias sobre las frecuencias en 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. 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.

<i>1</i>	<i>16</i>	<i>80</i>
IF	(SECTOR EQ1) WTFACTOR = 1.25	
IF	(SECTOR EQ2) AND (SECTOR LE 6) WTFACTOR = 2.75	
IF	(SECTOR GT 5) WTFACTOR = 3.0	
WEIGHT	WTFACTOR	

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

LAS INSTRUCCIONES DE MODIFICACION DE ARCHIVOS

Las instrucciones de modificación de archivos sirven para actualizarlos, agregando nuevos casos o datos, o borrando los que no vamos a utilizar más, o incluso reordenándolos de acuerdo a nuevas necesidades de investigación. Estas son las siguientes:

LIST FILEINFO. Es la expresión mnemónica que hace referencia a listar la información comprendida en el archivo SPSS. En otras palabras, esta tarjeta nos provee con un reporte completo de todas o de algunas definiciones de datos, utilizadas en la construcción de los archivos SPSS, según sean nuestras necesidades. Y puede estar orientada a conocer qué variables contiene el archivo, qué etiquetas están asociadas a ellas, su orden, etc.

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

Por último **WRITE CASES** y **WRITE FILEINFO** tienen funciones similares a estas instrucciones, pero su objetivo es proporcionar información del archivo a través de otros medios, como tarjetas o cinta magnética, por ejemplo, para que 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 en la corrida no se invoca un procedimiento estadístico cualquiera que éste sea, como FREQUENCIES, CROSSTABS, etc.; pues estos son los que permiten el acceso a los datos de un archivo SPSS.

EL CALCULO ESTADISTICO

Quizá la parte más conocida del SPSS, es el tratamiento estadístico con el que se manejan los resultados de una investigación empírica. Y es así porque indudablemente la Estadística es una herramienta indispensable en el manejo de datos de casi todas las áreas del saber, como lo confirma que forme actualmente parte de casi todos los programas de posgrado y de áreas tan diferentes como Medicina y Sociología. Sin embargo, hasta ahora no se ha cuestionado el por qué de esta importancia. No queremos desviarnos de nuestro tema principal; sólo mencionaremos dos de los elementos sobre los que descansa esta importancia: el primero, la Estadística forma parte del proceso de investigación, pues es con ella que se lleva a cabo el análisis de los datos. Y segundo, que la Estadística, por ser un conjunto de técnicas y métodos matemáticos, es de aplicación general.

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.

TAREAS QUE EL SPSS PUEDE REALIZAR

El SPSS en su versión más conocida (sexta) cuenta con dieciséis procedimientos estadísticos también llamados 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 (OPTION) y estadísticas (STATISTICS), que de hecho permiten cambiar formas de impresión, utilización de diferentes modelos o métodos, y proporcionar 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 lo es la obra *Estadística social*, de Blalock.¹⁰ Pero incluso cuenta con procedimientos como la construcción 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, por tanto, requieren la mayoría de las veces un tratamiento particular, como podría ser el análisis factorial o el discriminante.

En su presentación, el SPSS sigue lo que podría calificarse como el análisis empírico clásico: primero, la descripción de los datos y luego los procedimientos multivariados y los de estadística inferencial. Segundo, la lógica práctica del análisis de los datos se inicia con el análisis univariado (de variable por variable) y, posteriormente, —después de las transformaciones y recodificaciones de las variables—, el análisis de relaciones entre variables que puede ser bivariadas o multivariadas. Así la primera tarea que podría considerarse obligatoria es la construcción de distribución de frecuencias simples, relativas y acumuladas, las cuales tienen múltiples usos: checar códigos y así poder corregir posibles errores de captura; observar el comportamiento de cada una de las variables para su reconstrucción, eliminación o separación para estudios de pauta.

Para la realización de esta primera tarea estadística, el SPSS cuenta con dos subprogramas: FREQUENCIES Y CONDESCRIPTIVE.

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

La instrucción anterior instruye al sistema para generar las distribuciones de frecuencias simples y acumuladas para cada una de las variables mencionadas o de todas, haciendo uso de la expresión ALL. La instrucción OPTIONS, correspondiente al programa invocado, contiene nueve distintas opciones de salida. Permitiendo diferentes tipos de impresión de resultados, eliminación de etiquetas, construcción

10. Blalock, Humbert Jr. *Estadística Social*, México, FCE, 1975.

de histogramas, o 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 no se contrapongan. Así por ejemplo, la opción número tres instruye al programa para la impresión, dejando la parte derecha para hacer anotaciones pertinentes a los resultados; mientras que la número siete instruye al paquete para imprimir sólo las estadísticas de resumen (media, mediana, varianza, etc.) y la número cinco proporciona un reporte en forma condensada, en el que además no aparecerán las etiquetas de las variables.

Una instrucción *STATISTICS* produce distintas estadísticas de resumen: media, error estándar, mediana, moda, desviación estándar, varianza, kurtosis, asimetría, rango, valor mínimo y máximo observado; pueden ser solicitadas algunas o 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 subrayar 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; ésta es responsabilidad absoluta del usuario.

Otro subprograma que permite un análisis descriptivo más fino es el *BREAKDOWN*, literalmente romper hacia abajo. Produce algunas de las estadísticas de resumen señaladas anteriormente para variables continuas (variables criterio) y son calculadas para cada valor de otras variables (variables control), que incluso tengan el nivel nominal u ordinal. Así por ejemplo un *BREAKDOWN* podría calcular el ingreso promedio (criterio) de individuos o el ingreso mediano, tomando en consideración las combinaciones producidas entre otras variables como podrían ser: población, ocupación y sexo (control), bajo el supuesto "en este ejemplo", de trato salarial diferencial por población, ocupación y sexo. Este subprograma puede ser utilizado en combinación del *T-TEST*, con el cual se puede poner a prueba la significación estadística de las diferencias encontradas en el *BREAKDOWN*.

Los siguientes procedimientos 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, en 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 dos procedimientos no paramétricos: la Correlación de Spearman o el Coeficiente de Kendall, contenidos en el subprograma NOMPAR CORR. 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), con subrutinas especiales que nos permiten analizar relaciones no lineales, la utilización de variables mudas (DUMMY) y el análisis, o de senderos (PATH), para la interpretación causal de los fenómenos.

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

Cabe hacer una última aclaración: todos estos procedimientos fueron diseñados para el manejo de quinientas variables como máximo. Podrían antojarse pocas a primera vista, pero en la práctica no siempre manejamos tantas. En última instancia siempre es posible resolver cualquier problema de esta índole a través de muchas maneras, como podría ser eliminar variables, o reducirlas formando índices, etc. Como es posible contar con archivos mayores, el SPSS cuenta con las instrucciones para el manejo de archivos mayores.

CONSIDERACIONES FINALES

El lector interesado en estos temas deberá partir del principio de la existencia de un desarrollo vertiginoso e impresionante de la tecnología computacional. En nuestros días, se convierte rápidamente en "obsoleta" la versión nueva. En otras palabras, apenas comenzamos

a comprender o adaptarnos a las modificaciones de una versión, aparece en el mercado la nueva, "la actual", "la mejor":

Alertamos al lector en la posibilidad de enfrentarse no sólo a nuevas versiones del SPSS, sino a propuestas que van más allá de la mera inclusión de nuevos procedimientos. Sería el caso del SPSS^x, que es una versión especial para los sistemas IBM, cuyo componente principal no sólo es el SPSS original, sino su fusión con otro paquete estadístico, el SAS (STATISTICAL ANALYSIS), y con el enriquecimiento de su sintaxis con un lenguaje de programación, como el Fortran.

En general, las versiones derivadas del SPSS original o cualquier otra como el SPSS^x son iguales en esencia, sólo que cada vez más enriquecidas y versátiles.

Por todo lo anterior, recomendamos a los usuarios potenciales del paquete SPSS consultar y estudiar el manual correspondiente a la versión que vayan a manejar; fundamentalmente la sintaxis o lenguaje de programación y el o los procedimientos estadísticos conforme se vayan a utilizar.