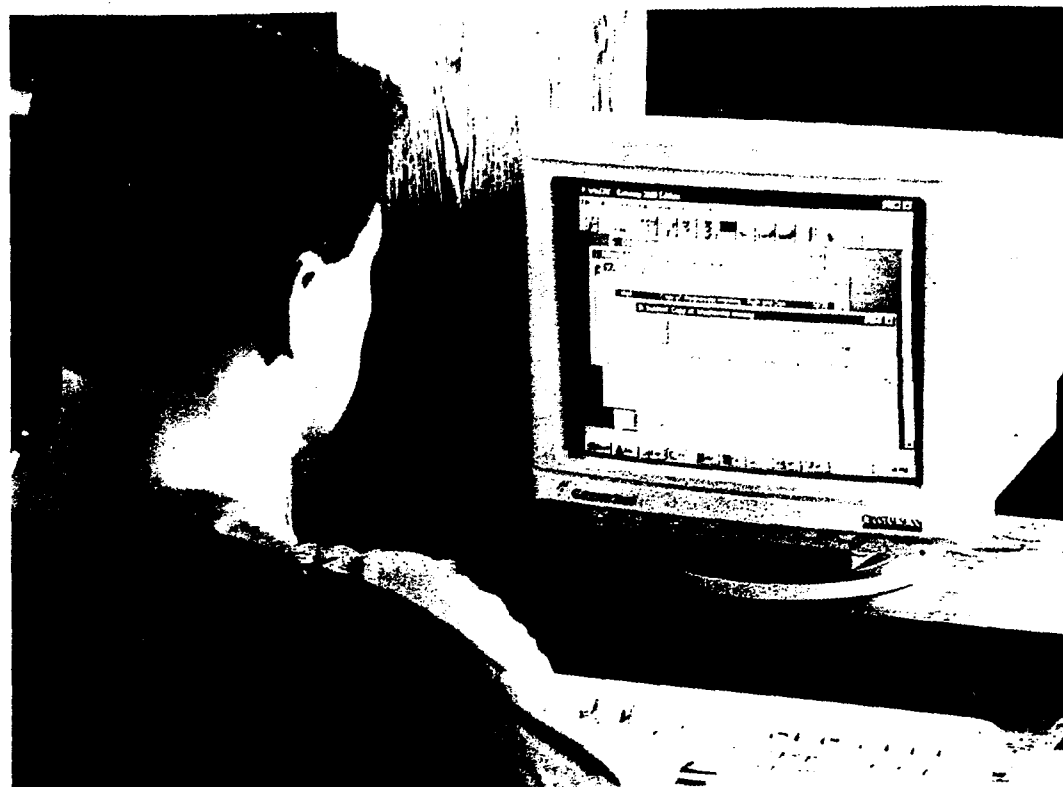




Cuantificación de los datos

Lo que aprenderá en este capítulo

Después de recopilar un conjunto de observaciones, es útil convertirlas en una forma que se preste al análisis por computadora. Aquí veremos la respuesta del científico social al microscopio, el láser y el ciclotrón.

En este capítulo...

Introducción

Las computadoras en la investigación social

Introducción a las computadoras
Microcomputadoras

Codificación

Elaboración de categorías de codificación

Construcción del libro de codificación

Opciones de codificación y captura de datos

Hojas de transferencia
Codificación al margen
Captura directa de datos

Datos capturados por los entrevistadores
Codificación para las hojas de lectura óptica
Uso directo de hojas de lectura óptica
Conexión con un programa de análisis de datos

Depuración de datos

Depuración de códigos posibles
Depuración de incongruencias

Puntos principales

Preguntas y ejercicios de repaso

Proyecto de continuidad

Lecturas adicionales

Introducción

Lo que el microscopio ha sido para la biología, lo que el telescopio para la astronomía, lo ha sido la computadora para la investigación social cuantitativa. Más aún, usted sigue este curso en una época en que la aportación de la computadora a las ciencias sociales está por descubrirse. En este capítulo nos introduciremos en esta aportación.

El propósito de este capítulo es describir los métodos para convertir los datos de las ciencias sociales en una *forma de lectura mecánica*, una forma que puedan interpretar y manipular las computadoras y otras máquinas usadas en el **análisis cuantitativo**. Si usted realiza un proyecto de investigación cuantitativa de manera más o menos paralela a la lectura de este libro, en este momento sus datos estarán en la forma de cuestionarios llenos, hojas de código de análisis de contenidos, etc. En la etapa que cubrimos en este capítulo, usted capturaría estos datos en diskettes, disco duro o cualquier otro dispositivo que lea y analice la computadora.

A causa del ritmo del desarrollo y la difusión actuales de las computadoras, es difícil anticipar qué equipos estarán a su disposición. Por eso, voy a darle un repaso de las etapas de la evolución de las computadoras en la investigación social. De esta manera, por lo menos rozaremos la herramienta con la que usted trabaja. Además, es útil conocer los equipos y las técnicas anteriores, puesto que

revelan la lógica del análisis de datos con más claridad que las técnicas y los equipos actuales avanzados (en el mismo sentido en que se aprenden más fácilmente los fundamentos del funcionamiento de los automóviles en un VW sedán que en un Maserati del año).

Al terminar esta exposición, me parece que verá que ha elegido una gran época para estudiar investigación social. Las herramientas disponibles y las que vendrán en el futuro inmediato hacen que esta clase de investigación sea más estimulante que nunca.

Las computadoras en la investigación social

Para nuestros propósitos, la historia de la computación en la investigación social comienza en Francia en 1801, aproximadamente en las mismas fechas en que germinaban las ciencias sociales modernas. Ese año, Joseph-Marie Jacquard revolucionó la industria textil de una forma que influiría en los más insospechados rincones de la vida.

Para facilitar el tejido de diseños intrincados, Jacquard inventó un telar automático que seguía las instrucciones perforadas en unas tarjetas. A medida que una serie de tarjetas pasaba por el "lector" del telar, entraban en las perforaciones unos ganchos de madera y el telar convertía la información en diseños tejidos. Para crear diseños nuevos, Jacquard sólo tenía que perforar los hoyos

convenientes en tarjetas nuevas y el telar respondía en consecuencia.

El punto que debemos reconocer aquí es que la información (por ejemplo, un diseño de tejido) se podía *codificar y almacenar* en la forma de orificios perforados en una tarjeta, y que luego la *recuperaba* una máquina que leía las perforaciones y realizaba una acción según el significado asignado a éstas.

La siguiente etapa en nuestra historia selecta de la computación ocurrió en Estados Unidos durante el censo de 1890. Como usted sabe, la mayoría de los países realizan un censo completo de su población cada 10 años. En Estados Unidos, los censos comenzaron en 1790, con la enumeración de menos de cuatro millones de ciudadanos. Sin embargo, conforme creció la población del nuevo país, también aumentó la tarea de medirla. El censo de 1880 arrojó una cifra de más de 62 millones de habitantes, pero la Oficina del Censo se tardó *nueve años* en terminar las tabulaciones. Era evidente que se necesitaba un adelanto tecnológico antes de realizar el censo de 1890. La oficina pidió sugerencias.

Herman Hollerith, ex empleado de la Oficina del Censo, tuvo una idea. Hollerith había trabajado en el censo de 1880. Como joven profesor de ingeniería en el Instituto Tecnológico de Massachusetts, propuso adaptar las tarjetas de Jacquard a la tarea de contar a los estadounidenses. Cuando se reunieran los totales locales, se perforarían en tarjetas. Entonces, una máquina tabuladora que Hollerith inventó leería las tarjetas y calcularía los conteos de la población de todo el país.

El sistema de Hollerith se puso a prueba junto con otras propuestas, y resultó ser el más rápido. Entonces, la Oficina del Censo rentó equipo por 750 000 dólares a la nueva Tabulating Machine Company de Hollerith y se obtuvo en seis semanas el total de la población estadounidense de 1890. Dicho sea de paso, la compañía de Hollerith siguió fabricando nuevos equipos, fusionada con otras empresas, y al cabo cambió su nombre por el de International Business Machines Corporation: IBM. Para la década de 1950, las tarjetas perforadas, conocidas como *tarjetas IBM*, se habían adoptado para almacenar y recuperar datos de investigaciones sociales, y todavía se emplean, aunque muy rara vez, con ese propósito.

Introducción a las computadoras

En la actualidad, casi todos los análisis de datos se realizan con computadoras, desde las grandes *mainframes* hasta las personales pequeñas. Las computadoras hacen más que contar y clasificar para realizar cálculos intrincados y entregar presentaciones elaboradas de los resultados. Los programas de computación populares examinan diversas variables al mismo tiempo y son capaces de calcular una variedad de estadísticas. Los datos almacenados en disquetes, discos duros o CD-ROM se leen mucho más rápidamente que las tarjetas perforadas. Las computadoras calculan estadísticas complicadas con bastante más velocidad y precisión que nosotros. En realidad, las primeras "computadoras" fueron seres humanos (llamados "computadores"); cuartos llenos de matemáticos haciendo cálculos de balística militar. Cuando las máquinas se encargaron de esa función, las llamaron *computadoras*.

Muchos programas actuales sirven en particular para analizar los datos de las ciencias sociales para nombrar algunos: ABTab, AIDA, A.SAT, BMDP, CRISP, DAISY, Data Desk, DATA-X, Dynacomp, INTERSTAT, MASS, MicroCase, Microquest, Microstat, Micro-SURVEY, Minitab, POINT FIVE, P-STAT, SAM, SAS, SNAP, SPSS, STATA, STAT80, Statgraf, Statpak, StatPro, STATS PLUS, Statview, Survey Mate, SURVTAB, SYSTAT y TECPACS. Cualquiera que sea el programa que utilice, la lógica básica del almacenamiento y el análisis es la misma.

Hasta finales de la década de 1970, todos los análisis por computadora se realizaban en máquinas grandes y caras, las llamadas *mainframes*, mantenidas por centros de cómputo y algunos análisis aún se efectuaban de esa manera. Para hacer uso de estas herramientas, había que suministrar los datos y las instrucciones del análisis que uno quería. Había que llevar los datos al centro en la forma de tarjetas perforadas o en cinta magnética.

Más tarde, dos adelantos mejoraron este proceso. Los sistemas de captura remota situaban las lectoras de tarjetas y las impresoras cerca de los investigadores, lo que les permitía enviar órdenes y recibir los resultados en sitios alejado del centro de cómputo.

El *tiempo compartido* fue otro avance. En cierto sentido, las computadoras siempre han operado en tiempo compartido, es decir, que varios usuarios

aprovechan la misma máquina. Sin embargo, al principio las instalaciones de cómputo se compartían en forma serial: la máquina corría su trabajo, luego el mío y luego el de otra persona. En cambio, las computadoras más adelantadas realizan varias tareas simultáneamente. Leen mi solicitud, analizan la de usted e imprimen la de alguien más al mismo tiempo. De hecho, la enorme velocidad de las computadoras permite comprimir operaciones pequeñas (que tardan unas milésimas de segundo) en medio de tareas más largas.

Esta capacidad hizo posible que las computadoras manejen simultáneamente las solicitudes de cientos de usuarios, lo que a menudo da la impresión de que cada uno tiene toda la atención de la máquina. Estas interacciones con las computadoras comprenden terminales, dispositivos legibles que imprimen en papel o bien despliegan sus operaciones en monitores de video parecidos a las pantallas de televisión.

En las ciencias sociales, el tiempo compartido se volvió más práctico con la llegada de las terminales portátiles, no mayores que las máquinas de escribir pequeñas. Estas terminales se comunican con la computadora mediante líneas telefónicas ordinarias. Así, uno se lleva la terminal portátil a casa o de viaje alrededor del mundo.

Los primeros sistemas de telecomunicaciones requerían que uno colocara el auricular del teléfono en dos receptores de caucho de la terminal, llamados *acopladores acústicos*. Hoy, los *módems* incorporados a la computadora o situados en un dispositivo externo se conectan directamente a un enchufe telefónico común. Una vez conectado, todo lo que uno teclee en la terminal se transmite por las líneas telefónicas a la computadora. Ésta hace lo que se le pide y devuelve los resultados a la terminal.

En la actualidad estamos más familiarizados con el tiempo compartido en la forma de redes internacionales de computadoras como Internet, CompuServe, America Online y la World Wide Web. Por ejemplo, para esta edición revisada de nuestro libro intercambié un sinnúmero de comunicaciones electrónicas con colegas de todo el mundo. En los foros electrónicos encontré ejemplos de investigación y docencia. Envié por correo electrónico partes del manuscrito para someterlo a revisión y recibir comentarios. En el apéndice C examinaremos la evolución del ciberespacio con mayores detalles.

Lo más importante para esta exposición es que las redes de computadora hacen posible que un in-

vestigador en su casa de Guayaquil obtenga una copia del conjunto de datos que almacena una computadora de Buenos Aires y que los analice con un programa que se encuentra en una computadora de México. Los resultados del análisis regresan al usuario en Guayaquil y también se pueden almacenar para que los examine un colega en Moscú. Estas situaciones serán un lugar común en su carrera de investigación, y, antes de que se retire, las recordará como primitivas.

Microcomputadoras

El adelanto más útil y estimulante a la fecha han sido las *microcomputadoras*, y es probable que tenga alguna experiencia con ellas. Estas máquinas son computadoras pequeñas completas, no mucho mayores que una máquina de escribir (o incluso menores). Las operaciones se despliegan en alguna clase de pantalla de video. La información se almacena en disquetes, discos duros o CD-ROM.

La revolución de las microcomputadoras ha progresado tan rápidamente que es fácil perder de vista cuántas cosas han cambiado en un tiempo más bien corto. En la década de 1960, cuando yo era estudiante de posgrado en la Universidad de California, en Berkeley, el centro de Investigación de Encuestas tenía una IBM 1620, que ocupaba más o menos el mismo espacio que seis o siete refrigeradores. Su memoria tenía una capacidad de 24K, o aproximadamente 24 000 caracteres de información. La IBM 1620 satisfacía las necesidades de 20 a 30 investigadores activos que cada año publicaban numerosos libros y artículos.

Justo en este momento estoy trabajando en una microcomputadora *laptop* que funciona con pilas y tiene aproximadamente el tamaño de un directorio telefónico. Su memoria es de 48 megabytes, *2000 veces mayor* que la vieja 1620. La computadora de bolsillo en la que anoto direcciones y números de teléfono tiene una memoria 83 veces mayor que la IBM 1620. El reloj que llevo en la muñeca tiene un vigésimo de la memoria de aquella enorme máquina que servía para las necesidades de 30 a 40 investigadores en el Centro de Investigación de Encuestas hace no tanto tiempo.

La IBM 1620 almacenaba datos en tarjetas perforadas que contenían 80 bytes de información y en cintas magnéticas con capacidad de algunos cientos de bytes. Casi todas las microcomputadoras actuales almacenan datos en discos duros en el

orden de millones (megabytes) o incluso miles de millones de bytes (gigabytes). Por ejemplo, mi *laptop* tiene un disco duro que almacena tres gigabytes.

Por último, la llegada de los CD-ROM ha expandido enormemente la capacidad de almacenamiento de las microcomputadoras. Un disco pequeño y delgado almacena cientos de millones de bytes de datos. Las siglas ROM de CD-ROM significan *read-only memory* ("memoria de sólo lectura"), lo que indica que uno puede recuperar los datos que almacenó ahí el fabricante pero no escribir nada nuevo. La nueva generación de discos compactos ópticos permite tanto la lectura como la escritura. El CD-ROM del *Statistical Abstract* de la Oficina del Censo estadounidense es un buen ejemplo del valor para los investigadores sociales de este adelanto. También están muy difundidas las enciclopedias, como *Encarta*, de Microsoft. Quienes realizan investigaciones históricas son quizá los más beneficiados, pues disponen con facilidad de grandes volúmenes de materiales sobre diversos periodos históricos y temas específicos.

Ahora ya tiene usted un panorama de la función que cumplen las computadoras en la investigación social. En el resto del capítulo estudiaremos las etapas y las opciones para adaptar los datos de forma que se presten al análisis por computadora. Presentaremos el proceso de codificación y enumeraremos los métodos para hacer que los datos se puedan leer mecánicamente.

Codificación

Para que las computadoras realicen su magia, deben estar en posición de leer los datos que uno reunió en la investigación. Más aún, las computadoras funcionan mejor con números. Si en una encuesta un entrevistado dice que el principal problema que encara su pueblo de Asturias es la "desintegración de la capa de ozono", la computadora no entiende esta respuesta y tenemos que traducirle: un proceso llamado *codificación*. En el capítulo 12, nuestra exposición del análisis de contenidos trató el proceso de una manera muy parecida a la que nos ocupa ahora. Recuerde que en el análisis de contenidos se deben idear métodos para señalar clasificaciones o atributos concretos a párrafos, editoriales, libros, canciones, etc. En el análisis de contenidos, la codificación es inherente a la recopilación de datos o a la observación.

Para realizar un análisis cuantitativo cuando se emplean otros métodos de investigación, a menudo hay que efectuar una codificación después de recopilar los datos. Por ejemplo, los cuestionarios con reactivos abiertos dan por resultado respuestas que no son numéricas, y es preciso codificarlas para analizarlas. Asimismo, un investigador de campo acaso desee emprender un análisis cuantitativo basado en sus notas cualitativas; digamos que se trata de usted, que quiere cuantificar las entrevistas abiertas que realizó a los participantes de algún acontecimiento social que investigaba.

Al igual que con el análisis de contenidos, aquí la tarea es reducir una gran variedad de unidades heterogéneas de información a un conjunto más limitado de atributos que compongan una variable. Por ejemplo, supongamos que un investigador realiza una encuesta en la que le pregunta a los entrevistados: "¿Cuál es su ocupación?" Las respuestas variarían considerablemente. Aunque sería posible asignar a cada ocupación un número exclusivo de código, este procedimiento no facilitaría el análisis, que depende de que varios sujetos tengan el mismo atributo.

La variable "ocupación" tiene muchos esquemas de codificación preestablecidos, pero ninguno es muy bueno. Un esquema distingue entre profesionistas, directivos, oficinistas, puestos de capacitación media, etc.; otro, sectores de la economía: manufactura, salud, educación, comercio, etc.; otros más combinan los dos.

El esquema en que codifique la ocupación debe ser apropiado para los conceptos teóricos examinados en el estudio. Para algunos estudios, codificar las profesiones como fabriles o de oficina será suficiente. Para otros, bastará distinguir entre empleados y quienes trabajan por su cuenta. Asimismo, en una investigación sobre la paz sólo sería deseable saber si la ocupación depende de una dependencia de Defensa o no.

Aunque el esquema de codificación debe adaptarse para cumplir los requisitos particulares del análisis, hay que tener presente un lineamiento general. Si se codifican los datos de modo que se conserven muchos detalles, las categorías podrían combinarse durante los análisis que no necesiten tantos pormenores. Sin embargo, si se codifican los datos en unas cuantas categorías generales no hay forma de recrear los detalles originales durante el análisis. Así, es un buen consejo que codifique sus datos con un poco más de detalle del que piensa utilizar en el análisis.

Elaboración de categorías de codificación

Hay dos planteamientos básicos en el proceso de codificación. Primero, se comienza con un esquema relativamente completo tomado del propósito de la investigación. Así, como indicamos, el investigador de la paz codificaría las ocupaciones de acuerdo con su relación con las dependencias de Defensa. O supongamos que usted practica la observación participante de una nueva religión. Usted ha tomado notas cuidadosas de las razones que esgrimen los miembros para afiliarse. Quizá usted tenga la impresión de que los nuevos miembros consideran esa religión como un sustituto de la familia. Entonces, revisaría sus notas con atención para codificar los comentarios de estos sujetos según sus afirmaciones acerca de este aspecto. Asimismo, podría codificar sus comentarios a partir del hecho de que tengan familia o no.

Si usted tiene la suerte de contar con asistencia en el proceso de codificación, su tarea sería perfeccionar las definiciones de las categorías de codificación y mostrarle a sus codificadores la manera de asignar cada respuesta a la categoría adecuada. Debe explicar el significado de tales categorías y dar ejemplos de cada una. Para verificar que sus codificadores entienden por completo lo que usted pretende, debe codificar varios casos, pedirles que codifiquen los mismos sin saber cómo lo hizo usted y luego comparar su trabajo con el suyo. Cualquier discrepancia indicará una comunicación imperfecta del esquema a los codificadores. Sin embargo, aunque exista un acuerdo completo entre usted y los codificadores, debe verificar el código de por lo menos algunos casos durante todo el proceso.

Por otro lado, si no tiene la fortuna de que lo asistan en la codificación, de todos modos debe conseguir alguna verificación de su propia confiabilidad en tanto que codificador. Nadie es perfecto, sobre todo el investigador tras la pista de un hallazgo. Supongamos que en su estudio de la religión nueva usted tiene la impresión de que quienes no tienen una familia serán los más proclives a ver en aquella un sucedáneo. El peligro es que cada vez que localice a un sujeto que no tiene familia, trate inconscientemente de encontrar pruebas en sus comentarios de que la nueva religión suple a la familia que le falta. Entonces, en la medida en que le sea posible, consiga que alguien codifique algunos casos para ver si hace las mismas asignaciones que usted (advierta que esto se relaciona con la intersubjetividad de las ciencias).

El segundo planteamiento sobre la codificación es apropiado cuando al principio uno no está seguro de la forma de codificar los datos; es decir, cuando uno no sabe qué variables representar entre los sujetos de estudio. Por ejemplo, supongamos que usted formula esta pregunta: "¿Qué opina de que las mujeres tengan el derecho a abortar por cualquier razón?" Aunque haya anticipado la codificación de las respuestas como positivas, negativas y neutras, es poco probable que conjeture de antemano toda la gama de respuestas posibles. En esta situación, sería útil preparar una lista de unas 50 a 100 respuestas reales a esta pregunta abierta. Entonces, repasaría la lista y advertiría las dimensiones que manifiestan las respuestas. Quizá encontraría que varias de las respuestas positivas contienen referencias a las libertades civiles, y quizá algunas de las respuestas negativas remitan a convicciones religiosas.

Cuando haya preparado un esquema de codificación basado en la lista de 50 a 100 respuestas, verificaría que cada una encajara en alguna de las categorías de codificación. Entonces estaría listo para codificar el resto de las respuestas. Si cuenta con asistentes codificadores, los comentarios anteriores sobre su capacitación y supervisión vienen a cuento aquí; en caso contrario, retome la recomendación de que alguien verifique su trabajo.

Al igual que los atributos que componen una variable, y que las categorías de respuesta en los cuestionarios de reactivos cerrados, las categorías de codificación deben ser exhaustivas y mutuamente excluyentes. Toda información codificada debe corresponder a una y sólo una categoría. Surgen problemas cada vez que una respuesta parece casar igualmente bien en más de una categoría o, por el contrario, cuando no encuentra lugar en ninguna.

Construcción del libro de codificación

El producto final del proceso de codificación es la conversión de los datos en códigos numéricos. Estos códigos representan atributos que componen variables y se les asignan nombres y localizaciones en el archivo de datos. El *libro de codificación* es el documento que señala la localización de las variables y que enlista las asignaciones de código a los atributos que las componen. Los libros de codi-

Figura 14.1

Parte de un libro de codificación

POSPOL	ASIS
En estas épocas oímos hablar mucho de liberales y conservadores. Voy a mostrarle una escala de siete puntos con las posturas políticas comunes dispuestas de extremadamente liberal —punto 1— a extremadamente conservador, —punto 7—. ¿Dónde se ubicaría usted en la escala?	¿Con qué frecuencia asiste a las ceremonias religiosas?
1. Extremadamente liberal	0. Nunca
2. Liberal	1. Menos de una vez al año
3. Ligeramente liberal	2. Una o dos veces al año
4. Moderado, en el centro	3. Varias veces al año
5. Ligeramente conservador	4. Alrededor de una vez al mes
6. Conservador	5. Dos a tres veces al mes
7. Extremadamente conservador	6. Casi cada semana
8. No sé	7. Cada semana
9. No respondió	8. Varias veces a la semana
	9. No sé; no respondió

ficación cumplen dos funciones esenciales. Primera, es la guía fundamental para la codificación. Segunda, es la guía para localizar las variables e interpretar los códigos del archivo de datos durante el análisis. Si usted decide correlacionar dos variables como parte del análisis de sus datos, el libro le indicará dónde encontrarlas y verá lo que significan los códigos.

La figura 14.1 es parte de un libro de codificación elaborado para dos variables por la Encuesta Social General. Aunque no hay un único formato correcto para los libros de codificación, en este ejemplo se presentan algunos de los elementos comunes.

Hay varios elementos que vale la pena notar en el fragmento del libro de codificación que se ilustra en la figura. Primero, cada variable se identifica con un nombre abreviado: POSPOL, ASIS. Por ejemplo, podemos determinar la asistencia a la iglesia de los entrevistados con la referencia ASIS. En este ejemplo, tomamos el formato establecido por la Encuesta Social General, que ha sido transferido a SPSS. Tenga presente que otros conjuntos de datos y programas de análisis pueden dar otro formato a las variables. Así, algunos optan por códigos numéricos en lugar de nombres abreviados. En cualquier caso, debe tener *algún* identificador que le permita localizar y utilizar la variable que busca.

Además, todo libro de codificación debe contener la definición completa de la variable. En el caso de los cuestionarios, consistiría en la redacción exacta de las preguntas formuladas, porque esto,

como hemos visto, influye en buena medida en las respuestas que se consiguen. En el caso de POSPOL, usted sabe que se extendió a los entrevistados una tarjeta con varias categorías políticas y se les pidió que eligieran la que les correspondiera mejor.

Su libro de codificación también debe indicar los atributos que componen cada variable. Así, en POSPOL los entrevistados caracterizarían sus posturas políticas como "extremadamente liberal", "liberal", "ligeramente liberal", etc. Por último, observe que cada atributo tiene también un número; por ejemplo, "extremadamente liberal" en POSPOL tiene la categoría de codificación "1". Estos códigos numéricos sirven para diversas manipulaciones de los datos; por ejemplo, digamos que usted decide combinar las categorías 1 a 3 (todas las respuestas "liberales"). Es más fácil hacerlo con números de código que con nombres largos.

Opciones de codificación y captura de datos

Hace años, cualquier captura de datos adoptaba la forma de perforación manual de tarjetas que se analizaban con el equipo de registro unitario, o bien se leían con computadoras para realizar análisis más complicados. Todo esto ha cambiado. Ahora los datos se teclean directamente en los archivos almacenados en los discos de las computadoras. Sin embargo, al igual que antes, esta actividad se

Figura 14.2

Parte de una hoja de transferencia de códigos

01	02	03	04	05	06	07	08	09	10	71	72	73	74	75	76	77	78	79	80
0	0	1	3	7	8	9	3	1	1	4	5	2	1	1	7	8	7	1	2
0	0	2	4	2	4	2	4	1	2	2	2	5	1	1	2	8	2	2	2
0	0	3	5	6	1	2	3	1	1	1	3	6	2	1	3	9	2	2	3
0	0	4	6	3	4	4	2	2	1	1	4	1	0	3	6	0	4	2	1

relaciona mucho con la codificación, y hay varios métodos para establecer este vínculo. Veamos algunas de las posibilidades.

Hojas de transferencia

El método tradicional de procesamiento de datos comprende la codificación de éstos y el vaciado de las asignaciones de código a unas *hojas de transferencia* o de *código*. Estas hojas se cuadrículaban en 80 columnas que equivalían a las columnas de las tarjetas, pero se pueden adaptar a las configuraciones adecuadas para el método de captura de datos elegido (la figura 14.2 ofrece una ilustración). Los codificadores escriben los números que corresponden a las categorías establecidas en las columnas apropiadas. Entonces, las hojas de código se usan para introducir los datos en las computadoras. Esta técnica aún es útil cuando se procesan cuestionarios particularmente complicados u otros documentos con datos.

Codificación al margen

La *codificación al margen* se utiliza a veces para sortear la necesidad de hojas de transferencia. El margen externo de cada página del cuestionario o del documento que contiene los datos se deja en blanco o se marca con espacios correspondientes a los nombres o números de las variables. En lugar de transferir las asignaciones de código a una hoja aparte, se anotan los códigos en los espacios apropiados al margen. Entonces, se introducen los datos a partir de los documentos codificados al margen.

Captura directa de datos

Si los cuestionarios u otras formas de recopilación de datos están bien diseñados, es posible capturar directamente los datos en la computadora sin recurrir a hojas de código ni a la codificación al margen. Un cuestionario precodificado contendría indicaciones de las columnas y los códigos que se deben asignar a preguntas y respuestas; con ello, los datos se introducirían de manera directa.

Datos capturados por los entrevistadores

Ya tratamos en el capítulo 10 el método más directo de captura de datos en las encuestas: las entrevistas telefónicas asistidas por computadora, o ETAC. Como recordará, los entrevistadores portan teléfonos de manos libres y se sientan frente a terminales de computadora que despliegan las preguntas que hay que formular y teclean las respuestas de los entrevistados. De esta manera, los datos se introducen directamente en los archivos en el momento en que se generan. Los datos de preguntas cerradas están listos para su análisis inmediato. Los datos de preguntas abiertas también se capturan, pero requieren pasar antes por otra etapa.

Digamos que un cuestionario pregunta a los entrevistados: "¿Cuál cree que es el principal problema que enfrenta el país?" La terminal de computadora solicitará al entrevistador que formule la pregunta. Luego, en lugar de aguardar como entrada un código numérico, esperaría a que el entrevistador tecleara todo lo que diga el entrevistado; por ejem-