

Codificación, entrada y verificación de datos

Víctor E. Bonilla Rodríguez, Ph.D.
Investigación y Evaluación Educativa (INEVA)
Departamento de Estudios Graduados, Facultad de Educación

Ciclo de Investigación Cuantitativa
Centro para la Excelencia Académica
Centro de Cómputos, Facultad de Ciencias Sociales
1 de abril de 2011, 8:30am-12:00pm

Objetivos del taller

- Conocer las ventajas de la codificación de datos.
- Distinguir entre los códigos asignados a una variable y los valores originales.
- Conocer las escalas de medición y su importancia al realizar análisis.
- Entrar datos en Excel.
- Pasar datos de Excel a SPSS.
- Distinguir entre las funciones de las diferentes pantallas de SPSS.
- Conocer y aplicar diferentes estrategias para verificar los datos en SPSS.
- Pasar los resultados de SPSS a Word.

Codificación de datos

- Asignar códigos numéricos a respuestas que no sean numéricas agiliza el proceso de entrada de las mismas ya que:
 - podemos utilizar el “keypad” (parte derecha del “keyboard”) para entrarlas en lugar de estar utilizando la segunda línea del “keyboard”. Esta es la forma en que los “tellers” de un banco entran las cantidades de las transacciones.
 - evita que cometamos muchos errores en la entrada de los datos. Por ejemplo, es más fácil escribir un número 1 para femenino y un 2 para masculino que escribir la palabra completa o las letras F o M cuando entramos los datos de la variable género.

Escala de medición

- A pesar de que asignemos códigos numéricos a todas las respuestas de las preguntas de, por ejemplo, un cuestionario, no podemos olvidar lo que significan dichas respuestas.
- En otras palabras, no podemos olvidar la naturaleza o escala de medición de las variables.
- Por ejemplo, aunque codifiquemos el género como 1 = femenino y 2 = masculino, no tiene sentido pedir al programado estadístico calcular el promedio de esta variable.
- Esto implica que la naturaleza de las variables es uno de los factores a considerar en la selección del análisis estadístico más apropiado.
- Debemos reconocer cuándo los valores implican orden, responden a variables continuas o a variables discretas.

Escala de medición (cont.)

- La clasificación más utilizada fue desarrollada por Stevens en el 1946.
- La clasificación establece 4 escalas de medición. Las escalas se presentan de menor a mayor precisión.
 - Nominal
 - Ordinal
 - Intervalar
 - Razón
- Las dos primeras, nominal y ordinal, representan variables discretas. Las dos últimas, intervalar y de razón, representan variables continuas.

Escala de medición (cont.)

- Nominal
 - Clasifica pero no establece orden
 - Ejemplos: género, partidos políticos
- Ordinal
 - Clasifica y establece orden.
 - Ejemplos: nota en letra, nivel educativo, Likert
- Intervalar
 - Clasifica, establece orden y posee intervalos equidistantes. No posee un cero absoluto.
 - Ejemplos: IQ, temperatura en Celsio y Farenheit
- Razón
 - Clasifica, establece orden, posee intervalos equidistantes y un cero absoluto
 - Ejemplos: peso, masa, longitud

Entrada de datos en Excel

- Entrar los datos en Excel presenta la ventaja de ser un programado que la mayoría de las personas tiene instalado en su computadora.
- Además, la mayoría de los programados estadísticos proveen la opción de importar datos de Excel.
- Por tanto, puede entrar los datos en Excel y, una vez termine esta fase, llevarlos al lugar donde tenga disponible SPSS para realizar los análisis.

Entrada de datos en Excel (cont.)

- Para facilitar la transición, visualice las filas como un récord para cada cuestionario y las columnas como las variables.
- Coloque el nombre de las variables en la primera fila.
- Coloque el número de identificación del cuestionario en la primera columna. Este número es asignado según recibe los cuestionarios. Recuerde que al enviarlos no pueden tener nada que indiquen que se puede identificar al que lo llena (confidencialidad y anonimato).
- Cada celda (intersección entre la fila y la columna) contiene la respuesta obtenida en un cuestionario específico para una variable en particular.
- No entre fórmulas ni otra información que no sea parte de los datos.

Pasando los datos de Excel a SPSS

- Para leer los datos de un archivo de Excel en SPSS debe realizar los siguientes pasos:
 - Seleccione el menú de “File”
 - Seleccione la opción de “Open”
 - Seleccione la opción de “Data”
 - En la pantalla que aparece, cambie la opción de “Files of type:” de “SPSS (*.sav)” a “Excel (*.xls)”.
 - Busque el lugar donde guardó el archivo, selecciónelo y presione el botón de “Open”.
 - En la pantalla que aparece, asegúrese que la cajita que lee “Read variable names from the first row of data.” esté seleccionada.
 - Oprima el botón de “OK”.

Pantallas de SPSS

- SPSS cuenta con cuatro pantallas principales: 1) “Data View”, 2) “Variable View”, 3) “SPSS Viewer” y 4) “Syntax Editor”.
- “Data View”
 - En esta pantalla se entran los datos.
 - Las filas son los récords y las columnas son las variables.
- “Variable View”
 - En esta pantalla se definen las variables.
 - Puede colocar nombres más largos a sus variables (“variable labels”). Por ejemplo, escribir la pregunta completa como nombre para la variable.
 - Además, puede asignar nombres a los valores a una variable (“value labels”). Por ejemplo, 1 = “femenino”.

Pantallas de SPSS (cont.)

- “SPSS Viewer”
 - Aparece cuando se realiza un análisis.
 - Es el lugar donde se presentan los resultados. La pantalla se divide en dos partes: 1) la tabla de contenido (lado izquierdo) y 2) los resultados (lado derecho).
 - La tabla de contenido puede ser utilizada para saltar de una sección a otra de los resultados.
 - Si realiza todos sus análisis en una misma sesión, sin salvar y cerrar el “output”, todos los análisis aparecerán en el “output” en el orden en que se realizaron.

Pantallas de SPSS (cont.)

- “Syntax Editor”
 - Permite ver las instrucciones que utiliza SPSS para realizar los análisis (e.g., desplegar una tabla de frecuencia).
 - Esta pantalla aparece si presiona el botón que lee “Paste” en cualquier análisis.
 - Puede guardar la sintaxis utilizada para ejecutar los análisis si lo desea. Esto le permitirá recordar lo que realizó, incluirlo como texto en su sección de análisis o volver a realizar el mismo análisis en otra ocasión sin tener que seleccionar todas las opciones nuevamente.

Verificación de datos

- Utilizar códigos numéricos para la entrada de datos, no sólo facilita la entrada de los mismos, sino que también facilita la identificación de errores en el proceso.
- Existen dos formas sencillas de determinar errores en la entrada de datos, cuando los errores se encuentran fuera del rango de posibles valores. Por ejemplo, entrar un cero o un tres como respuesta a la variable género, cuando esta variable fue codificada como 1 ó 2.
- El primer método implica el uso de “Value labels” que se encuentra bajo el menú de “View” en SPSS.
- Esta opción cambia la presentación de los datos en la pantalla de “Data View” de los números a los “labels” asignados a cada respuesta en la pantalla de “Variable View”. De esta forma, cualquier valor que se quede con un número, en vez de la palabra, contiene un error.

Verificación de datos (cont.)

- Otro método para identificar errores en la entrada de datos es solicitar al programado que le presente tablas de frecuencia.
- Utilizando tablas de frecuencia podemos encontrar instancias en las que nos equivocamos ya que no entramos datos o entramos un dato fuera del rango de posibles valores.
- Una vez que encontramos un error, podemos ordenar la base de datos (de forma ascendente o descendente según sea más conveniente) respecto a esa variable para facilitar la localización del error y contrastarlo con el cuestionario al que pertenece.

Verificación de datos (cont.)

- Al organizar la base de datos de otra forma se evidencia la utilidad de asignar a cada cuestionario un número según entremos los datos del mismo a la base de datos y colocar el número del cuestionario en la primera columna de la base de datos como una variable.
- Si no hubiese colocado un número a cada cuestionario, no sería capaz de parrear la información de la base de datos con la de los cuestionarios para poder corregir los errores.
- Además, el número de las filas que provee SPSS le indica si entró los datos de todos los cuestionarios. Por ejemplo, si sabe que eran 50 cuestionarios y sólo llenó hasta la fila 49 en SPSS, será obvio que olvidó entrar alguno.

Verificación de datos (cont.)

- Obteniendo tablas de frecuencia
 - Las tablas de frecuencia se obtienen bajo el menú de “Analyze”.
 - De este menú debe seleccionar “Descriptive Statistics” y luego “Frequencies”.
 - En la pantalla que aparece debe seleccionar la(s) variable(s) de la(s) cual(es) desea obtener la tabla de frecuencia y pulsar el botón de “OK”.
- Ordenando (“sorting”) los datos
 - Para ordenar los datos respecto a una variable, seleccione “Sort Cases” del menú de “Data”.
 - En la pantalla debe seleccionar la variable y si desea un orden ascendente o descendente. Luego debe presionar el botón de “OK”.

Verificación de datos (cont.)

- Un tipo de error más difícil de detectar es el que se encuentra dentro del rango de posibles valores. Por ejemplo, el género es 1 y usted entra un 2. Sin embargo, como el género puede ser 1 ó 2, el valor entrado, a pesar de ser incorrecto, se encuentra dentro del rango de posibles valores.
- Si la cantidad de cuestionarios no es muy grande y los cuestionarios no son muy extensos, se recomienda que verifique todos los datos entrados.
- Cuando el volumen de datos es muy grande, la literatura recomienda seleccionar de forma aleatoria un 10% de los cuestionarios para verificarlos completamente (Kent, 2001, p. 62). Si no encuentra errores, no hay mucha probabilidad de que haya cometido muchos errores.

Verificación de datos (cont.)

- Para seleccionar una muestra aleatoria de los cuestionarios utilizando SPSS:
 - Vaya al menú de "Data" y seleccione la opción de "Select Cases".
 - En la pantalla que aparece seleccione "Random sample of cases". Oprima el botón que lee "Sample". En la pantalla que aparece entre el porcentaje de casos que desea seleccionar y oprima el botón de "OK".

Pasando los resultados de SPSS a Word

- Una vez haya concluido el análisis de los datos puede pasarlos a Word.
- Esto le permite tener acceso a sus resultados independientemente de que posea o no el programado SPSS.
- Para pasar los resultados de SPSS a Word debe:
 - Seleccionar la porción del “output” (tablas o gráficas) que desea pasar a Word.
 - Seleccionar del menú “Edit” la opción de “Copy Objects” (o presionar CTRL-K).
 - Abra un documento nuevo en Word.
 - Seleccione del menú “Edit” la opción de “Paste” (o presione CTRL-V).

Referencias

- Fowler, F. J., Jr. (2002). *Survey research methods* (3rd ed.). Thousand, Oaks, CA: Sage Publications.
- Kent, R. (2001). *Data construction and data analysis for survey research*. New York, NY: Palgrave.
- Levesque, R. & SPSS Inc. (2006). *SPSS programming and data management, 3rd edition: A guide for SPSS and SAS Users*. Chicago, IL: SPSS Inc.
- Norusis, M. J. (2007). *SPSS 15.0 guide to data analysis*. Upper Saddle River, NJ: Prentice Hall.
- Stevens, S. S. (1946). *On the theory of scales of measurement*. *Science*, 103, 677-680.
- Tabachnick, B. G. & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.