

Pautas para el diseño de una base de datos en “EXCEL” o “LibreOffice Calc” para su análisis con SPSS

Para el análisis estadístico de un conjunto de datos recogidos por un investigador se utilizan diversos programas específicamente diseñados para este fin (SPSS, R, STATA, etc.). Cuando se introducen datos en estos programas, la estructura siempre es rectangular, con filas y columnas, de tal forma que las filas recogen los casos (registros) y las columnas las variables (campos). En el siguiente ejemplo se puede ver la imagen la estructura de una base de datos de SPSS:

| | | Variables | | | | | | | |
|-------|----|-----------|------|------|-------|--------|-------|--------|---------|
| | | ID | edad | SEXO | ESTAN | INTERV | DURAC | EXITUS | MICROOR |
| Casos | 1 | 1003 | 40 | 1 | 10 | 0 | 178 | 0 | 16 |
| | 2 | 1004 | 73 | 1 | 6 | 1 | 236 | 0 | 15 |
| | 3 | 1010 | 77 | 1 | 9 | 0 | 135 | 0 | 14 |
| | 4 | 1012 | 21 | 1 | 4 | 1 | 144 | 0 | 5 |
| | 5 | 1015 | 53 | 1 | 11 | 1 | 211 | 1 | 2 |
| | 6 | 1016 | 70 | 1 | 12 | 0 | 250 | 0 | 7 |
| | 7 | 1017 | 32 | 0 | 9 | 0 | 172 | 0 | 7 |
| | 8 | 1018 | 63 | 0 | 10 | 0 | 216 | 0 | 1 |
| | 9 | 1031 | 59 | 0 | 10 | 0 | 168 | 0 | 1 |
| | 10 | 1036 | 51 | 0 | 8 | 1 | 166 | 0 | 7 |
| | 11 | 1061 | 52 | 1 | 5 | 0 | 241 | 0 | 15 |
| | 12 | 1074 | 23 | 1 | 1 | 1 | 218 | 1 | 3 |
| | 13 | 1076 | 70 | 1 | 12 | 1 | 189 | 0 | 5 |
| | 14 | 1078 | 50 | 0 | 8 | 0 | 224 | 0 | 3 |
| | 15 | 1079 | 51 | 1 | 5 | 0 | 183 | 0 | 16 |
| | 16 | 1081 | 58 | 0 | 8 | 0 | 178 | 0 | 9 |
| | 17 | 1090 | 76 | 1 | 7 | 0 | 262 | 0 | 5 |

Ejemplo 1

Estos programas traen incorporado un editor / visor de datos, cuya forma es similar a una hoja de cálculo estándar (EXCEL, LibreOffice Calc, etc.), aunque sus funcionalidades son mucho más limitadas. Podemos resumir así la estructura de datos de estos programas:

- Las filas son casos. **Cada fila representa un caso o una observación.** Por ejemplo, cada individuo que responde a un cuestionario es un caso.
- Las columnas son variables. **Cada columna representa una variable** o una característica que se mide. Por ejemplo, cada elemento de un cuestionario es una variable.
- Las celdas contienen valores. **Cada celda contiene un valor único** de una variable para cada caso. La celda se encuentra en la intersección del caso y la variable. Las celdas sólo contienen valores de datos.
- El archivo de datos es rectangular. Las dimensiones del archivo de datos vienen determinadas por el número de casos y de variables.

La hoja de cálculo es la herramienta que se usa con más frecuencia para la recogida de datos por parte de los investigadores. Sin embargo, para que uno de estos archivos pueda ser importado por SPSS, debe tener una estructura determinada cumpliendo una serie de requisitos que garanticen que no se produzcan errores al ser leídos.

Todas las recomendaciones que se exponen a continuación son válidas tanto para ‘Microsoft EXCEL’ como para ‘LibreOffice Calc’. Sin embargo, si se usa este último software, es recomendable guardar el archivo con el formato ‘Microsoft Excel 97/200/XP/2003’ (*.xls) o ‘XML de Microsoft Excel 2007/2010’ (*.xlsx).

PREPARACIÓN DE UNA HOJA DE CÁLCULO PARA SER IMPORTADA POR SPSS

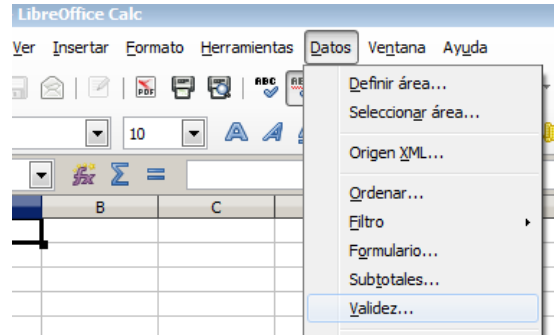
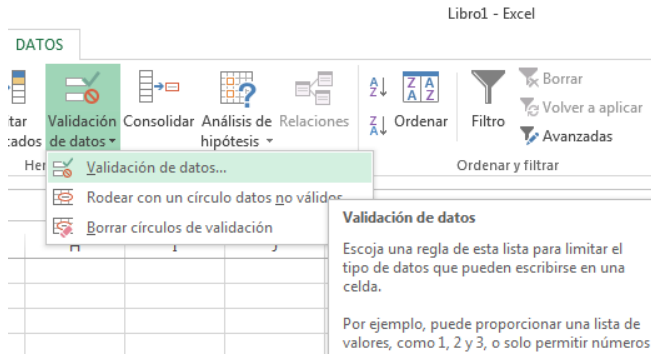
1. Tal y como se comentó anteriormente, la **estructura será rectangular**. Las filas, con la excepción de la primera, se corresponderán con los casos y las columnas con las variables.
2. **En la primera fila se especificarán los nombres de las variables**, que serán compatibles con las convenciones de SPSS. Los nombres de las variables sólo podrán estar en la primera fila, no extienda los nombres a la segunda fila.
 - a. Los nombres de variables deben empezar por una letra. No podrán contener espacios en blanco ni caracteres especiales (por ejemplo, ñ, ¡, ´, *). Tampoco podrán terminar con un punto.
 - b. Los nombres de variables no podrán tener más de 64 caracteres.
 - c. No se deben usar como nombre de variables palabras clave internas de SPSS, como: ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH.
 - d. No se pueden repetir los nombres de variables, estos deben ser únicos. SPSS no distingue mayúsculas de minúsculas, así que nombres de variables como *EDAD* y *Edad* se consideran idénticos.
 - e. La mayor parte de los errores en los nombres de las variables son corregidos automáticamente por SPSS en el proceso de importación, asignando un nombre compatible con sus convenciones y añadiéndole una etiqueta con el nombre erróneo original.
 - f. Es recomendable utilizar nombres de variables claros e intuitivos.
 - g. Es recomendable que la primera variable (columna A), asigne a cada caso (fila) un identificador único, de tal forma que sea fácil identificarlo.
3. Cada fila (con la excepción de la primera, que contiene los nombres de las variables) se corresponderá con un caso. Evite las filas en blanco.
4. Variables cuantitativas: Las celdas deben tener formato numérico, para evitar errores comunes recuerde que:
 - a. La celda solo puede contener un número, no añada texto (como unidades de medida, aclaraciones, etc., si lo hace la importación lo tratará como un error.
 - b. Utilice el separador decimal correcto. El separador decimal (,) o (.) es configurable. Si utiliza uno incorrecto se interpretará el contenido como texto y en la importación se tratará como un error. Normalmente lo notará porque cambiará la alineación del contenido de la celda (a la izquierda si es texto, a la derecha si es número).
5. Codificación de variables cualitativas: Es recomendable registrar las variables cualitativas en formato numérico, evita errores y es más rápido el proceso de registro. Tenga en cuenta que SPSS permite etiquetar con texto valores numéricos.
 - a. Variables cualitativas nominales: son aquellas variables cualitativas que no tienen ningún orden natural. Se le asignará un código numérico simple a cada una de las categorías, sin importar el orden. Ejemplo: la variable 'sexo' se puede codificar con 0 y 1, asignando el código 0 al sexo femenino y el código 1 al sexo masculino; la variable 'hábito tabáquico' con



categorías 'no fumador', 'exfumador' y 'fumador', puede codificarse como 0 (no fumador), 1 (exfumador) y 2 (fumador).

- b. Variables cualitativas ordinales: son aquellas variables cualitativas con un cierto orden natural, aunque el tamaño de los intervalos no suele estar definido, por ejemplo la valoración del grado de intensidad de un proceso como 'leve', 'intermedio', 'moderado', 'grave' o la fuerza muscular entre los grados 1 (sin movimiento) y 5 (fuerza normal). Se recomienda asignarles un código numérico simple que respete el orden natural. En el caso de la valoración de grado de intensidad que se puso antes como ejemplo, podría codificarse como 0 (leve), 1 (intermedio), 2 (moderado) y 3 (grave).
6. Configuración de las fechas: las celdas donde se registren fechas deben tener asignado este formato. Hay que tener en cuenta que algunos formatos de fecha pueden variar en función de la configuración regional del sistema operativo.
 7. Valores perdidos (missing values): si una celda donde debe recogerse el valor de una variable está vacía, SPSS considerará ese valor como *perdido* solo si la variable es numérica. Ya que SPSS permite tratar a determinados valores como *perdidos* excluyéndolos del análisis, es preferible registrarlos con un valor singular, por ejemplo 0 (varón) 1 (mujer) 9 (no consta). En ocasiones este método nos permitirá distinguir distintas causas de ausencia de datos, por ejemplo 0 (leve), 1 (intermedio), 2 (moderado), 3 (grave), 8 (no procede) y 9 (no consta). El mismo criterio se puede aplicar a variables cuantitativas.
 8. Otras recomendaciones:
 - a. Asegúrese de que las celdas de la hoja de cálculo que están fuera del rectángulo de datos están vacías. No incluya ninguna nota aclaratoria, texto, etc.
 - b. Utilice una sola hoja, no extienda una base de datos en varias hojas de cálculo.
 - c. Es recomendable elaborar un documento donde se recoja:
 - i. Si no se ha elaborado un protocolo de investigación, el significado de cada una de las variables, sus unidades de medida, opciones de respuesta, criterios para declarar un valor como perdido, etc.
 - ii. En el caso de variables cualitativas codificadas como numéricas, la etiqueta que identifica la característica asociada a cada uno de los valores numéricos asignados.
 - d. Asegúrese de introducir una variable que le permita identificar con rapidez un caso para corregir posibles errores cometidos al registrar el dato en la hoja de cálculo. Cuando se trate de datos de carácter personal (i.e. datos clínicos), al enviarlos para su análisis, debe garantizar que nadie, excepto el investigador, puede identificar la persona a la que pertenecen los datos registrados a través del identificador utilizado.
 - e. Utilice herramientas que garanticen que el formato del valor introducido es el correcto. Tanto 'Microsoft EXCEL' como 'LibreOffice Calc' permiten configurar un determinado rango de celdas de tal forma que solo se permita introducir datos con un formato determinado. Para ello, antes de introducir los datos, seleccione el rango de celdas y, usando, dentro del menú 'DATOS', la opción 'Validación de datos' ('Validéz' en 'LibreOffice Calc'), limite el tipo de datos que pueden introducirse en la celda.





9. Ejemplo: en la siguiente imagen puede ver cómo podría verse la base de datos del ejemplo 1 en una hoja de cálculo de EXCEL:

Variable Identificador

Variables

Nombre de las variables

Casos

| | A | B | C | D | E | F | G | H | I |
|---|------|------|------|-------|--------|-------|--------|---------|---|
| 1 | ID | edad | SEXO | ESTAN | INTERV | DURAC | EXITUS | MICROOR | |
| 2 | 1003 | 40 | 1 | 10 | 0 | 178 | 0 | 16 | |
| 3 | 1004 | 73 | 1 | 6 | 1 | 236 | 0 | 15 | |
| 4 | 1010 | 77 | 1 | 9 | 0 | 135 | 0 | 14 | |
| 5 | 1012 | 21 | 1 | 4 | 1 | 144 | 0 | 5 | |
| 6 | 1015 | 53 | 1 | 11 | 1 | 211 | 1 | 2 | |
| 7 | 1016 | 70 | 1 | 12 | 0 | 172 | 0 | 7 | |
| 8 | 1017 | 32 | 0 | 9 | 0 | 172 | 0 | 7 | |

Valor de la variable SEXO par el caso 1012: 1 (Varón)