



Fichas para investigadores

¿Cómo hacer un análisis cuantitativo de datos de tipo descriptivo con el paquete estadístico SPSS?

Autora:

Vilà Baños, Ruth. Departament de Mètodes d'investigació i Diagnòstic en Educació. Universitat de Barcelona.

El análisis cuantitativo de datos mediante la informática estadística

En la actualidad es difícil hablar de procesos estadísticos de análisis cuantitativo de datos sin hacer referencia a la informática aplicada a la investigación. Estos recursos informáticos se basan a menudo en paquetes de programas informáticos que tienen por objetivo ayudar al/la investigador/a en la fase de análisis de datos. En estos momentos uno de los paquetes más perfeccionados y completos es el SPSS (Statistical Package for the Social Sciences).

El SPSS es un paquete de programas para llevar a cabo el análisis estadístico de los datos. Constituye una aplicación estadística muy potente, de la que se han ido desarrollando diversas versiones desde sus inicios, en los años setenta. En esta ficha las salidas de ordenador que se presentan corresponden a la versión 11.0.1. No obstante, aunque la forma ha ido variando desde sus inicios, su funcionamiento sigue siendo muy similar entre las diferentes versiones.

Antes de iniciarnos en la utilización de las aplicaciones del SPSS es importante familiarizarse con algunas de las ventanas que más usaremos. Al entrar al SPSS lo primero que nos encontramos es el editor de datos. Esta ventana visualiza, básicamente, los datos que iremos introduciendo. El editor de datos incluye dos opciones: la *vista de los datos* y la de las *variables*. Estas opciones pueden seleccionarse a partir de las dos pestañas que se presentan en la parte inferior. La *vista de datos* contiene el menú general y la matriz de datos. Esta matriz está estructurada ubicando los casos en las filas y las variables en las columnas.

La matriz de datos

En una investigación, una vez finalizado el proceso de recogida de datos, toma protagonismo la creación de la matriz de datos (que se encuentra en el *visor de edición de datos* del SPSS). Una matriz de datos es una estructura en forma de tabla que contiene los valores de cada sujeto en las diferentes variables. Habitualmente se colocan las variables en las columnas y los individuos en las filas, aunque no tienen por qué coincidir el número de variables con el de columnas, o el número de filas con el número de individuos.

Crear la matriz de datos implica definir las variables que se consideran para el estudio e introducir los datos recogidos. La vista de variables nos permite hacer especificaciones



sobre las variables que contiene nuestra matriz de datos. En esta ventana las diferentes variables se encuentran situadas en las filas, mientras que las características a definir de las mismas se encuentran en las columnas. Concretamente, podemos definir los siguientes aspectos de cada una de las variables del estudio:

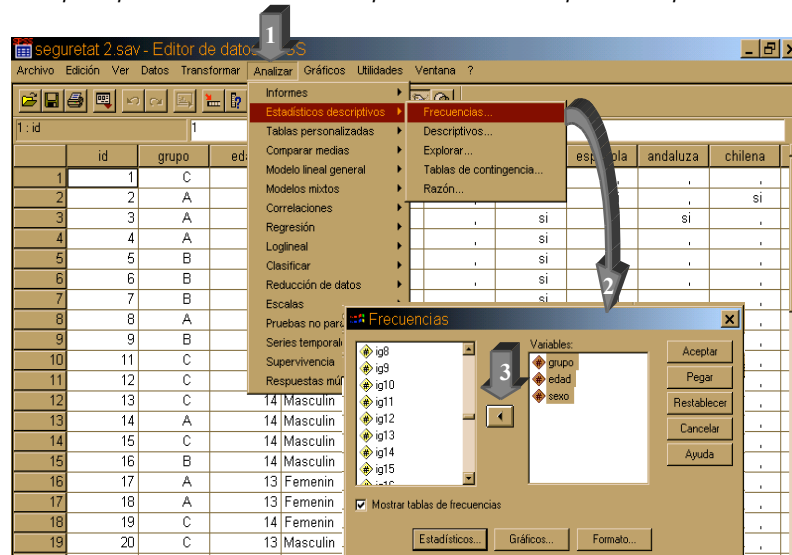
- El **nombre de la variable** lo utilizaremos para visualizarlo en la matriz y, por tanto, nos interesa un nombre corto (máximo de 8 dígitos) que nos identifique cada una de las variables.
- El **tipo de variable** nos permite distinguir si la variable es numérica o alfanumérica.
- En la **amplitud** de la variable podemos definir el número de dígitos que contempla esta variable en la matriz de datos.
- Los **decimales** suponen definir el número de decimales que contiene la variable.
- En la **etiqueta de la variable** podemos proporcionar un nombre más largo y específico que identifique el contenido de la variable que generemos. Esta etiqueta es la que aparecerá en las salidas de resultados. En caso de no definirla se utiliza por defecto el *nombre de la variable*.
- Los **valores** de la variable se utilizan por aquellas variables cualitativas que disponen de categorías. Por ejemplo, el sexo tiene las categorías "femenino" y "masculino". En este caso, hacemos clic en la opción de valores y en el cuadro de diálogo que surja podemos especificar los *valores numéricos* y su correspondiente *etiqueta de valor*. Después de cada una de las especificaciones es necesario darle a *agregar*.
- Los valores **perdidos** son aquellos valores que no se tienen en cuenta para el análisis de los resultados. El SPSS identifica dos tipos: los perdidos por el sistema (cualquier casilla en blanco dentro de la matriz de datos) y los perdidos por el usuario. En este último caso es el usuario el que define cuáles son estos valores. En el cuadro de diálogo que aparece podemos escoger entre diferentes opciones: podemos definir valores perdidos que no sean consecutivos (valores perdidos discretos), o bien podemos definir un rango de valores perdidos más un valor discreto opcional. Por ejemplo si en el sexo tenemos que 1 es femenino y 2 es masculino, podemos definir un rango de 3 a 999999 y un valor discreto 0, como valores perdidos.
- La opción **columna** nos permite definir o modificar la amplitud que ocupa la columna para cada variable dentro de la vista de datos y no afecta en ningún caso la amplitud definida previamente.
- La **alineación** hace referencia a cómo se sitúan los datos en la columna correspondiente dentro de la vista de datos (a la derecha, izquierda o centradas).
- Por último, debemos especificar el tipo de **escala de medida** que utilizaremos para cada variable. Aparece un menú en el que podemos escoger la escala correspondiente: Escala, Ordinal o Nominal.

Una vez que hemos definido todas las características de las variables del estudio podemos comenzar a introducir los datos dentro del editor de datos.

Una vez elaborada la matriz de datos, pero de forma previa al análisis propiamente dicho, es más que recomendable hacer una primera revisión de los posibles errores que contenga la matriz y que puedan alterar los resultados del estudio que se pretende hacer. Para ello, disponemos de los procesos de depuración de la matriz de datos. Estos procedimientos consisten en corregir dos posibles fuentes de error: errores en la corrección del instrumento de medida o bien errores en la transcripción de los valores en la matriz de datos.



Fig. 1 Pasos para pedir estadísticos descriptivos en el SPSS para la depuración de la matriz



Depurar la matriz de datos consiste básicamente en calcular algunos estadísticos descriptivos de todas las variables (en el SPSS se pueden pedir desde la opción del menú *analizar/estadísticos descriptivos/frecuencias*), como por ejemplo:

- *Las frecuencias absolutas* de cada una de las categorías o valores que toma la variable para verificar que no hay ningún valor ajeno a los valores lógicos de la variable. Por ejemplo con la variable *sexo*, podemos tener las dos categorías *femenino* y *masculino*. Si además encontramos un valor numérico, podemos deducir que se trata de un error.
- *El número de individuos* de cada variable para verificar que sea correcto y que no se ha olvidado ni duplicado ningún individuo.
- *El mínimo y el máximo* (en la opción de *estadísticos* del cuadro de diálogo) para verificar que no existan valores extremos que sobrepasen o no lleguen al rango lógico de la variable. Si tenemos la variable *edad* y la muestra es alumnado de primaria, se trataría de verificar que no exista ningún valor inferior a 6 años o superior a 12, dado que seguramente sería un error.

La depuración de la matriz es básica, ya que si no detectamos y corregimos estos posibles errores de transcripción o corrección podrían producirse graves errores en los resultados finales.

Por otro lado, algunos de estos procesos de depuración de la matriz se inician de forma previa a la introducción de datos, con la misma definición de las variables cuando se consideran los valores perdidos de cada una. Los valores perdidos del/la usuario/a se especifican para cada variable por parte de la persona investigadora. Considerando el rango posible de valores para las cuantitativas continuas, se puede definir un intervalo de valores que se encuentren fuera de los límites posibles de cada variable. Aquellos valores que manifestamos que son perdidos, no se consideran para el análisis de datos, ya que podrían alterarlo.



El análisis descriptivo

Después de la depuración de la matriz de datos es frecuente dar prioridad a un análisis exploratorio EDA (*Exploratory Data Analysis*) basado en el análisis descriptivo de cada variable, que permite una mayor comprensión del fenómeno. Por este motivo, tiene una especial importancia la detección de *outliers* o sujetos con puntuaciones que se distancian notablemente de los modelos establecidos para el resto de individuos de la muestra.

La estadística descriptiva univariable da cuerpo al análisis exploratorio de los datos, describiendo las características de un hecho, objeto o fenómeno. Este análisis se puede efectuar desde una perspectiva más visual mediante tablas y representaciones gráficas; y desde una perspectiva más analítica a través de los indicadores estadísticos.

Tablas y representaciones gráficas

El SPSS proporciona tablas de frecuencias de las variables, que constan de los siguientes elementos: valores válidos o las categorías, su frecuencia absoluta (*frecuencia*), la frecuencia percentual (*porcentaje*), la frecuencia percentual sin contabilizar los valores perdidos (*porcentaje válido*) y la frecuencia percentual acumulada (*porcentaje acumulado*) basada en el *porcentaje válido*. Para pedir esta tabla desde el SPSS se hace mediante la opción *analizar / estadísticos descriptivos / frecuencias*.

Además de las tablas de frecuencias también podemos hacer uso de las tablas de contingencia y otras tablas donde cruzar más de una variable. Una tabla de contingencia es una tabla donde situar las frecuencias en el interior, en función de dos variables con dos o más categorías cada una. Por ejemplo, en la tabla 1, se presenta una tabla de contingencia sobre las frecuencias observadas de las diferentes categorías profesionales, en función del sexo.

Tabla 1 Ejemplo de tabla de contingencia obtenida con el SPSS

		Sexo del encuestado		Total
		Hombre	Mujer	
Categoría ocupacional	Direcctivo o profesional liberal	160	179	339
	Empleado técnico, administrativo o comercial	122	334	456
	Servicios	57	147	204
	Agricultura, forestal y pesca	26	10	36
	Producción de precisión, manufactura o reparación	136	27	163
	Operario, fabricación y mano de obra en general	120	100	220
Total		621	797	1418

El SPSS nos ofrece crear tablas de contingencia desde la opción *analizar / estadísticos descriptivos / tablas de contingencia*; donde inclusive tenemos la opción de pedir gráficos agrupados en función de las variables que se representan en la tabla.

También se pueden representar estadísticos en el interior de una tabla, en función de diferentes variables categóricas. Imaginemos que nos gustaría hacer una tabla donde representar el promedio (*media aritmética*) del número de hijos que tienen las personas en función de su categoría profesional. Este ejemplo es el que presentamos en la tabla 2.



Tabla 2 Ejemplo de tabla donde se resumen las medias aritméticas de una variable en función de otra

Categoría ocupacional	Direcctivo o profesional liberal	Nº de hijas e hijos	3
	Empleado técnico,	Nº de hijas e hijos	3
	Servicios	Nº de hijas e hijos	5
	Agricultura, forestal y pesca	Nº de hijas e hijos	4
	Producción de precisión,	Nº de hijas e hijos	4
	Operario, fabricación y mano	Nº de hijas e hijos	5

El SPSS nos ofrece crear este tipo de tablas, a través de *analizar/tablas/tablas básicas*. No obstante, también se pueden crear a través de la opción de *tablas generales*, el proceso es muy similar. En ambos casos, cuando se pide la variable a *resumir* se trata de ofrecer la variable cuantitativa de la que deseamos obtener la media aritmética, en el ejemplo anterior, correspondería a la variable *nº de hijos/as*. Por otro lado, los subgrupos los haríamos gracias a variables habitualmente cualitativas como en el ejemplo, *la categoría profesional*.

Las técnicas de representación gráfica son otro instrumento para reducir la información original. La pérdida de información es mayor, pero presenta la ventaja de establecer diferencias entre dos distribuciones a simple vista. Para la representación gráfica es necesario diferenciar si la variable es cualitativa (fruto de la observación como el sexo o el estado civil) o cuantitativa (fruto de una medida como la edad o el peso). Así, para las variables cualitativas son adecuados los gráficos siguientes: *Ciclograma o de sectores o el diagrama de barras*. Las variables cuantitativas, en cambio, se representan gráficamente, entre otros, con el *polígono de frecuencias o el histograma*. Todos ellos se pueden obtener en el SPSS desde la opción del menú *gráficos*.

Indicadores estadísticos

Las representaciones gráficas ofrecen al investigador/a un método elemental para el análisis de las propiedades del objeto de estudio; se necesitan otros métodos más precisos para el estudio de las características más importantes de las distribuciones de frecuencias. Para ello disponemos de los indicadores estadísticos, que nos permitirán dar respuesta a cuestiones como las siguientes: ¿cómo están los valores ordenados? (*indicadores de posición*), ¿cómo se concentran los valores? (*indicadores de tendencia central*), ¿cómo se dispersan o qué variabilidad hay? (*indicadores de dispersión*) y ¿qué forma adopta la distribución? (*indicadores de forma*). Básicamente, con uno solo de estos indicadores no podemos llegar a describir a la perfección toda una serie numérica, sino que como mínimo, a la hora de sintetizar una distribución, debemos hacer uso de todos estos diferentes tipos de indicadores estadísticos.

Desde el SPSS se puede hacer este análisis en la opción *Frecuencias*¹ para los siguientes estadísticos descriptivos:

- Indicadores de tendencia central: media aritmética, error típico de la media, moda² y mediana.

¹ La opción *Frecuencias* se obtiene desde la opción *Analizar* del menú principal donde se ejecutan todos los análisis estadísticos, en la opción de *Estadísticos descriptivos*.



- Indicadores de dispersión: varianza, desviación típica, rango, mínimo y máximo.
- Indicadores de posición: cuartiles, deciles y percentiles³.
- Indicadores de forma: asimetría y su error típico, curtosis y su error típico.

Hay diversas maneras de obtener estos cálculos desde el SPSS, una de las más sencillas es desde la opción *analizar/estadísticos descriptivos/frecuencias*, siguiendo los pasos que se detallan en la figura 2.

Fig.2 Pasos para calcular indicadores estadísticos desde el SPSS

The figure shows the SPSS interface with the following components:

- Step 1:** The 'Análisis' menu is open, and 'Frecuencias...' is selected.
- Step 2:** The 'Frecuencias' dialog box is open, showing selected variables: 'Sexo del encuestado', 'Región de los Estados Unidos', 'Nivel de felicidad (felicidad)', '¿Su vida es emocionante?', 'Número de hermanos', 'Número de hijos (hijos vivos)', 'Número de años de escolaridad', and 'Número de años de experiencia profesional'. The 'Mostrar tablas de frecuencias' checkbox is checked.
- Step 3:** The 'Frecuencias: Estadísticos' sub-dialog box is open, showing the following options:
 - Valores percentiles:** Cuartiles, Puntos de corte para 10 grupos iguales, Percentiles: 68. The 'Añadir' button is highlighted.
 - Tendencia central:** Media, Mediana, Moda, Suma.
 - Dispersión:** Desv. típica, Varianza, Amplitud, Mínimo, Máximo, E.T. media.
 - Distribución:** Asimetría, Curtosis.

² El SPSS nos da un único valor modal, tanto si se trata de distribuciones unimodales, bimodales, como multimodales. Concretamente, cuando estamos delante de distribuciones con más de una moda, el output ofrece el valor modal más pequeño.

³ El output siempre da percentiles, aunque tiene la opción de pedir cualquiera de los indicadores deciles o cuartiles, ya que su equivalencia es sencilla.



Para profundizar en estos elementos, se recomienda consultar algunos manuales de estadística básica que explican con mayor detalle aspectos de estadística descriptiva (Por ejemplo: Visauta i Batallé, 1991).

Documentación

Alvarez, M. (2000). *Análisis estadístico con SPSS: procedimientos básicos*. Bilbao: Deusto.

Bisquerra, R. (1987). *Introducción a la estadística aplicada a la investigación educativa. un enfoque informático con los paquetes BMDP y SPSS-X*. Barcelona: PPU.

Boadas, J. (1995). *Fundamentos de estadística*. Barcelona: UPC.

Bryman, A., y Cramer, D. (1995). *Quantitative data analysis for social scientist*. London: Routledge.

Calvo, F. (1994). *Estadística aplicada*. Bilbao: Deusto.

Camacho, J. (2000). *Estadística con SPSS*. Madrid: Ra-Ma.

Cuadras, C., Echeverría, B., Mateo, J., y Sánchez, P. (1996). *Fundamentos de estadística. Aplicación a las ciencias humanas*. Barcelona: PPU.

Domènech, J. M., y Granero, R. (2002a). *Anàlisi de dades per a recerca en psicologia. Fonaments*. Barcelona: Signo.

Domènech, J. M., y Granero, R. (2002b). *Anàlisi de dades per a recerca en psicologia. Models estadístics bàsics (Vol. 2)*. Barcelona: Signo.

Filgueira, E. (2001). *Análisis de datos con SPSSWIN*. Madrid: Alianza editorial.

Gil, J. (2003). La estadística en la investigación educativa. *Revista de Investigación Educativa*, 21(1), 231-248.

Gondar, J. E. (2002). *Técnicas estadísticas con SPSS*. Madrid: Data mining institute.

Hartwig, F., y Dearing, B. E. (1987). *Exploratory data analysis*. Beverly Hills: Sage.

Hoaglin, D., Mosteller, F., y Tukey, J. W. (1991). *Fundamentals of exploratory analysis of variance*. New York: Wiley.

Hopkins, K. D., Hopkins, B. R., y Colass, G. V. (1997). *Estadística básica para las ciencias sociales y del comportamiento*. México: Prentice Hall.

Lizasoain, L., y Joaristi, L. (1998). *SPSS para windows*. Madrid: Paraninfo.

Lomax, R. G. (2001a). *An introduction to statistical concepts for education and behavioral sciences*. New Jersey: LEA.



Lomax, R. G. (2001b). *Statistical concepts. A second course for education and the behavioral sciences*. New Jersey: LEA.

Norusis, M. J. (2000). *SPSS 10.0: guide to data analysis*. New Jersey: Prentice Hall.

Pardo, A., y Ruiz, M. A. (2002). *SPSS 11. Guía para el análisis de datos*. Madrid: McGraw-Hill.

Spiegel, M. (1991). *Estadística*. Madrid: McGrawHill.

Visauta, B. (1997). *Análisis estadístico con SPSS para windows*. (Vol. 1. Estadística básica). Madrid: McGraw-Hill.

Visauta, B., y Batallé, P. (1991). *Métodos estadísticos aplicados. Estadística descriptiva*. Barcelona: PPU.

Vilà, R., y Bisquerra, R. (2004). El análisis cuantitativo de los datos. En R. Bisquerra (Ed.), *Metodología de la investigación educativa*. Madrid: La Muralla.