



Fitxes per a investigadors

Com fer una anàlisi quantitativa de dades de tipus descriptiu amb el paquet estadístic de l'SPSS?

Autora:

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

L'anàlisi quantitativa de dades mitjançant la informàtica estadística

En l'actualitat és difícil parlar de processos estadístics d'anàlisi quantitativa de dades sense fer referència a la informàtica aplicada a la recerca. Aquests recursos informàtics es basen sovint en paquets de programes informàtics que tenen l'objectiu d'ajudar al/a la investigador/a en la fase d'anàlisi de dades. En aquests moments un dels paquets més perfeccionats i complets és l'SPSS (Statistical Package for the Social Sciences).

L'SPSS és un paquet de programes per tal de dur a terme l'anàlisi estadística de les dades. Constitueix una aplicació estadística força potent, de la qual s'han anat desenvolupant diverses versions des dels seus inicis, als anys setanta. En aquest manual les sortides d'ordinador que es presenten pertanyen a la versió 11.0.1. No obstant això, tot i que la forma ha anat variant des dels inicis, pel que fa al funcionament segueix essent molt similar entre les diferents versions.

Abans d'iniciar-nos en la utilització de les aplicacions de l'SPSS és important familiaritzar-se amb algunes de les finestres que més farem servir. En entrar a l'SPSS el primer que ens trobem és l'editor de dades. Aquesta finestra visualitza, bàsicament, les dades que anirem introduint. L'editor de dades inclou dues opcions: la *Vista de les dades* i la de les *variables*. Aquestes opcions es poden seleccionar a partir de les dues pestanyes que es presenten en la part inferior. La *vista de dades* conté el menú general i la matriu de dades. Aquesta matriu s'estructura amb els casos a les files i les variables a les columnes.

La matriu de dades

En una recerca i una vegada finalitzat el procés de recollida de dades, pren protagonisme la creació de la matriu de dades (que es troba en el *Visor d'edició de dades* de l'SPSS). Una matriu de dades és una estructura en forma de taula que conté els valors de cada subjecte en les diferents variables. Habitualment es col·loquen les variables en columnes i els individus en fileres, tot i que no ha de coincidir obligatòriament el nombre de variables amb el de columnes, o el nombre de files amb el nombre d'individus.

Crear la matriu de dades implica definir les variables que es consideren per a l'estudi i introduir les dades recollides. La vista de variables ens permet fer especificacions sobre les variables que conté la nostra matriu de dades. En aquesta finestra les diferents



variables es troben situades a les files, mentre que les característiques per definir d'aquestes variables es troben a les columnes. Concretament, podem definir els següents aspectes de cada una de les variables de l'estudi:

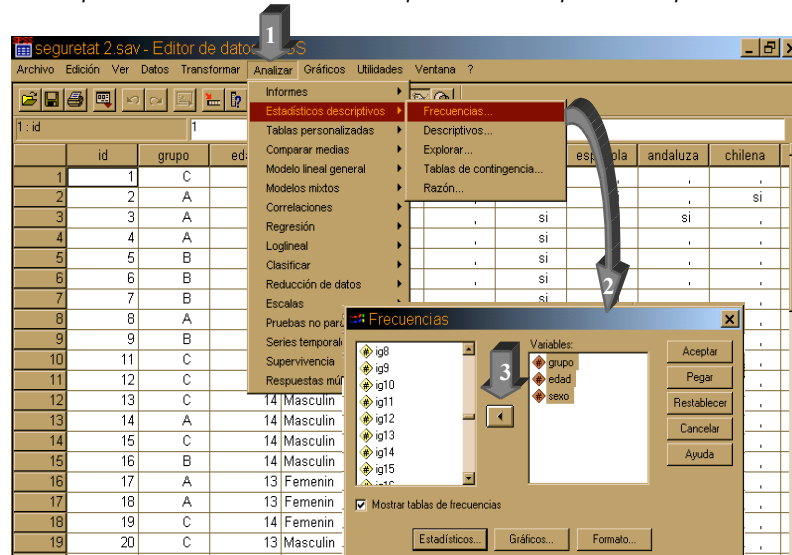
- El **nom** de la variable, que utilitzarem per visualitzar-lo en la matriu i, per tant, ens interessa un nom curt (màxim de 8 dígits) que ens identifiqui cada una de les variables.
- El **tipus** de variable ens permet distingir si la variable és numèrica o alfanumèrica.
- A l'**amplada** de la variable podem definir el nombre de dígits que contemplen aquesta variable en la matriu de dades.
- Els **decimals** suposen definir el nombre de decimals que conté la variable.
- A l'**etiqueta** de la variable podem proporcionar un nom més llarg i específic que identifiqui el contingut de la variable que generem. Aquesta etiqueta és la que apareixerà a les sortides de resultats. En cas que no es defineixi s'utilitza per defecte el nom de la variable.
- Els **valors** de la variable s'utilitzen per aquelles variables qualitatives que disposen de categories. Per exemple, el sexe té les categories «femení» i «masculí». En aquest cas, fem clic a l'opció de valors i al quadre de diàleg que sorgeix podem especificar els «valors numèrics» i la corresponent «etiqueta de valor». Després de cada una de les especificacions és necessari prémer «afegir».
- Els valors **perduts** són aquells valors que no es tenen en compte per a l'anàlisi dels resultats. L'SPSS n'identifica dos tipus: els perduts pel sistema (qualsevol casella en blanc dins de la matriu de dades) i els perduts per l'usuari. En aquest darrer cas és l'usuari el que defineix quins són aquests valors. Al quadre de diàleg que apareix podem escollir entre diferents opcions: podem definir valors perduts que no siguin consecutius (valors perduts discrets), o bé podem definir un rang de valors perduts més un valor discret opcional. Per exemple si en el sexe tenim que 1 és femení i 2 és masculí, podem definir un rang de 3 a 999999 i un valor discret 0, com a valors perduts.
- L'opció **columna** ens permet definir o modificar l'amplada que ocupa la columna per a cada variable dins la vista de dades i no afecta en cap cas l'amplada definida prèviament.
- L'**alineació** fa referència a com se situen les dades a la columna corresponent dins la vista de dades (a la dreta, a l'esquerra o centrades).
- Per últim, hem d'especificar el tipus d'**escala de mesura** que farem servir per a cada variable. Apareix un menú on podem escollir l'escala corresponent: Escala, Ordinal o Nominal.
-

Un cop hem definit totes les característiques de les variables de l'estudi podem començar a introduir les dades dins de l'editor de dades.

Una vegada elaborada la matriu de dades però de forma prèvia a l'anàlisi pròpiament, és més que recomanable fer una primera revisió de les possibles errades que contingui la matriu i que poden alterar els resultats de l'estudi que es planteja fer. Per això, disposem dels processos de depuració de la matriu de dades. Aquests procediments consisteixen a corregir dos possibles fonts d'error: errades en la correcció de l'instrument de mesura o bé errades en la transcripció dels valors a la matriu de dades.



Fig. 1 Passos per demanar estadístics descriptius a l'SPSS per a la depuració de la matriu



Depurar la matriu de dades consisteix bàsicament a calcular alguns estadístics descriptius de totes les variables (a l'SPSS es poden demanar des de l'opció del menú *Análisis*>*Estadístics descriptius*>*Freqüències*), com per exemple:

- Les *freqüències absolutes* de cada una de les categories o valors que pren la variable per verificar que no hi ha cap valor aliè als valors lògics de la variable. Per exemple amb la variable «sexe», podem tenir les dues categories, femení i masculí. Si a més a més trobem un valor numèric, podem deduir que es tracta d'una errada.
- El *nombre d'individus* de cada variable per verificar que sigui correcte i que no s'ha oblidat ni duplicat cap individu.
- El *mínim* i el *màxim* (en l'opció d'*estadístics* del quadre de diàleg) per verificar que no existeixin valors extrems que sobrepassin o que no arribin al rang lògic de la variable. Si tenim la variable «edat» i la mostra és «alumnat de primària», es tractaria de verificar que no existeixi cap valor inferior a 6 anys o superior a 12, ja que segurament seria una errada.

La depuració de la matriu és bàsica, ja que si no detectem i corregim aquestes possibles errades de transcripció o correcció es podrien produir greus errors en els resultats finals.

Per altra banda, alguns d'aquests processos de depuració de la matriu s'inicien prèviament a la introducció de dades, amb la mateixa definició de les variables quan es consideren els valors perduts de cadascuna. L'investigador o investigadora especifica per a cada variable els valors perduts del/de la usuari/a. Considerant el rang possible de valors per les quantitatives contínues, es pot definir un interval de valors que es troben fora dels límits possibles de cada variable. Els valors que manifestem que són perduts, no es consideren per a l'anàlisi de dades, ja que podrien alterar-la.



L'anàlisi descriptiva

Després de la depuració de la matriu de dades és freqüent donar prioritat a una anàlisi exploratòria EDA (*Exploratory Data Analysis*) basada en l'anàlisi descriptiva de cada variable, que permet una major comprensió del fenomen. Per aquest motiu, té una especial importància la detecció d'*outliers* o subjectes amb puntuacions que es distancien notablement dels models establerts per la resta d'individus de la mostra.

L'estadística descriptiva univariable dóna cos a l'anàlisi exploratòria de les dades, tot descrivint les característiques d'un fet, objecte o fenomen. Aquesta anàlisi es pot efectuar des d'una perspectiva més visual mitjançant taules i representacions gràfiques; i des d'una perspectiva més analítica a través dels indicadors estadístics.

Taules i representacions gràfiques

L'SPSS proporciona taules de freqüències de les variables, que consten dels següents elements: els valors vàlids o les categories, la seva freqüència absoluta (*freqüència*), la freqüència percentual (*percentatge*), la freqüència percentual sense comptabilitzar els valors perduts (*percentatge vàlid*) i la freqüència percentual acumulada (*percentatge acumulat*) basada en el *percentatge vàlid*. Per demanar aquesta taula des de l'SPSS es fa mitjançant l'opció *Analitzar>Estadístics descriptius>Freqüències*.

A més de les taules de freqüències també podem fer ús de les taules de contingència i altres taules on es pugui creuar més d'una variable. Una taula de contingència és una taula on situar les freqüències a l'interior, en funció de dues variables amb dues o més categories cada una. Per exemple, a la taula 1 es presenta una taula de contingència sobre les freqüències observades de les diferents categories professionals, en funció del sexe.

Taula 1 Exemple de taula de contingència obtinguda amb l'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

L'SPSS ens ofereix crear taules de contingència des de l'opció *Analitzar>Estadístics descriptius>Taules de contingència*, on fins i tot tenim l'opció de demanar gràfics agrupats en funció de les variables que es representen a la taula.

També es poden representar estadístics a l'interior d'una taula, en funció de diferents variables categòriques. Imaginem que ens agradaria fer una taula on representar la mitjana (*mitjana aritmètica*) del nombre de fills que tenen les persones en funció de la seva categoria professional. Aquest exemple és el que presentem a la taula 2.



Taula 2 Exemple de taula on es resumeixen les mitjanes aritmètiques d'una variable en funció de l'altra

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

L'SPSS ens ofereix crear aquest tipus de taules, a través d'*Analitzar>Taules>Taules bàsiques*. No obstant això, també es poden crear a través de l'opció de *Taules generals*, el procés és molt similar. En ambdós casos, quan es demana la variable a *Resumir* es tracta d'oferir la variable quantitativa de la qual voldrem fer la mitjana aritmètica, en l'exemple anterior correspondria a la variable «núm. de fills o filles». Per altra banda, els subgrups els faríem gràcies a variables habitualment qualitatives com en l'exemple, la «categoria professional».

Les tècniques de representació gràfica són un altre instrument per reduir la informació original. La pèrdua d'informació és major, però presenta l'avantatge d'establir diferències entre dues distribucions a simple vista. Per a la representació gràfica és necessari diferenciar si la variable és qualitativa (fruit de l'observació com el sexe o l'estat civil) o quantitativa (fruit d'una mesura com l'edat o el pes). Així, per a les variables qualitatives són adequats els gràfics següents: *Ciclograma o de sectors o el diagrama de barres*. Les variables quantitatives en canvi, es representen gràficament, entre altres, amb el *Polígon de freqüències o l'histograma*. Tots es poden aconseguir a l'SPSS des de l'opció del menú *Gràfics*.

Indicadors estadístics

Les representacions gràfiques ofereixen a l'investigador/a un mètode elemental per a l'anàlisi de les propietats de l'objecte d'estudi; es necessiten altres mètodes més precisos per a l'estudi de les característiques més importants de les distribucions de freqüències. Per això disposem dels indicadors estadístics, que ens permetran donar resposta a qüestions com les següents: com es troben els valors ordenats (*indicadors de posició*)? Com es concentren els valors (*indicadors de tendència central*)? Com es dispersen o quina variabilitat hi ha (*indicadors de dispersió*)? I, quina forma adopta la distribució (*indicadors de forma*)? Bàsicament, només amb un d'aquests indicadors no podem arribar a descriure a la perfecció tota una sèrie numèrica, sinó que com a mínim, a l'hora de sintetitzar una distribució, haurem de fer ús de tots aquests diferents tipus d'indicadors estadístics.

Des de l'SPSS es pot fer aquesta anàlisi a l'opció *Freqüències*¹ pels següents estadístics descriptius:

- Indicadors de tendència central: mitjana aritmètica, error típic de la mitjana, moda² i mediana.

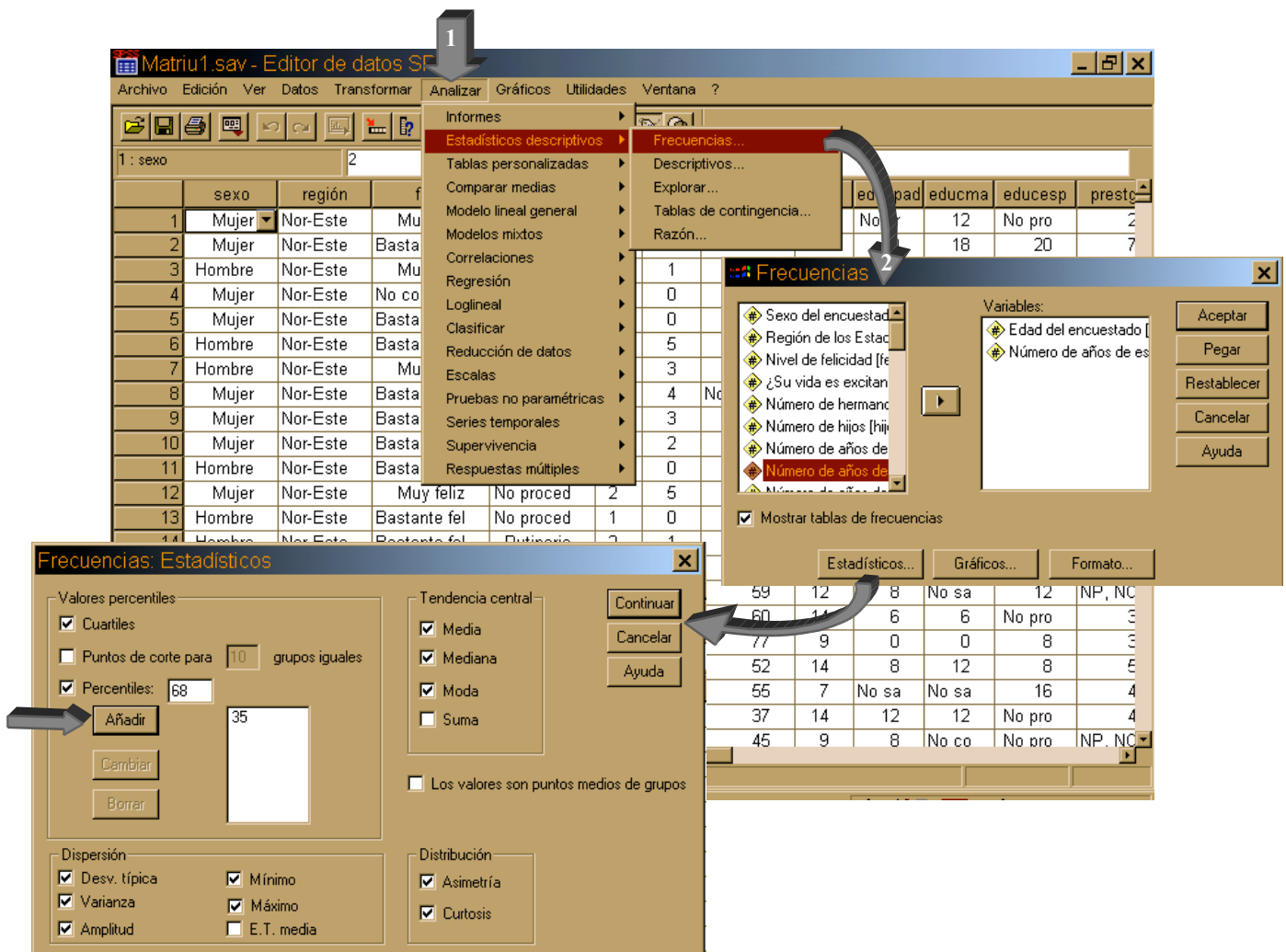
¹ L'opció *Freqüències* s'obté des de l'opció *Analitzar* del menú principal on s'executen totes les anàlisis estadístiques, en l'opció d'*Estadístics descriptius*.



- Indicadors de dispersió: variança, desviació típica, rang, mínim i màxim.
- Indicadors de posició: quartils, decils i percentils.³
- Indicadors de forma: asimetria i el seu error típic; curtosi i el seu error típic.

Hi ha diverses maneres d'obtenir aquests càlculs des de l'SPSS, una de les més senzilles és des de l'opció *Analtzar > Estadístics descriptius > Freqüències*, seguint els passos que es detallen a la figura 2.

Fig.2 Passos per calcular indicadors estadístics des de l'SPSS



Per aprofundir en aquests elements, es recomana que consulteu alguns manuals d'estadística bàsics que expliquen amb major detall aspectes d'estadística descriptiva (Per exemple: Visauta i Batallé, 1991).

² L'SPSS ens dona un únic valor modal, tant si es tracta de distribucions unimodals, bimodals, com multimodals. Concretament, quan estem davant de distribucions amb més d'una moda, l'*output* ofereix el valor modal més petit.

³ L'*output* sempre dona percentils, tot i que té l'opció de demanar qualsevol dels indicadors, decils o quartils, ja que l'equivalència és senzilla.



Documentació

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., i 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., i Sánchez, P. (1996). *Fundamentos de estadística. Aplicación a las ciencias humanas*. Barcelona: PPU.

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

Domènech, J. M., i 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., i Dearing, B. E. (1987). *Exploratory data analysis*. Beverly Hills: Sage.

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

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

Lizasoain, L., i 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., i 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., i Batallé, P. (1991). *Métodos estadísticos aplicados. Estadística descriptiva*. Barcelona: PPU.

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