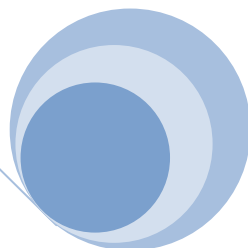




UNIVERSITAT DE BARCELONA



Estadística descriptiva aplicada a l'educació

**Ormazábal, F.J. (Coord.) Vila, R.; Mateo, M.; Torrado, M.;
Berlanga, V.; del Barrio, J.; Ruiz, A.**

*Departament de Mètodes d'Investigació i Diagnòstic en
Educació (MIDE)
Facultat de Pedagogia
Universitat de Barcelona*



Aquesta publicació compta amb la següent llicència de Creative Commons:



Estadística descriptiva aplicada a l'educació està subjecte a una llicència de Reconeixement-NoComercial-SinObraDerivada 3.0 (CC BY-NC-ND 3.0)

Para citar la obra:

Ormazábal, F.J. (coord.) Vila, R.; Mateo, M.; Torrado, M.; Berlanga, V.; Del Barrio, J.; Ruiz, A. (2012) *Estadística descriptiva aplicada a l'educació*. Barcelona: Universitat de Barcelona. Dipòsit Digital <http://hdl.handle.net/2445/21387>

ÍNDEX

Estadística descriptiva	
1.1 Tabulació i representació gràfica	4
1.1.1 La tabulació	4
1.1.2 Representació gràfica	11
1.2 Indicadors estadístics	19
1.2.1 Indicadors de posició	19
1.2.2 Indicadors de tendència central	21
1.2.3 Indicadors de dispersió	22
1.2.4 Indicadors de forma	25
1.2.5 Els indicadors estadístics al SPSS	28
1.2.6 Com descriure una variable utilitzant els indicadors estadístics	30

1.1 Tabulació i representació gràfica

1.1.1 La tabulació

La tabulació és el procediment d'organitzar les dades en una taula. Forma part d'una tècnica d'ordenació de dades. Concretament, les taules de freqüències consten dels següents aspectes, que poden variar en funció del tipus de variable:

- **Freqüència absoluta:**

- **Individual (fi):** és el nombre de vegades que apareix una determinada dada.
- **Acumulada (fa):** és la suma de les freqüències absolutes individuals de les dades anteriors més l'actual.

$$fa_j = \sum_{k=1}^j fi_k$$

Per exemple: $fa_4 = fi_1 + fi_2 + fi_3 + fi_4$

L'última freqüència absoluta acumulada coincideix amb el total de la mostra (n)

- **Freqüència relativa:**

En proporció:

- **Individual (pi):** és el nombre de vegades que apareix una dada (fi) en relació al total de la mostra (n). És el tant per u.

$$pi = \frac{fi}{n}$$

- **Acumulada (pa):** és la freqüència absoluta acumulada (fa) dividida per la grandària de la mostra (n).

$$pa = \frac{fa}{n}$$

O bé podem dir que és la suma de les freqüències relatives individuals de les dades anteriors més l'actual.

$$pa_j = \sum_{k=1}^j pi_k$$

Per exemple: $pa_4 = pi_1 + pi_2 + pi_3 + pi_4$

L'última freqüència relativa en proporció acumulada val sempre 1

En percentatge:

- **Individual (Pi):** és la freqüència relativa (pi) en base a 100 (percentatge).

$$P_i = \frac{f_i}{n} \cdot 100 = p_i \cdot 100$$

- **Acumulada (Pa):** és la suma de les freqüències relatives individuals de les dades anteriors, més l'actual.

$$P_{a_j} = \sum_{k=1}^j P_{i_k}$$

Per exemple: $P_{a_4} = P_{i_1} + P_{i_2} + P_{i_3} + P_{i_4}$

L'última freqüència relativa en percentatge acumulada val sempre 100

Per exemple, si recollim les següents dades sobre el nombre d'amics d'altres cultures, podem organitzar la taula de la següent forma:

Dades: 2 3 3 3 3 4 4 4 5 6 (n = 10)

Tabulació:

X	fi	fa	pi	pa	Pi	Pa
2	1	1	0'10	0'10	10	10
3	4	5	0'40	0'50	40	50
4	3	8	0'30	0'80	30	80
5	1	9	0'10	0'90	10	90
6	1	10	0'10	1	10	100

Taula 1. Exemple de taula de freqüències.

- **Taules de freqüència amb SPSS:**



L'SPSS proporciona taules de freqüències de les variables, que consten dels següents aspectes:

- Els valors vàlids o les categories
- La seva freqüència absoluta individual (*frecuencia*)
- La freqüència relativa individual en percentatge (*porcentaje*)
- La freqüència relativa individual en percentatge sense comptabilitzar els valors perduts (*porcentaje válido*)
- La freqüència relativa acumulada en percentatge (*porcentaje acumulado*) basada en el percentatge vàlid.

A la taula 2 presentem la taula obtinguda amb SPSS sobre la variable nombre d'amics. Per demanar aquesta taula des del SPSS es fa mitjançant l'opció *Analizar / Estadísticos descriptivos / Frecuencias*¹.

Nombre d'amics					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	2,000	1	10,0	10,0	10,0
	3,000	4	40,0	40,0	50,0
	4,000	3	30,0	30,0	80,0
	5,000	1	10,0	10,0	90,0
	6,000	1	10,0	10,0	100,0
	Total	10	100,0	100,0	

Taula 2. Exemple de taula de freqüències obtinguda amb l'SPSS.

A la taula 3 es presenten com queden les freqüències de percentatge vàlid i percentatge acumulat en el cas d'una variable (nombre de fills) quan hi ha valors perduts.

Nombre de fills					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0	419	27,6	27,8	27,8
	1	255	16,8	16,9	44,7
	2	375	24,7	24,9	69,5
	3	215	14,2	14,2	83,8
	4	127	8,4	8,4	92,2
	5	54	3,6	3,6	95,8
	6	24	1,6	1,6	97,3
	7	23	1,5	1,5	98,9
	8 o més	17	1,1	1,1	100,0
	Total	1509	99,5	100,0	
Perdidos	No contesta	8	,5		
Total		1517	100,0		

Taula 3. Exemple de taula de freqüències obtinguda amb l'SPSS.

¹ Aquest procediment es veu amb més detall en la figura 2.3 en aquest mateix capítol.

- **Altres taules**

A més a més de les taules de freqüències també podem fer ús de les taules de contingència i altres taules on 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, en la taula 2.4 es presenta una taula de contingència sobre les freqüències observades de les diferents categories professionals en funció del sexe. En aquest cas, es tracta d'una taula de 6 x 2, donat que en la variable *categoria professional* hi ha sis categories (*directiu, empleat, serveis, agricultura, producció i operari*) i la variable sexe té dues categories (*home i dona*).

Tabla de contingencia Categoría ocupacional * Sexo de l'enquestat

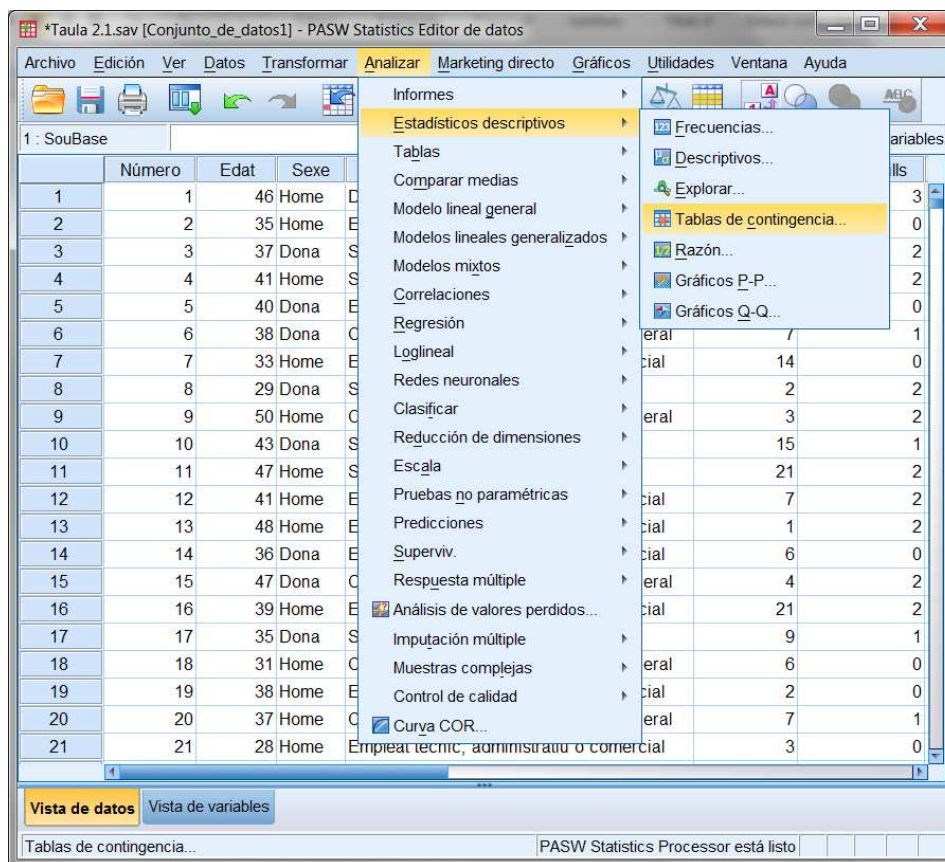
Recuento

		Sexe de l'enquestat		Total
		Home	Dona	
Categoría ocupacional	Directiu o professional liberal	160	179	339
	Empleat tècnic, administratiu o comercial	122	334	456
	Serveis	57	147	204
	Agricultura, forestal i pesca	26	10	36
	Producció de precisió, manufactura o reparació	136	27	163
	Operari, fabricació i mà d'obra en general	120	100	220
Total		621	797	1418

Taula 4. Exemple de taula de freqüències obtinguda amb l'SPSS.



L'SPSS ens ofereix crear taules de contingència. En la figura 1 podeu veure el procés pel qual es demana una taula de contingència a l'SPSS.



I llavors:

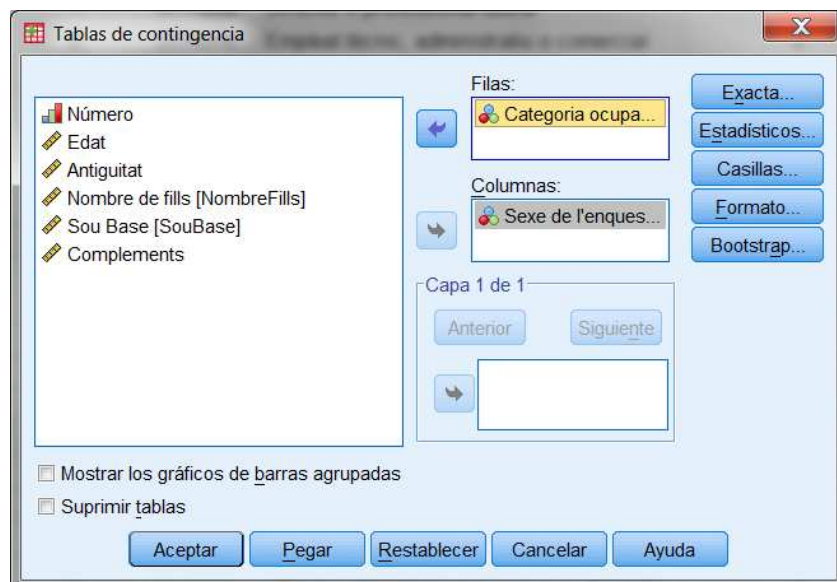


Figura 1. Les taules de contingència des de l'SPSS.

També podem representar estadístics a l'interior d'una taula, en funció de diferents variables categòriques. Imaginem que ens agradaria fer una taula on representar el valor mig (anomenat mitjana aritmètica²) del nombre de fills que tenen les persones en funció de la seva categoria professional. És a dir, quina mitjana de fills tenen les i els professionals de la direcció, de l'administració, del sector serveis, etc. Aquest tipus de taula és el que presentem a la taula 5.

		Nombre de filles i fills
		Media
Categoria ocupacional	Directiu o professional liberal	3
	Empleat tècnic, administratiu	3
	Serveis	5
	Agricultura, forestal i pesca	4
	Producció de precisió,	4
	Operari, fabricació i mà	5

Taula 5. Mitjanes aritmètiques sobre el nombre de fills/es segons la categoria professional dels pares.

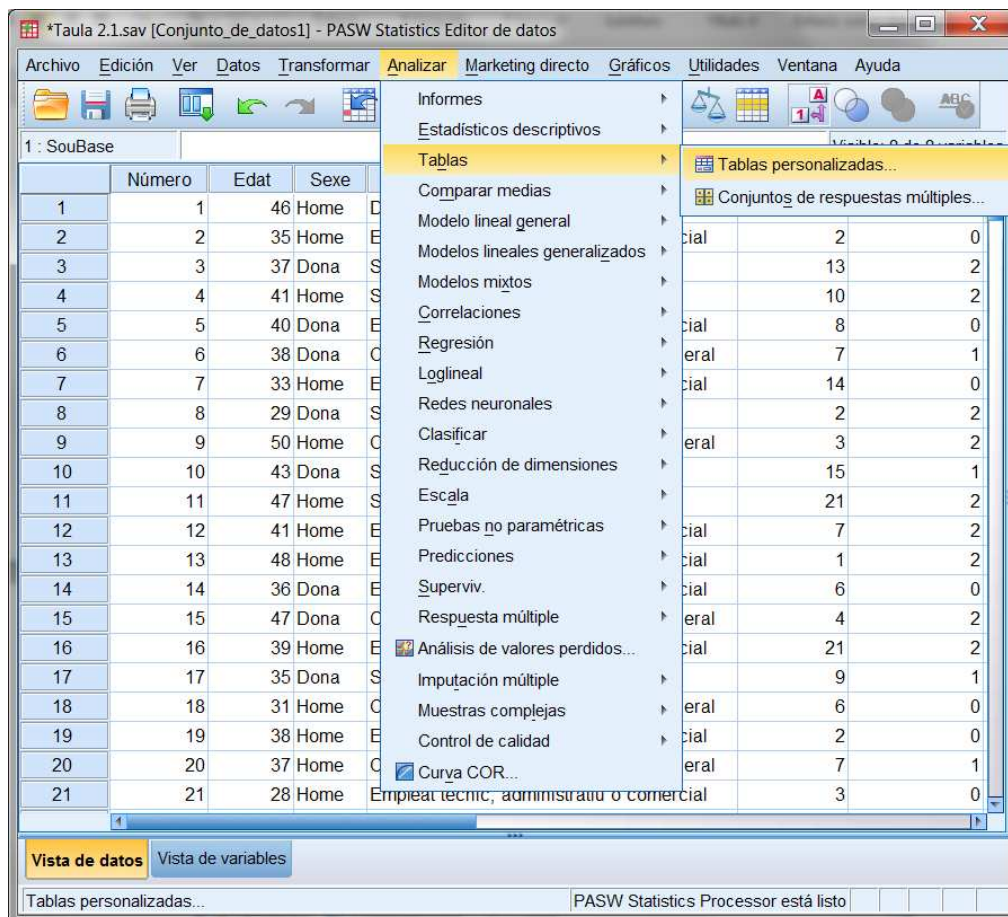
També podríem demanar més variables per fer subgrups, com per exemple el sexe. I així, obtindríem la mitjana de fills, segons la categoria professional, d'homes i dones. Podem arribar a demanar taules tant complexes com ens interressi.



L'SPSS ens ofereix crear aquest tipus de taules. En la figura 2 podeu veure un dels possibles processos per demanar una taula, a través de les *taules personalitzades*.

Els subgrups els faríem arrossegant al rectangle *Filas* alguna variable qualitativa (la categoria professional en l'exemple) i arrossegant al rectangle *Columnas* la variable quantitativa a resumir (el nombre de filles i fills en l'exemple). En l'exemple que es veu a la figura 2.2 es demana la mitjana aritmètica de la variable *edat* en funció del sexe; obtindrem així una taula que ens indiqui les mitjanes d'edat d'homes i dones. Des d'aquesta opció podem fer taules amb els estadístics que volem.

² La mitjana aritmètica és el resultat de sumar tots els valors i dividir-ho entre el nombre de valors; és un valor mitjà. Aquest indicador estadístic i altres es tractaran en el següent apartat en aquest mateix tema.



I llavors:

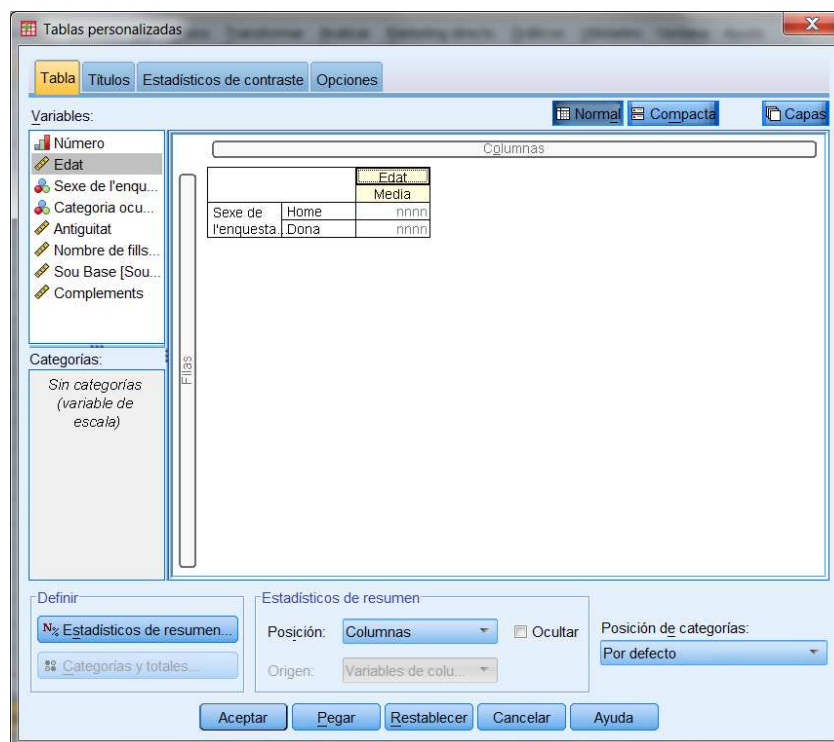


Figura 2. Les taules d' estadístics descriptius des de l'SPSS.

1.1.2 Representació gràfica

Les tècniques de **Representació Gràfica** són un altre instrument per a presentar la informació original. La pèrdua d'informació és major, però presenta l'avantatge d'establir diferències entre dues distribucions a simple vista.

- Representació Gràfica:

Ordenades:
freqüències
individuals o
acumulades



Abscisses: Categories o
valors de la variable

Per a la representació gràfica és necessari diferenciar si la variable és qualitativa, quantitativa continua o discontinua. A més, la correcta realització d'un gràfic necessita que s'hi inclogui la font d'origen de les dades i el títol que indiqui la variable a la que correspon la informació.

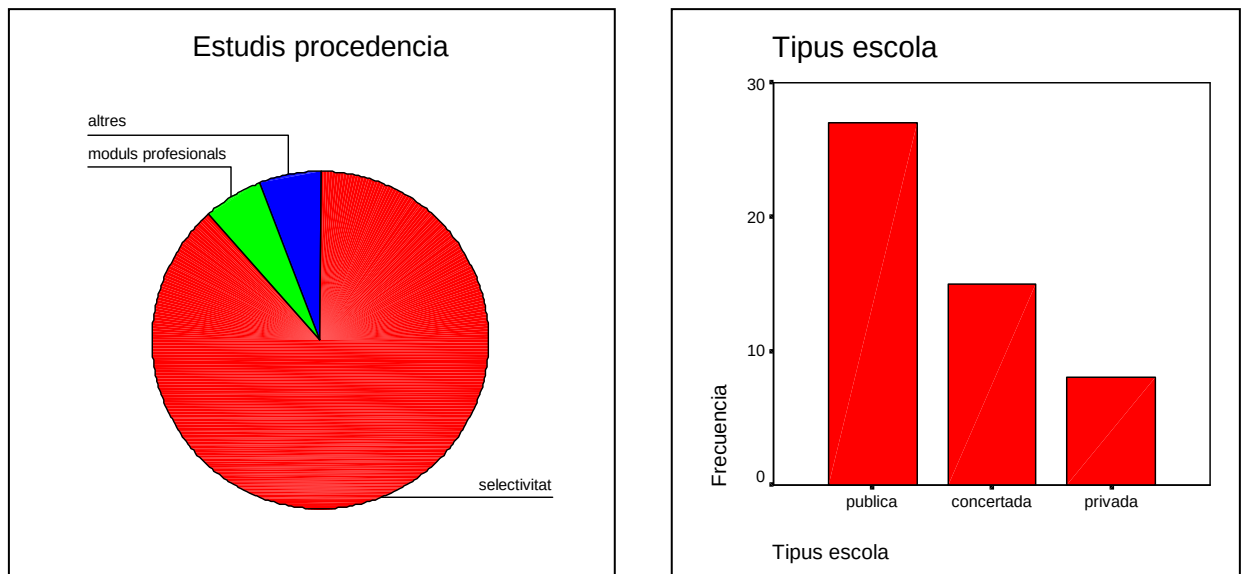
a) Variables qualitatives:

Per a les variables qualitatives la representació permet els gràfics següents:

- **Diagrama de barres:** gràfic de barres separades entre sí, amb la mateixa distància i amplitud de base, amb longituds proporcionals a les freqüències.
- **Ciclograma o de sectors:** cercle que es divideix en sectors circulars proporcionals a la freqüència de cada categoria.
- **Diagrama de Rectangles:** la superfície d'un rectangle es reparteix proporcionalment a les freqüències.
- **Pictograma:** dibuixos amb la mida proporcional a les freqüències.



Alguns dels gràfics que ofereix l'SPSS són el de barres i el de sectors, tal com podem veure al gràfic 2.1. Aquests es poden aconseguir des de l'opció del menú *Gràficos*. A la figura 2.3 podem consultar altres vies per demanar aquests gràfics.



Gràfic 1. Gràfics de sectors i de barres obtinguts amb l'SPSS

b) Variables quantitatives discretes:

Les variables quantitatives discretes o discontinues es representen gràficament, entre d'altres, amb el *diagrama lineal*. Està format per blocs separats sobre cadascun dels valors de la variable i s'eleva fins la freqüència o el nombre de vegades que es repeteix. Molt semblant al diagrama de barres, però enlloc de rectangles simètrics, s'utilitzen línies verticals.

c) Variables quantitatives contínues:

Les variables quantitatives contínues es poden representar gràficament amb els següents gràfics:

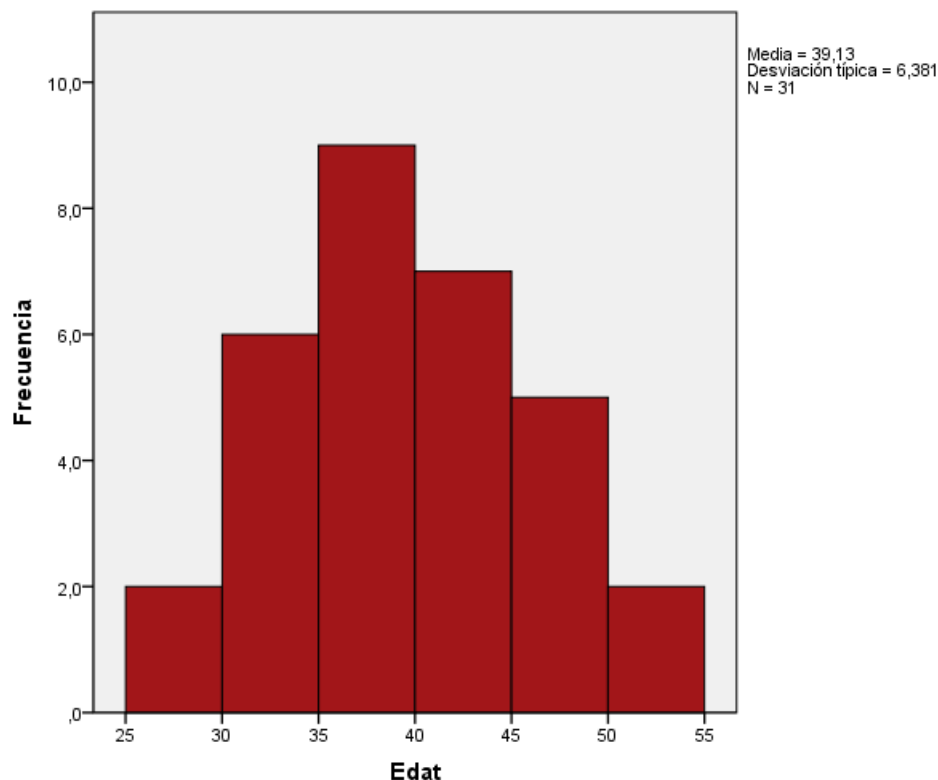
- **Polígon de Freqüències:** és un gràfic de les freqüències de classe en relació a la marca de classe (x'), on els punts són units en una línia. A l'eix de les ordenades es poden representar les freqüències absolutes, les relatives o les percentuals i a l'eix de les abscisses, les marques de classe.
- **Polígon de Freqüències Acumulades (Ogiva):** és una gràfica que representa les freqüències acumulades. Es representa igual que el polígon, però en l'eix de les ordenades hi ha les freqüències absolutes acumulades o les freqüències percentuals acumulades. A les abscisses, el límit inferior (li) del primer interval i els superiors (Li) de la resta d'intervals. És important per representar gràficament el nombre de casos que es tro-

ben per sobre o per sota d'un determinat valor. Exemple: conèixer el percentatge d'aprovats o que hagin tret menys d'un notable.

- **Histograma:** consisteix en un conjunt de rectangles units, amb un eix d'ordenades format per les freqüències absolutes, relatives o percentuals, i a l'eix de les abscisses hi representen el límit inferior del primer interval i els Límits superiors de la resta dels intervals.



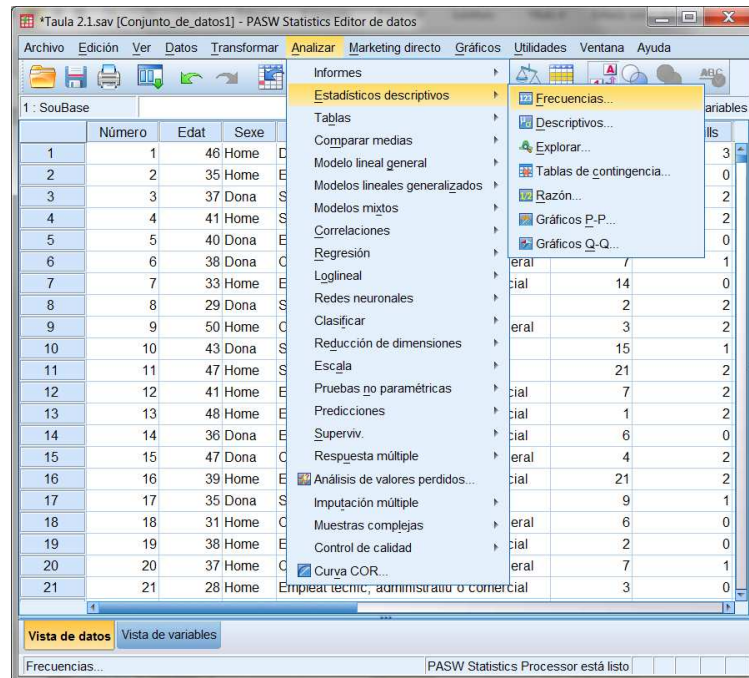
Un dels gràfics quantitativs més habitual que ofereix l'SPSS és l'histograma (gràfic 2) que es pot aconseguir des de l'opció del menú *Gràficos*. A la figura 3 podeu consultar altres vies per demanar aquest gràfic.



Gràfic 2. Histograma obtingut mitjançant l'SPSS.

Per últim, en la figura 3 queda reflectit un altre procediment a seguir per tal de sol·licitar els gràfics i les taules exposats en aquest tema. Com podem observar hem d'accedir a l'opció *Estadísticos descriptivos* del menú i dins aquesta seleccionar *Frecuencias*, per tal de demanar tant la taula com els gràfics corresponents a les variables que ens interessin.

Per tant, l'SPSS ofereix en la finestra d'Estadística Descriptiva i en concret en el càlcul de Freqüències, tres tipus de gràfics bàsics: barres i sectors que s'utilitzen per les variables qualitatives i histogrames per les variables quantitatives continues.



I llavors:

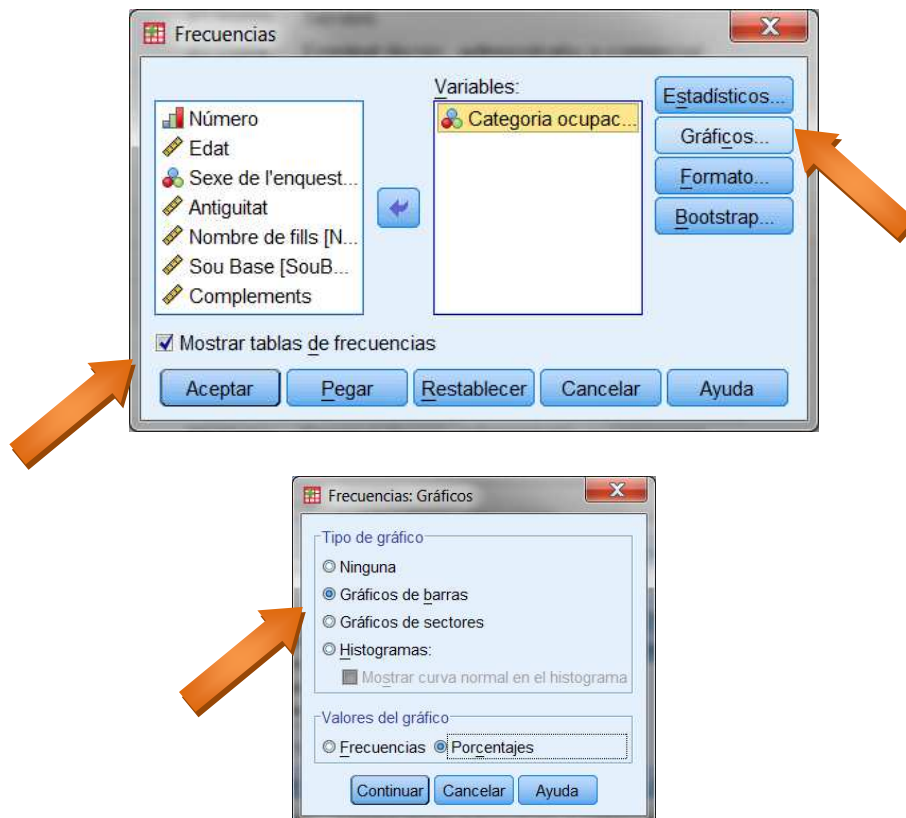


Figura 3. Procés per demanar gràfics junt amb les taules de freqüències.

Altres tipus de gràfics: diagrama de caixes

Als anys vuitanta, amb la proposta de noves tècniques d'estadística descriptiva assistida per ordinador, es va començar a parlar d'*Anàlisi exploratori de dades* (EDA).

Aquest és un conjunt de tècniques innovadores, tant a nivell de gràfiques com d'anàlisi, que té com a objectiu aconseguir un coneixement previ de les dades a partir d'un examen visual de les mateixes. Per tant, entre d'altres aspectes, recull la utilitat i necessitat de plantejar la construcció de diferents tipus de gràfics que siguin bons sintetitzadors de la informació proporcionada per les dades.

Des d'aquesta perspectiva es considera el **Diagrama de Caixa** com un gràfic que permet obtenir una impressió global dels aspectes més rellevants de la distribució d'una variable.

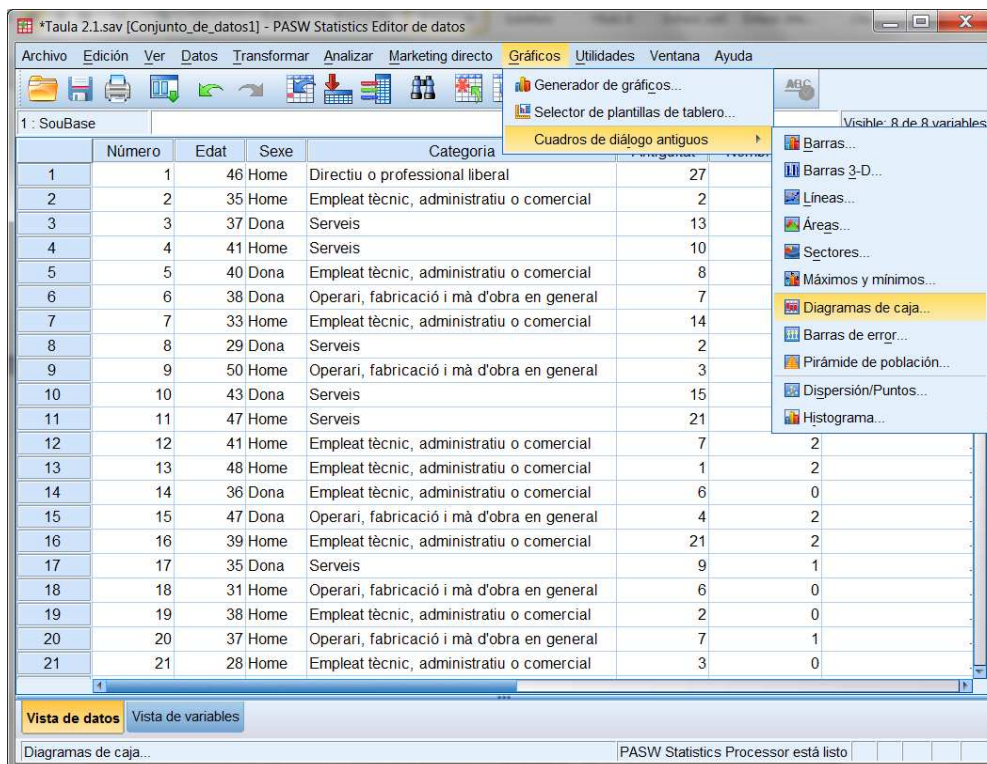
En aquest gràfic es representen estadístics bàsics de la distribució: mediana, quartils i els valors extrems. És a dir, que amb una simple observació de la gràfica podem obtenir una sèrie d'indicadors estadístics³: Amplitud, rang interquartílic, mediana, simetria (posició que ocupa la mitjana respecte els quartils).

Per tal d'obtenir aquest gràfic seguirem les passes assenyalades a la figura 4. Essent el primer que hem de fer accedir a l'opció *Gráficos* del menú i dins d'aquesta seleccionar el diagrama desitjat (*diagrama de cajas*).

Un cop seleccionem el *Diagrama de cajas* del menú se'ns obrirà un nou quadre de diàleg en el que haurem d'especificar si realitzarem una gràfica simple o agrupada, essent la diferència entre ambdues que amb la última les dades es poden agrupar en funció dels valors d'una tercera variable.

Escollirem l'opció *Simple* i tot seguit haurem de decidir-nos entre *Resúmenes para grupos de casos* o bé *Resúmenes para distintas variables*. La primera opció ens permet definir les variables amb les que pretenem elaborar el gràfic, essent necessari que una sigui quantitativa (*Variable*) i l'altra qualitativa (*Eje de categorías*). Mentre que amb la segona opció podrem definir una o més variables sense necessitat de representar-les en funció d'una altra sinó a elles mateixes.

³ Aprofundirem sobre els indicadors estadístics en el punt 2.2 d'aquest mateix tema.



I llavors:

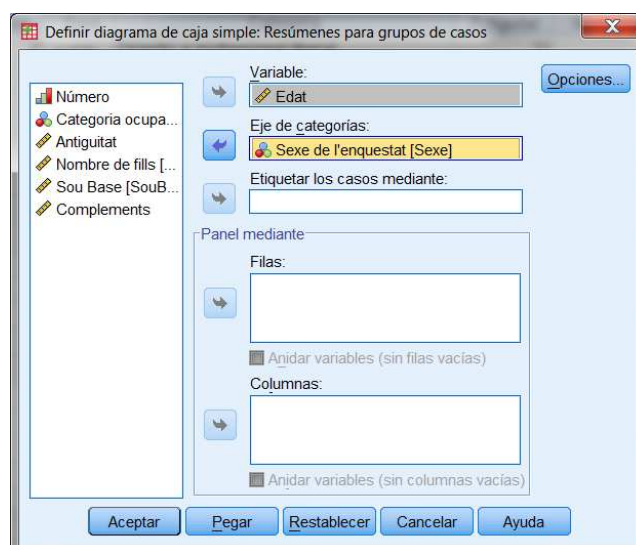
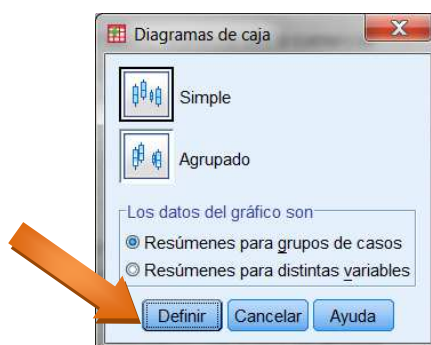
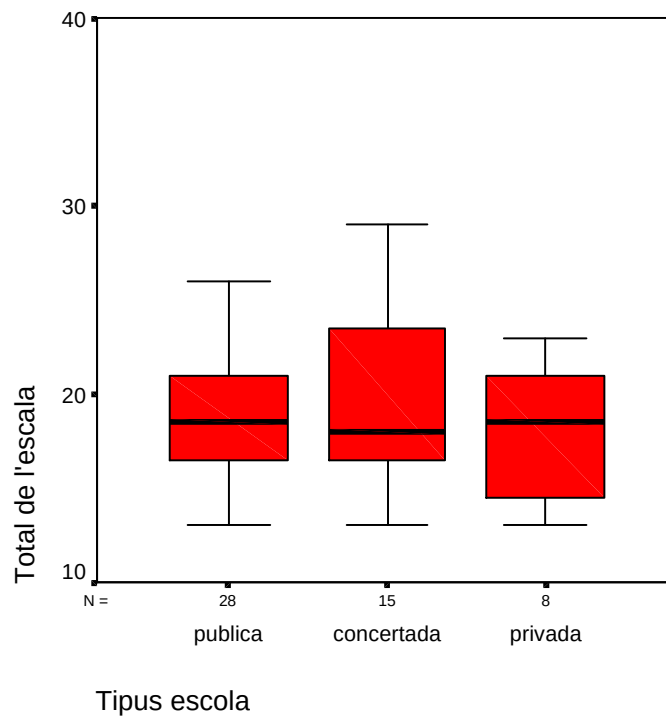


Figura 4. Procés per demanar diagrames de caixa.

Al gràfic 3 podem observar un Diagrama de caixa elaborat amb les dues variables: Total de l'escala i Tipus d'escola.



Gràfic 3. Diagrama de caixa obtingut mitjançant l'SPSS.

Per tal de poder interpretar un diagrama de caixa haurem de tenir coneixement dels indicadors estadístics que veurem en el proper apartat. La informació, però, que podem obtenir d'aquest tipus de gràfic és:

- D'una banda, la *posició de la mediana*, tendència central del grup. Aquesta ve marcada per la línia que es troba dins la caixa, havent-nos de fixar si es troba en el centre (distribució simètrica) o bé està pròxim al límit inferior (asimetria positiva) o superior (asimetria negativa) de la caixa.
- D'altra part, ens trobem que els límits d'aquesta caixa venen definits pels *quartils 1 i 3* (centils 25 i 75), obtenint-se també d'aquestes dades el *rang interquartí·lic*.
- També hem de tenir en compte que les línies que surten de la caixa central fan referència als *valors màxim i mínim* de la distribució.
- Pel que fa als *casos extrems*, en funció de la distància que prenguin respecte el cinquanta per cent central de la distribució, distingirem entre

dos tipus: els remots i els extrems. Aquells casos que estan més allunyats s'anomenen remots (*far outliers*) i estan representats per un asterisc (*), en canvi aquells que estan més propers, tot i trobar-se lluny de la resta del grup, s'anomenen extrems (*outliers*) i es representen per un cercle (0). Sempre, acompanyant a aquests símbols anirà el número d'identificació dels subjectes concrets.

Tots aquests elements es troben resumits en la següent figura, amb l'objectiu de facilitar la seva comprensió i aplicació pràctica.

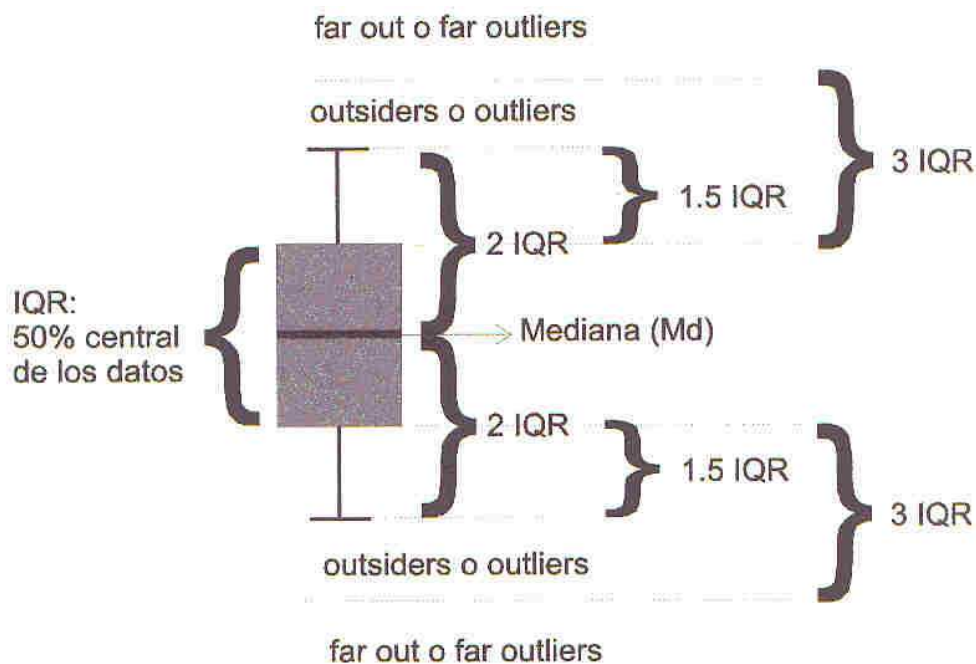


Figura 5. Elements per interpretar un gràfic de caixa.

Per últim, hem de dir que aquests gràfics són molt útils per a comparar la distribució dels valors entre diferents grups, per exemple en el gràfic 3 podem comparar fàcilment les distribucions de l'escola pública, privada i concertada.

1.2 Indicadors Estadístics

Una vegada que sabem com es recullen, organitzen i presenten les dades, ens queda estudiar com es resumeixen. 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.

Bàsicament, amb un sol indicador 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 donar resposta a les següents qüestions:

- 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)
- Quina forma adopta la distribució? (indicadors de forma)

A continuació presentem cada un d'aquests elements:

1.2.1 Indicadors de posició

Els *indicadors de posició* ens mostren la posició que ocupa un valor de la variable en la distribució. Divideixen la distribució en diverses parts iguals (per exemple en 100 parts) i indiquen el lloc que li correspon a un valor concret, en relació al grup. Per exemple, ens ajuden a respondre a qüestions com les següents:

- Quin percentatge d'alumnes no passa de l'aprobat (és a dir, d'una puntuació de 5)?
- Quin percentatge està per sobre del notable?
- I entre el notable i l'aprobat?
- Etc.

Segons el nombre de parts en que dividim una distribució ordenada, podem trobar-nos amb els següents indicadors:

- **Percentils o centils (C):** es divideix en 100 parts iguals i donen la situació de l'individu dins del grup en tant per cent. Indiquen el percentatge de la distribució que queda per sota d'un determinat valor.

- **Decils (D):** el mateix que els centils, però dividint la distribució en 10 parts iguals.
- **Quartils (Q):** es divideix la distribució en 4 parts iguals anomenats "quarts" i cada part és equivalent al 25% de la distribució. Els quartils són els tres valors que limiten els quatre quarts.

Un exemple: si cerquem el percentil 48 del rendiment acadèmic d'un grup d'estadística i el resultat que obtenim és 6'35; això vol dir que el 48% de l'alumnat d'aquest grup ha obtingut una nota que és igual o inferior a 6'35.

En la figura 6, podem veure les equivalències entre els diferents indicadors de posició, abans esmentats.

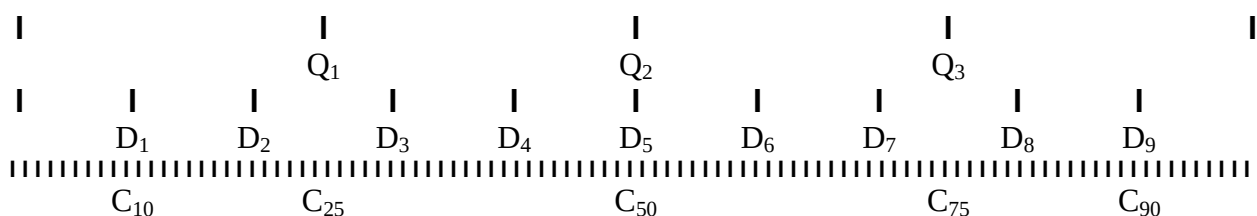


Figura 6. Equivalències entre alguns indicadors estadístics.

COM ES CALCULA?

1. Localitzar l'interval on es troba el centil buscat (Pa). Essent aquest el primer del qual el percentatge acumulat sigui igual o superior al centil que busquem.
2. Aplicar una fórmula:

$$C_{?} = li + \frac{\frac{? \cdot n}{100} - f_{a_{i-1}}}{f_i} \cdot i \quad D_{?} = li + \frac{\frac{? \cdot n}{10} - f_{a_{i-1}}}{f_i} \cdot i \quad Q_{?} = li + \frac{\frac{? \cdot n}{4} - f_{a_{i-1}}}{f_i} \cdot i$$

On:

? = Percentil, decil o quartil que cerquem

li = Límit inferior de l'interval

$f_{a_{i-1}}$ = Freqüència Absoluta Acumulada de l'interval anterior al cercat

f_i = Freqüència Absoluta Individual de l'interval

i = Amplitud de l'interval



Amb l'SPSS podem demanar tots els quartils, tots els decils (*puntos de corte para 10 grupos iguales*), els percentils que desitgem i altres indicadors (*puntos de corte para ? grupos iguales*). No obstant, els outputs sempre ens ofereixen percentils, donat que l'equivalència entre els indicadors és força senzilla.

1.2.2 Indicadors de tendència central

S'anomenen *indicadors de tendència central* d'una distribució perquè fan referència al lloc on es concentren les dades (normalment, en la part central de la distribució: valors mitjans, valors centrals...).

Hi ha diverses tipologies, sent les més comunes: la mitjana aritmètica, la mediana i la moda. Cadascuna té avantatges i desavantatges segons les dades i l'objectiu que es persegueix.

- La **Mitjana Aritmètica** ($\bar{X}$): Es correspon amb la idea usual que tenim del valor mig i equival a la suma de totes les dades d'una distribució dividida pel nombre de casos. Es representa pel símbol $\bar{X}$ per a la mostra i μ per a la població.

COM ES CALCULA?

Amb la calculadora científica: Tecla $\bar{X}$

Es calcula segons es troben les dades:

1. Dades sense agrupar: es sumen tots els valors de les dades i es divideixen entre el nombre de dades.

$$\bar{X} = \frac{\sum X_i}{n}$$

2. Dades agrupades: cada dada es multiplica per la seva freqüència i la suma es divideix pel nombre total de casos.

$$\bar{X} = \frac{\sum X_i \cdot f_i}{n}$$

3. Dades agrupades en intervals: cada marca de classe es multiplica per la seva freqüència i la suma es divideix pel nombre total de casos.

$$\bar{X} = \frac{\sum X_i' \cdot f_i}{n}$$

- La **Mediana (Md)**: És el valor que ocupa la posició central d'una sèrie estadística ordenada. En altres paraules, és el valor que deixa igual nombre de valors per sobre que per sota, deixant a cada costat el 50% de la distribució.

COM ES CALCULA?

Degut a què aquest indicador equival al centil 50, la forma de calcular la mediana seria fent servir la fórmula assenyalada pels centils i substituint ? (centil, decil o quartil que cerquem) per 50 (veure apartat d'Indicadors de posició).

- La **Moda (Mo)**: És la dada que té la freqüència més alta, o bé que es repeteix un major nombre de vegades. Segons el nombre de dades que tenen la freqüència més alta, podem parlar de distribucions:
 - *Unimodals*: Una única dada amb la freqüència més alta.
 - *Bimodals*. Dues dades no consecutives amb la freqüència més alta.
 - *Multimodals*. Més de dues dades no consecutives amb la freqüència més alta.
 - *Amodals*. No apareix cap moda clara.

Si hi ha dues dades que corresponen a la freqüència més alta, però són consecutives, es fa la mitjana per a calcular la Moda. Sent la distribució, en aquest cas, unimodal.



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, en l'output se'ns indica quan es tracta de distribucions multimodals i ens ofereix el valor modal més petit.

1.2.3 Indicadors de dispersió

Per a descriure una distribució necessitem saber alguna cosa més que la tendència d'agrupació de les dades o la posició que ocupen. Les mesures de tendència central no ens donen informació sobre els valors més elevats o els més baixos, només sobre on tendeixen a agrupar-se les dades, però no si estan més o menys disperses.

En aquest sentit, les mesures de dispersió ens informen de com es distribueixen els valors al voltant dels centrals (si estan lluny o propers). Es a dir, ens indiquen la homogeneïtat o heterogeneïtat de les dades. Per tant, amb aquests valors podem arribar a descriure com és la variabilitat o dispersió d'un grup.

Estudiarem els següents indicadors:

- **Rang, Recorregut o Amplitud (R):** És la diferència entre els valors màxim i mínim de la distribució.

COM ES CALCULA?

Utilitzem les següents fórmules:

$$R = V_{\text{màx}} - V_{\text{mín}}$$

$$\text{Intervals: } R = V_{\text{màx}} - V_{\text{mín}} + 1$$

Aquesta mesura és senzilla de calcular i ens dóna una visió intuïtiva de la dispersió de la distribució. No obstant, no té en consideració els valors intermedis, únicament es basa en els extrems. Per aquest motiu, el rang normalment s'utilitza en combinació amb altres indicadors que tinguin en compte un major volum d'informació. Per exemple, amb el rang interquartíl·lic (RQ, RI) que veurem més endavant.

- **Variància (S^2):** La importància d'aquest indicador és fonamental per l'estadística inferencial. Per ara, podem afirmar que a major variància, major dispersió. La variància és la mitjana de les diferències al quadrat de cada valor respecte a la mitjana de la distribució.

COM ES CALCULA?

Amb la calculadora: tecla σ^2 (S és l'estadístic i σ el paràmetre)

Càlcul similar a la desviació mitjana, però amb els quadrats de les diferències:

$$S^2 = \frac{\sum (X - \bar{X})^2}{n}$$

Dades sense agrupar

$$S^2 = \frac{\sum (X - \bar{X})^2 \cdot f_i}{n}$$

Dades agrupades

$$S^2 = \frac{\sum (X' - \bar{X})^2 \cdot f_i}{n}$$

Dades agrupades en intervals

Com podem veure, les unitats de la variància no són les mateixes que les de la mostra, ja que estem elevant les diferències al quadrat. Això implica certes dificultats en interpretar la magnitud de la dispersió del grup. Per superar aquesta dificultat disposem de la *desviació típica*, que veurem a continuació.

- La **Desviació Estàndard, típica o tipus (S, S_x)**: La desviació típica es la mesura de dispersió més utilitzada en l'estadística descriptiva. En expressió matemàtica és l'arrel quadrada de la variància.

En una corba normal⁴, la desviació típica equival a la distància compresa entre la mitjana i el punt d'inflexió (que modifica la corba de còncava a convexa).

COM ES CALCULA?

Amb la calculadora científica: Tecla σ

O bé, calculant la variància i la seva arrel quadrada:

$$\sigma = \sqrt{S^2}$$

- **Rang Interquartíl·lic (RQ, RI)**: És la diferència entre els quartils tercer i primer. És un índex de dispersió que ens pot servir per evitar l'existència de buits d'informació. El rang interquartíl·lic ens permet obtenir informació quan els valors extrems no són gaire significatius. Ens dona bàsicament, informació sobre el 50% central de la distribució.

⁴ Una corba normal correspon a la representació gràfica d'una variable que segueix la Llei Normal. En el següent tema aprofundirem en la llei normal.

COM ES CALCULA?

Utilitzem la següent fórmula:

$$RI = Q_3 - Q_1$$

- **Coefficient de Variació (CV) de Pearson:** Expressa la relació que hi ha entre la desviació típica d'una mostra i la seva mitjana aritmètica.

És una unitat de mesura relativa, no depèn de les unitats de mesura de la variable, per tant permet la comparació entre diferents mostres. Es tracta de saber quina mitjana de dues mostres és més representativa, en relació a les seves desviacions típiques.

Un CV gran ens indica que la mostra està molt desviada, per tant la mitjana no és massa representativa. Si és petita, indica que la mitjana és un bon indicador i que la mostra es troba força concentrada al voltant d'aquesta mesura.

COM ES CALCULA?

Utilitzem la següent fórmula:

$$CV = \frac{S_x}{X} \cdot 100$$

1.2.4 Indicadors de forma

Els indicadors de forma són aquells que ens donen informació sobre la forma de la distribució. Ens indiquen si els valors es distribueixen simètricament a dreta i esquerra de la mitjana aritmètica (simetria), així com, l'alçada de la corba que representa aquesta distribució (curtosi). A continuació, detallem un xic més aquests indicadors:

- **Asimetria:** Indica si la distribució dels valors al voltant de la mitjana aritmètica (a la seva dreta i esquerra) és més o menys simètrica.

COM ES CALCULA?

Es pot valorar a nivell gràfic a partir de la representació dels valors, o bé de forma més exacta, a partir del coeficient d'asimetria de *Pearson* (As) o també amb qualsevol de les fórmules següents que ens proporcionen un valor aproximat d'asimetria:

$$As_1 = \frac{\bar{X} - Mo}{S} \quad As_2 = \frac{3(\bar{X} - Md)}{S}$$

La interpretació del resultat és la següent:

$As < -0,5 \rightarrow$ la distribució és asimètrica a la dreta (asimetria negativa)

$-0,5 < As < 0,5 \rightarrow$ es considera simètrica, perquè en molt poques ocasions el coeficient serà igual a zero.

$As > 0,5 \rightarrow$ la distribució és asimètrica a l'esquerra (asimetria positiva)



L'SPSS utilitza una fórmula exacta (coeficient d'asimetria de Pearson), mantenint, però, la mateixa interpretació dels resultats. Gràficament, ho podem representar, tal com es veu a la figura 7.

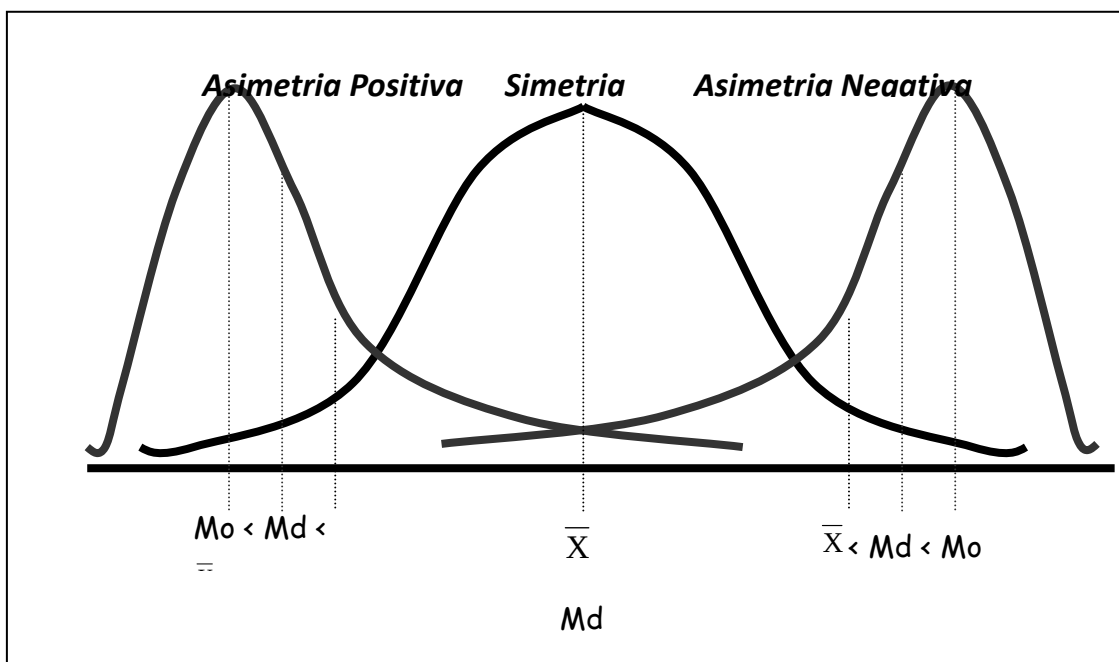


Figura 7. Indicador de forma d'Asimetria.

- **Curtosi:** És l'estudi de l'apuntament de la corba en la part central de la distribució, en relació a la corba normal.

COM ES CALCULA?

Podem mesurar-la utilitzant la fórmula exacte de *Pearson* o bé la fórmula aproximada del coeficient de curtosi percentí·lic:

$$C_u = \frac{Q_3 - Q_1}{2 \cdot (C_{90} - C_{10})}$$

La interpretació del resultat és la següent:

$C_u < 0,263 \rightarrow$ Platicúrtica (plana)

$0,263 < C_u < 0,363 \rightarrow$ Mesocúrtica (normal)

$C_u > 0,363 \rightarrow$ Leptocúrtica (aixecada)



Amb l'SPSS s'utilitza la fórmula exacte de *Pearson* i s'interpreta de la següent manera: és mesocúrtica quan el coeficient és 0. Leptocúrtica més gran de 0 i platicúrtica menor de 0. Però pot acceptar-se amb els següents marges de flexibilitat:

$C_u < - 0,5 \rightarrow$ Platicúrtica

$- 0,5 < C_u < 0,5 \rightarrow$ Mesocúrtica

$C_u > 0.5 \rightarrow$ Leptocúrtica

Gràficament, ho podem representar, tal com es veu a la figura 8.

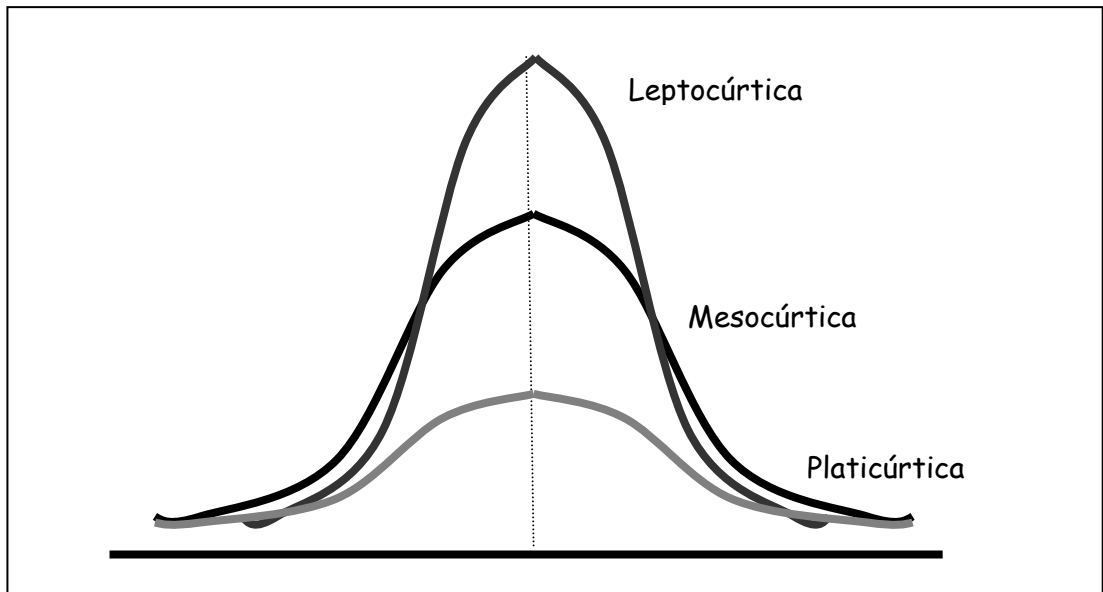
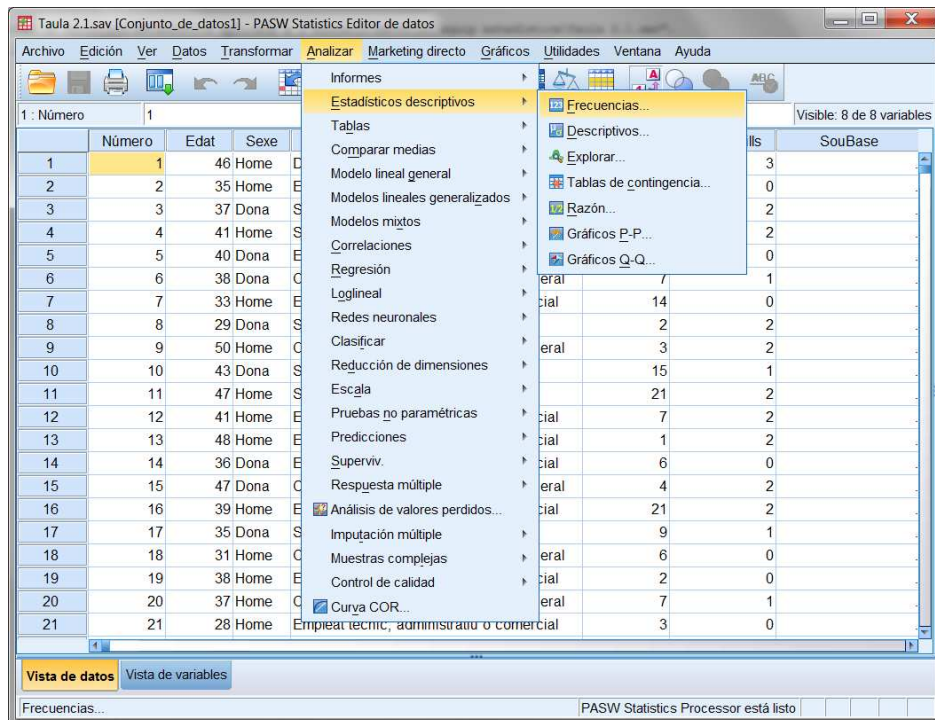


Figura 8. Indicador de forma de curtosi.

1.2.5 Els indicadors estadístics a l'SPSS

Bona part dels indicadors que hem anat explicant al llarg d'aquest tema es calculen des del paquet estadístic SPSS. Hi ha diverses maneres d'obtenir aquests càlculs des de l'SPSS, una de les més senzilles és des de l'opció *frecuencias*, seguint les passes que es detallen a la figura 9.



I llavors:

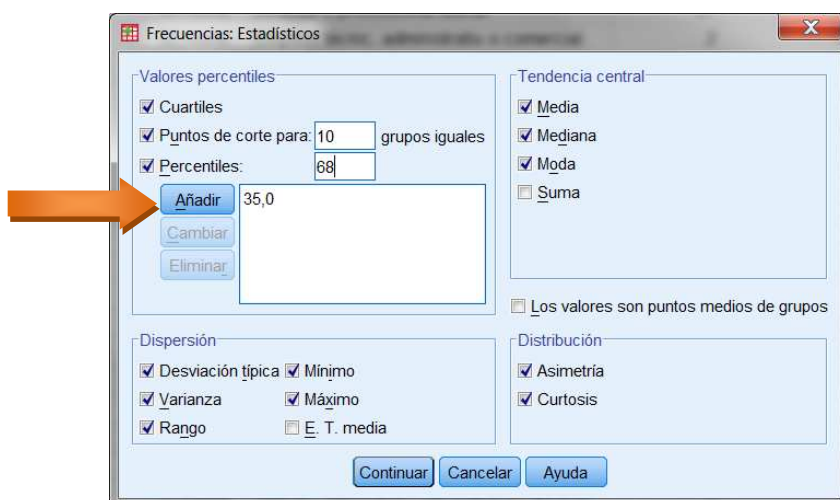


Figura 9. Procés per demanar indicadors estadístics.

Proposem una especial atenció a l'organització del quadre de diàleg *Frecuencias: Estadísticos*. Aquest quadre es divideix en 4 parts, distingint els quatre tipus d'indicadors que hem treballat:

- A la part superior esquerra s'identifiquen alguns *indicadors de posició*: quartils, decils⁵ i percentils⁶.
- A la part superior dreta s'identifiquen alguns *indicadors de tendència central*: mitjana aritmètica, mediana i moda.
- A la part inferior esquerra s'identifiquen alguns *indicadors de dispersió*: la desviació típica, variància, el rang, així com els valors mínim i màxim de la distribució.

⁵ Recordem que els decils poden obtenir-se indicant "10" en els: *puntos de corte para ____ grupos iguales*.

⁶ Podem demanar qualsevol percentil concret, tenint en compte d'*afegir* cada vegada que en demanem un de nou.

- A la part inferior dreta s'identifiquen alguns *indicadors de forma*: l'asimetria i la curtosi.

Una vegada sabem com demanar i calcular aquests indicadors estadístics, cal aprofundir en com interpretar els resultats obtinguts, aspecte que proposem en el següent apartat.

1.2.6 Com descriure una variable utilitzant els indicadors estadístics

En aquest apartat ens proposem posar a la pràctica tots els conceptes que s'han anat treballant al llarg d'aquest tema (els indicadors estadístics). D'aquesta manera, utilitzarem un exemple per poder aplicar els indicadors de tendència central, de posició, de dispersió i de forma. Pretenem extreure una descripció coherent i integrada de les dades, facilitant la interpretació dels resultats dels diferents càlculs estadístics aplicats.

De les diferents variables de la matriu de dades de l'exemple anterior, hem seleccionat la variable edat. Donat que es tracta d'una variable quantitativa contínua, podem utilitzar els diferents indicadors estadístics treballats, per fer la descripció de l'edat del grup. Concretament, fent ús de l'SPSS demanem els següents indicadors:

- Tendència central: mitjana, mediana i moda
- Posició: decils i quartils
- Dispersió: desviació típica, variància, rang o amplitud, valor mínim i valor màxim.
- Forma: asimetria i curtosi

En la taula 6 trobem els resultats que s'obtenen de l'SPSS sobre els indicadors abans esmentats.

Estadísticos		
EDAT		
N	Válidos	1188
	Perdidos	42
Media		17,9545
Mediana		18,0000
Moda		18,00
Desv. típ.		,90656
Varianza		,82186
Asimetría		1,673
Error típ. de asimetría		,071
Curtosis		5,354
Error típ. de curtosis		,142
Rango		9,00
Mínimo		15,00
Máximo		24,00
Percentiles	10	17,0000
	20	17,0000
	25	17,0000
	30	18,0000
	40	18,0000
	50	18,0000
	60	18,0000
	70	18,0000
	75	18,0000
	80	18,0000
	90	19,0000

Taula 6. Output d'indicadors estadístics de la variable edat.

D'aquestes dades podem realitzar la següent proposta de descripció. En primer lloc, s'ha de destacar que disposem d'informació sobre l'edat de 1188 dels 1230 subjectes; tenim, doncs, 42 valors perduts⁷. La mitjana del grup és de 17,96 anys, sent els 18 anys l'edat més freqüent. De fet, el valor que deixa a dreta i esquerra el 50% de la distribució, també són els 18 anys. Aquesta proximitat entre els tres indicadors de tendència central podria fer-nos pensar que es tracta d'una distribució simètrica. No obstant, el coeficient d'asimetria ens indica que la distribució és asimètrica positiva. Aquest fet, doncs, pot ser donat perquè es tracta d'un grup molt poc dispers, això ho podem verificar calculant el coeficient de variació de Pearson (5 %), tot i que també podem observar el resultat del coeficient de curtosis que ens indica que la distribució és leptocúrtica, és a dir, que la majoria dels individus es troben al voltant de la mitjana. És degut a aquesta poca dispersió del grup que tot i que l'edat oscil·la entre els 15 i els 24 anys, el 80% té entre 17 i 18 anys. Concretament, la mitjana de les diferències entre les edats de les persones del grup, no arriba a un any.

⁷ Recordem que hi ha valors perduts pel sistema (caselles buides) i valors perduts per l'usuari (definitos per l'usuari com a no vàlids).