

Pràctica 2. Anàlisi descriptiva univariàble: variables qualitatives

L'objectiu d'aquesta pràctica és introduir l'ús del programari SPSS per a l'anàlisi descriptiva d'una única variable quan aquesta està mesurada a nivell nominal i ordinal (variables categòriques o qualitatives). Per tal de fer una anàlisi descriptiva bàsica procedirem a obtenir taules de distribucions de freqüències, gràfics per a representar la informació de les taules (de barres, de sectors i de Pareto), així com els estadístics de resum adients per a aquest tipus de variables: la moda, per a les variables nominals, i a més la mediana i els percentils, per a les variables ordinals.

► Recordeu activar les opcions:

- “Mostrar comandos en anotaciones” a la pestanya de “Visor”.
- “Nombre y etiquetas” per a les variables i “Valores i etiquetas” per als valors a l'apartat de “Etiquetado de tablas pivote” de la pestanya de “Etiquetas de resultados”.

► En aquesta pràctica treballarem amb l'arxiu **GSS93 reducido.sav** del programari, ubicat a la carpeta de l'SPSS.


- Nombre de casos (persones enquestades): _____
- Nombre de variables (característiques de les persones): _____

1. Anàlisi descriptiva de variables qualitatives nominals

Considerarem la variable qualitativa nominal **SEXO**.

Codificada de tipus “numèrica” amb valors 1 i 2. Aquests valors s'etiqueten amb: 1 ‘Hombre’ 2 ‘Mujer’.

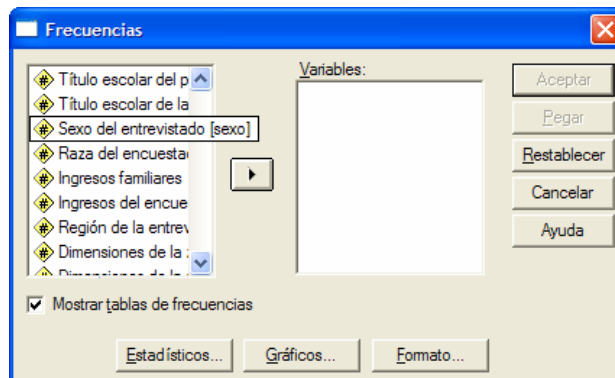
► Observeu la columna de la variable a l'editor de dades. Els valors són l'1 i

el 2. Si activeu l'opció “Etiquetas de valor” , apareixeran les etiquetes. Aquesta és una forma ràpida i visual per a determinar si una variable és qualitativa: quan necessitem etiquetes per a saber el significat del valor numèric la variable és qualitativa, si el valor té un significat per sí mateix no li calen, i la variable serà quantitativa o numèrica.

► A continuació demanaren: (1) la taula de distribució de freqüències, (2) un gràfic de barres i (3) l'estadístic de la moda. A través del menú:

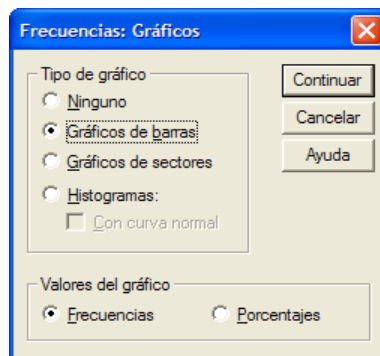
Analizar / Estadísticos descriptivos / Frecuencias...

Ens apareix el quadre de diàleg del procediment “Frecuencias”:



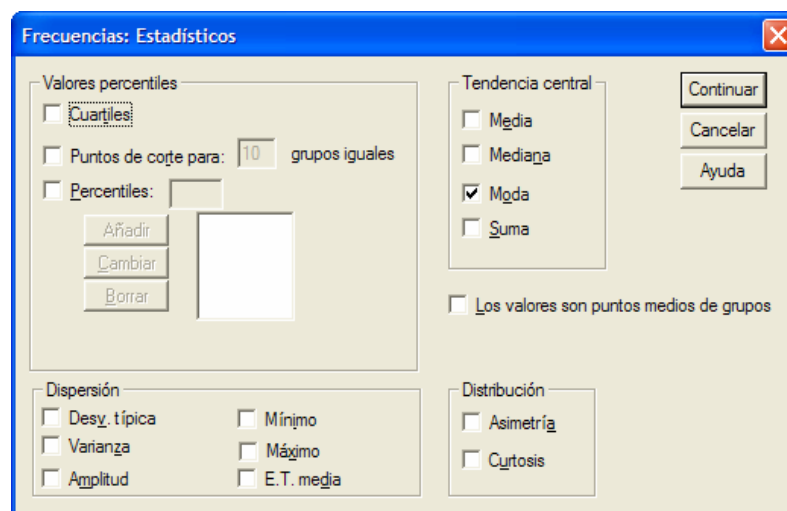
(1) Seleccionem la variable **sexo** i la col·loquem al requadre de “Variables”. D’aquesta manera obtenim la taula de distribucions de freqüències. L’opció “Mostrar tablas de frecuencias” està activada.

(2) Per a demanar el gràfic de barres cliquem sobre el botó “Gráficos...”. Ens apareix el quadre de diàleg on seleccionarem “Gráfico de barras” (també podríem haver escollit “Gráfico de sectores”):



I cliquem sobre “Continuar”.

(3) Per a demanar l’estadístic cliquem sobre el botó “Estadísticos...”. Ens apareix el quadre de diàleg on seleccionarem la **Moda** entre les opcions de mesures de tendència central:



I cliquem sobre “Continuar”. Finalment cliquem “Aceptar” al quadre de diàleg principal i observem els resultats següents:

Frecuencias

Estadísticos

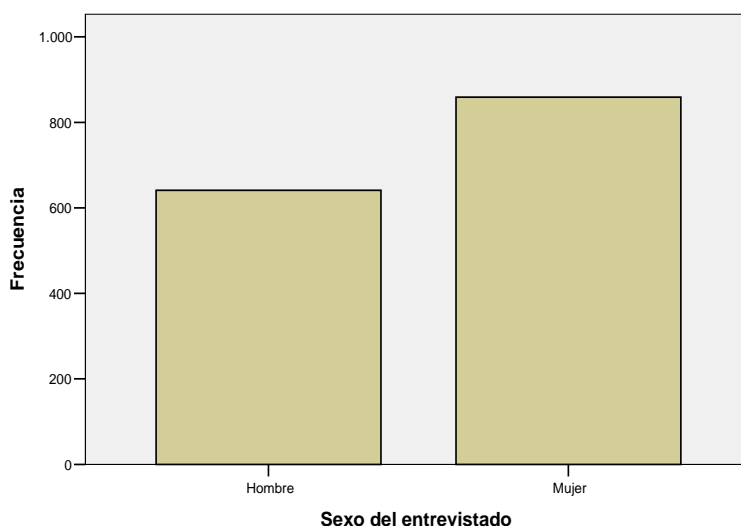
Sexo del entrevistado

N	Válidos	1500
	Perdidos	0
Moda		2

sexo Sexo del entrevistado

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Hombre	641	42,7	42,7	42,7
	2 Mujer	859	57,3	57,3	100,0
	Total	1500	100,0	100,0	

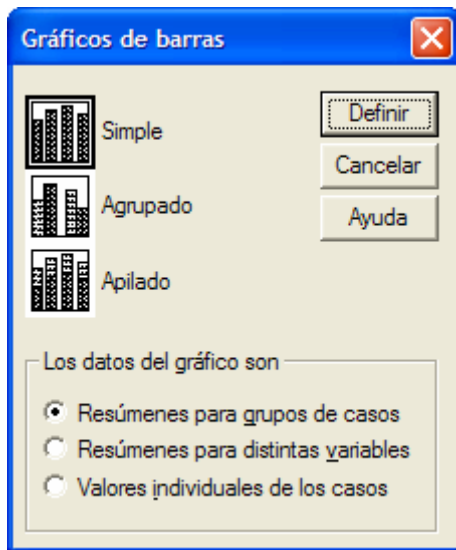
Sexo del entrevistado



- Les freqüències fan referència al total de casos (1500) i no hi ha valors perduts.
 - La distribució de freqüències ens dóna que hi ha més dones que homes a la mostra i això es tradueix gràficament en la més gran alçada de la barra corresponent a les dones. La moda és 2.
- A través del menú “**Gráficos**” es pot obtenir igualment el **gràfic de barres** (o el de sectors):

Gráficos / Barras...
Gráficos / Sectores...

Si demanem un diagrama de barres, ens apareix el quadre següent:



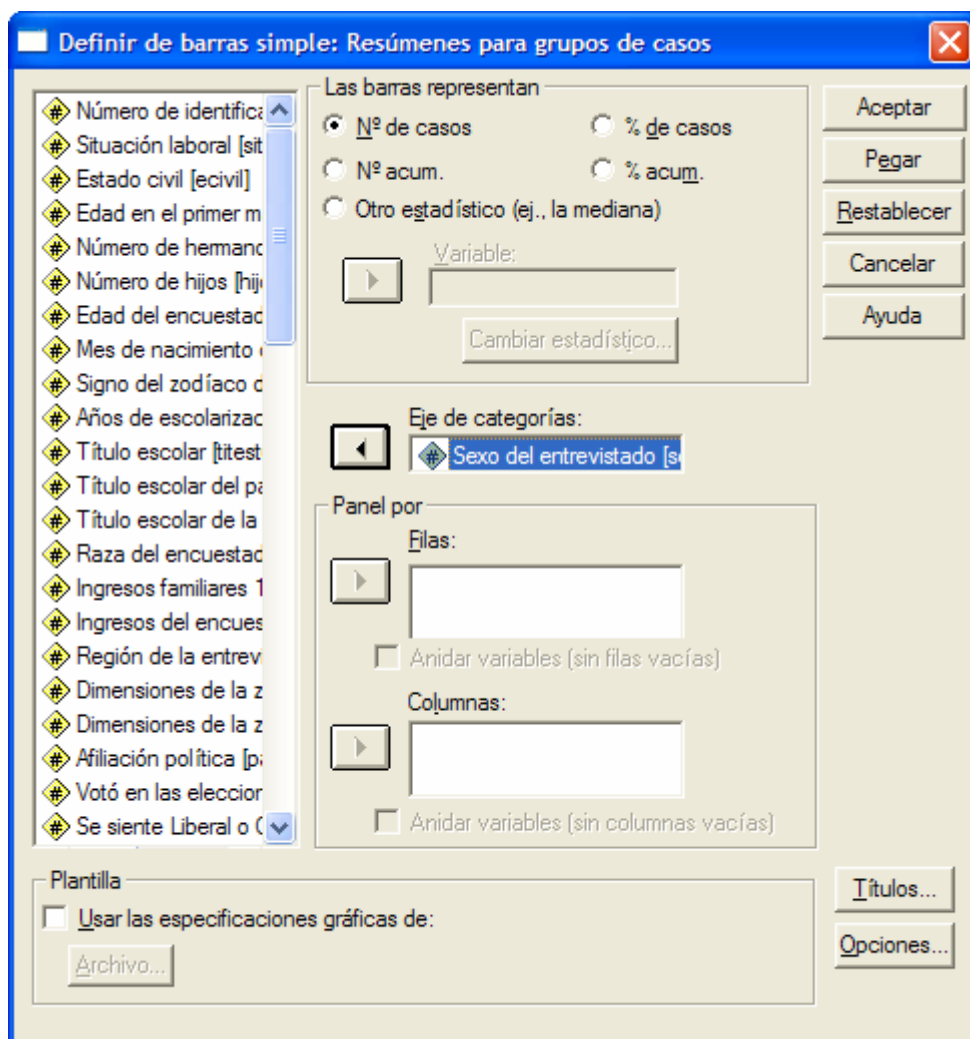
Optem per:

- “Simple”: una sola barra per a cada categoria, cas o variable, de l'eix de categories.
- “Resúmenes para grupos de casos”: el gràfic resumeix una única variable dins dels subgrups definits per una variable categòrica.

Són les opcions per defecte.

A continuació cliquem sobre “Definir”.

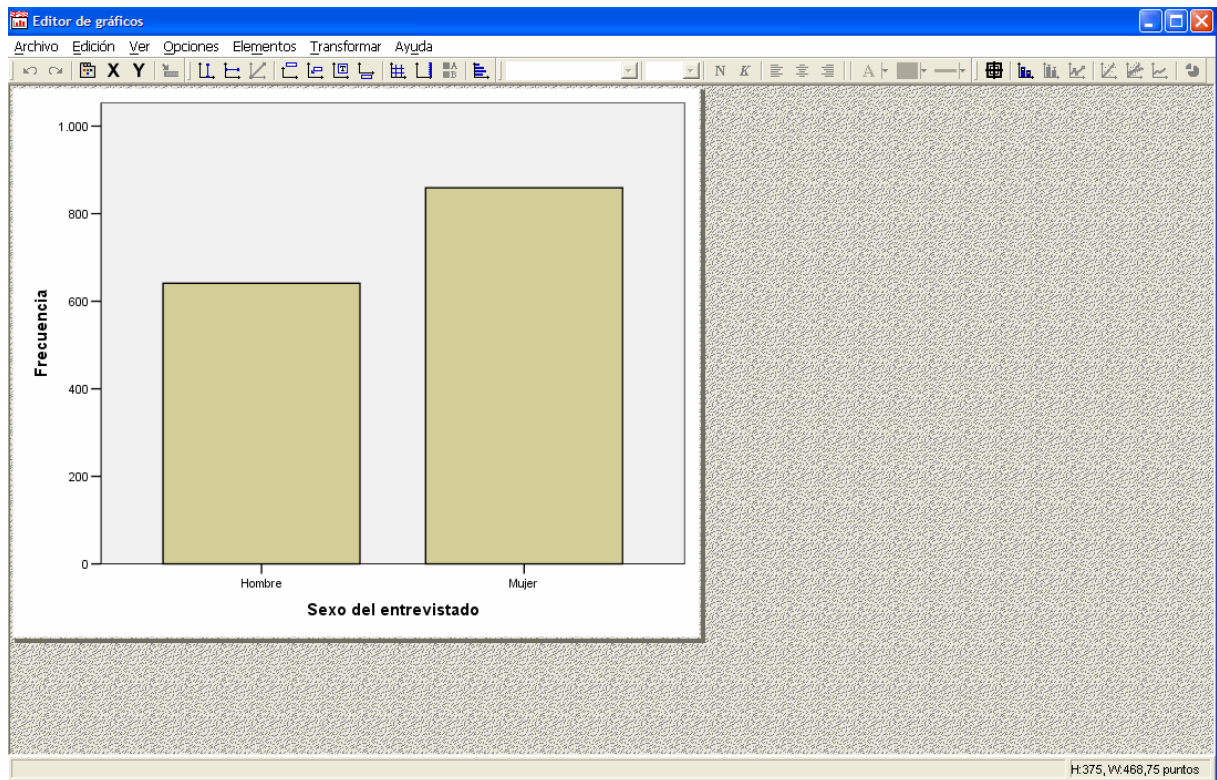
Al quadre de diàleg col·locarem la variable **sexo** com a “Eje de categorías”:



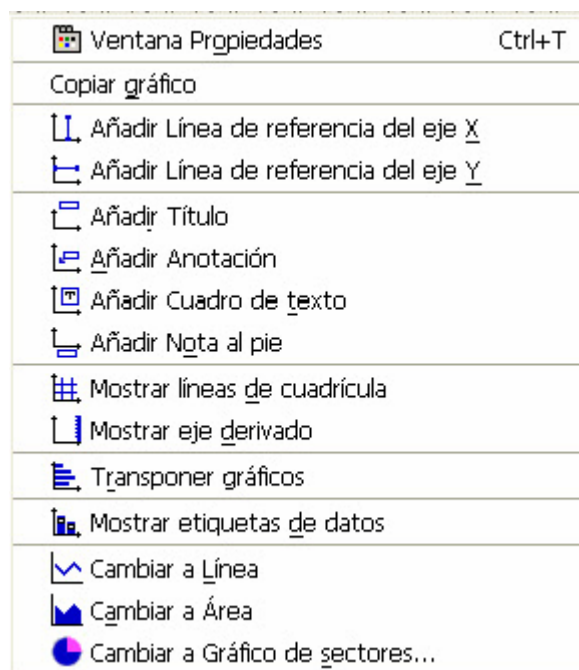
I clicarem sobre “Aceptar”. Obtenim el mateix gràfic de barres anterior.

► Ara procedirem a editar el gràfic barres per tal de modificar la seva aparença a través de la configuració d'algunes opcions o propietats del mateix.

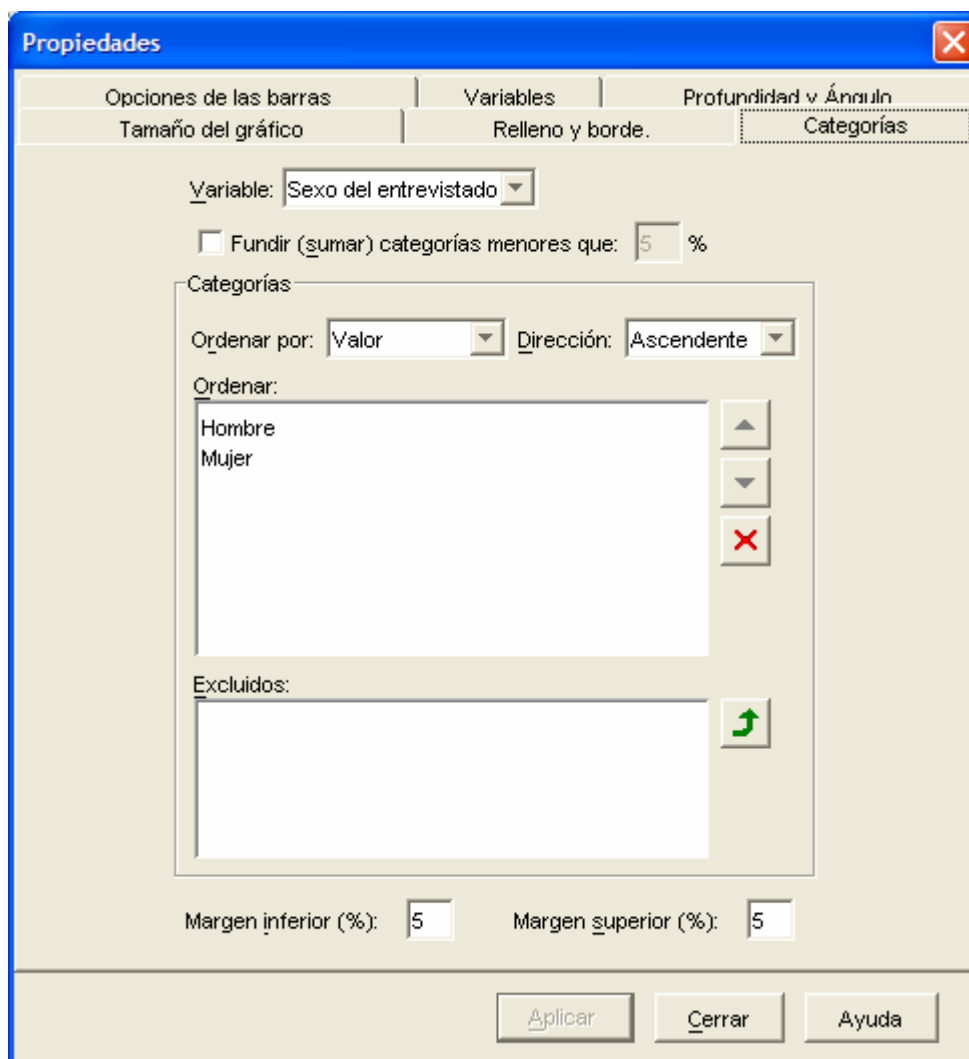
- Per a editar-lo fem doble-clic sobre el gràfic a l'editor de resultats. Se'ns obre una nova finestra, la de l'"**Editor de gràficos**".



- El primer que farem és clicar el botó dret del ratolí per a desplegar el menú contextual i clicarem sobre "Ventana de propiedades" (Ctrl+T)



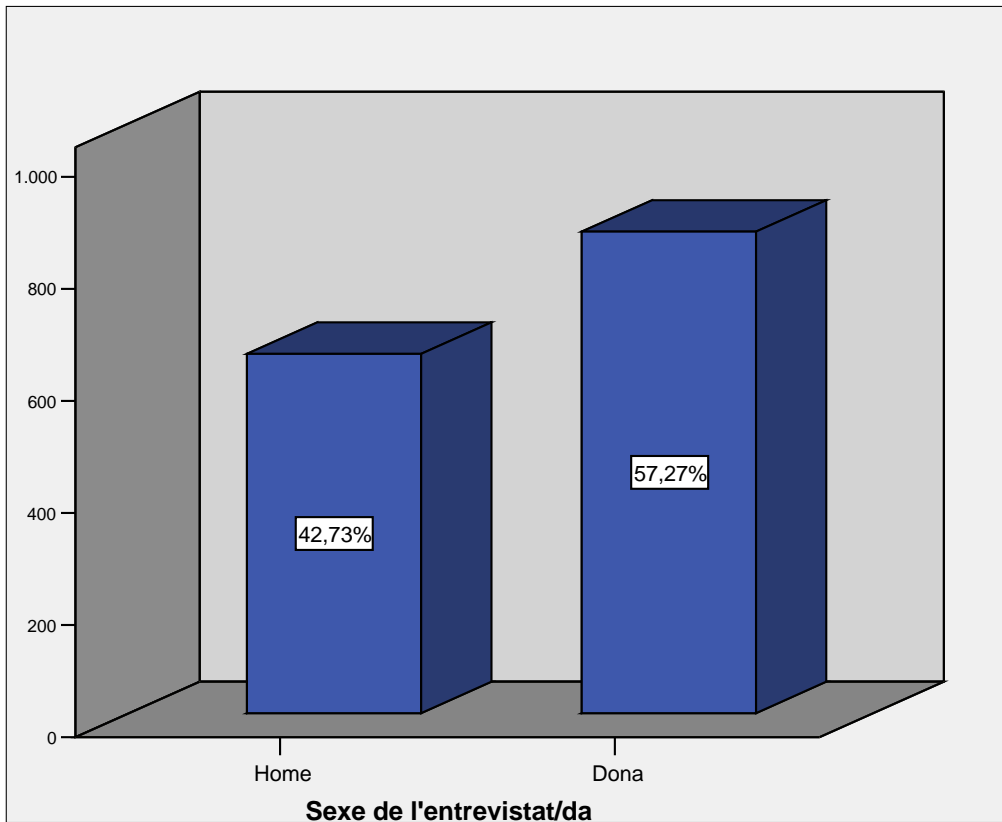
Ens apareix una finestra amb les diverses opcions d'edició del gràfic. Aquestes opcions canvien en funció de la part del gràfic que se selecciona. Si cliquem una vegada sobre les barres del gràfic aquestes es mostren seleccionades i la finestra de propietats té l'aparença següent:



- Realitzarem els canvis següents:
 - Al menú contextual escollirem “Mostrar etiquetas de datos”. Se superposen sobre les barres les freqüències absolutes de cada categoria. Podem optar per afegir les etiquetes dels valors de la variable o els percentatges de cada categoria.
 - Podem afegir un títol. Al menú contextual escollirem “Añadir título”. I escriurem, per exemple, “Gràfic de barres de la variable sexe”.
 - Podem canviar les etiquetes de la variable i dels valors, per exemple, per a escriure’ls en català, tot clicant sobre el text.
 - Si seleccionem les barres, a les propietats, podem canviar el tipus de lletra, els colors, l’amplada de les barres, la profunditat,...

Després d’aplicar alguns d’aquests canvis el gràfic adopta, per exemple, aquest aspecte:

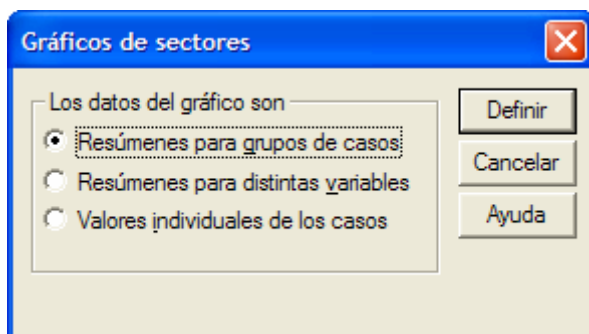
Gràfic de barres de la variable sexe



► A través del menú “**Gràfics**” ara demanarem un **gràfic de sectors**:

Gràfics / Sectores...

Ens apareix el quadre següent:

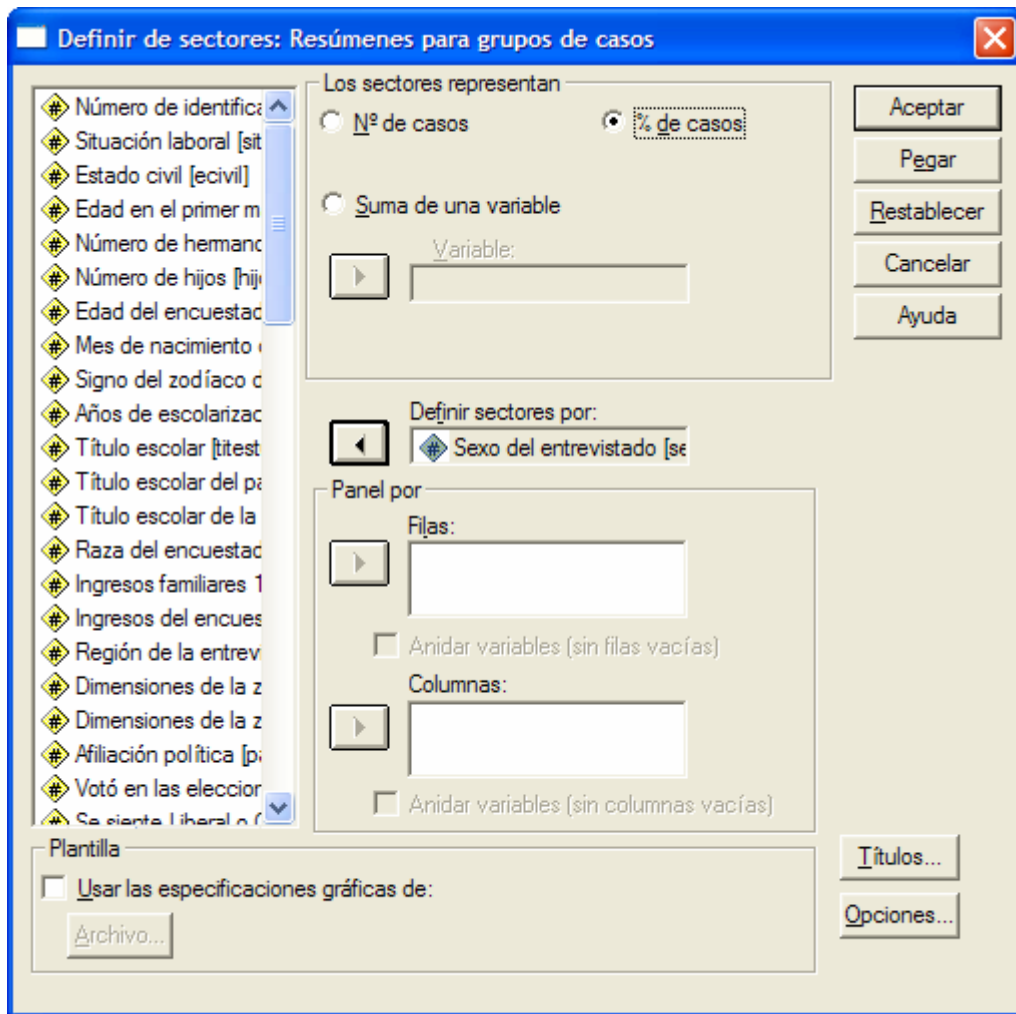


Optem per:

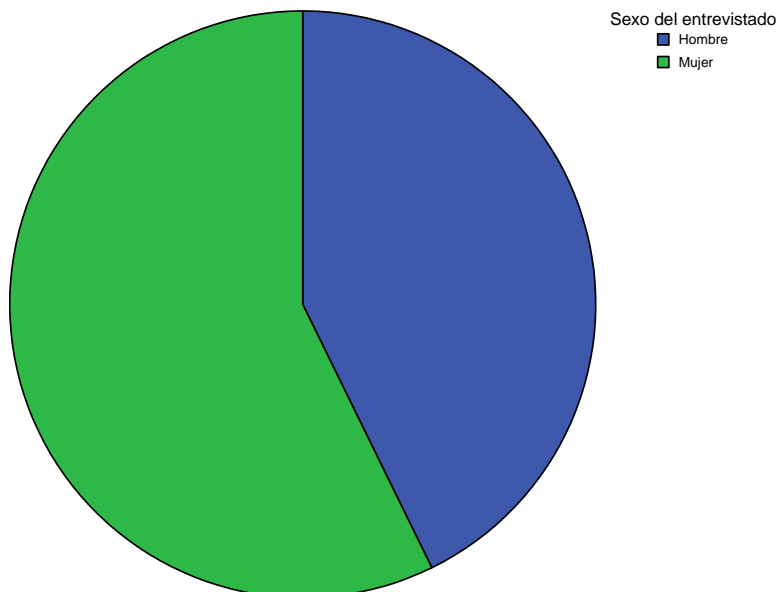
- “Resúmenes para grupos de casos”. Com abans, el gràfic resumeix una única variable dins dels subgrups definits per una variable categòrica.

I cliquem sobre “Definir”.

Al quadre de diàleg col·locarem la variable **sexo** a “**Definir sectores por:**” i marcarem l’opció “**% de casos**”:



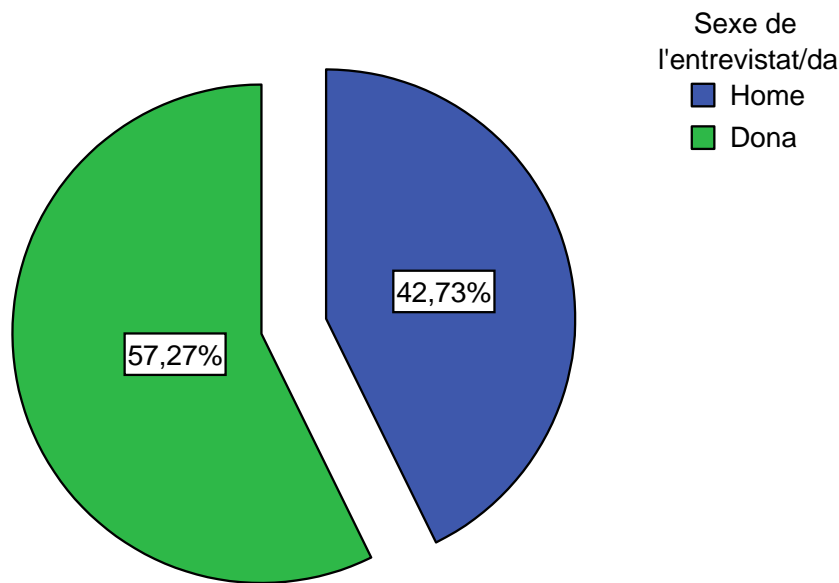
En clicar sobre “Aceptar” obtenim el resultat següent (el mateix que hauríem obtingut si l’haguéssim demanat al procediment “Frecuencias”):



- Podeu editar el gràfic per a realitzar els canvis següents:
 - Al menú contextual escollim “Mostrar etiquetas de datos”. Es superposen sobre els sectors els percentatges de cada categoria.
 - Podem afegir un títol. Al menú contextual escollirem “Añadir título”. I escriurem, per exemple, “Gràfic de sectors de la variable sexe”.
 - Podem canviar les etiquetes de la variable i dels valors, per exemple, per a escriure’ls en català, tot clicant sobre el text.
 - Podem triar l’opció “Desgajar sector” del menú contextual.

El gràfic adopta l’aparença següent:

Gràfic de sectors de la variable sexe

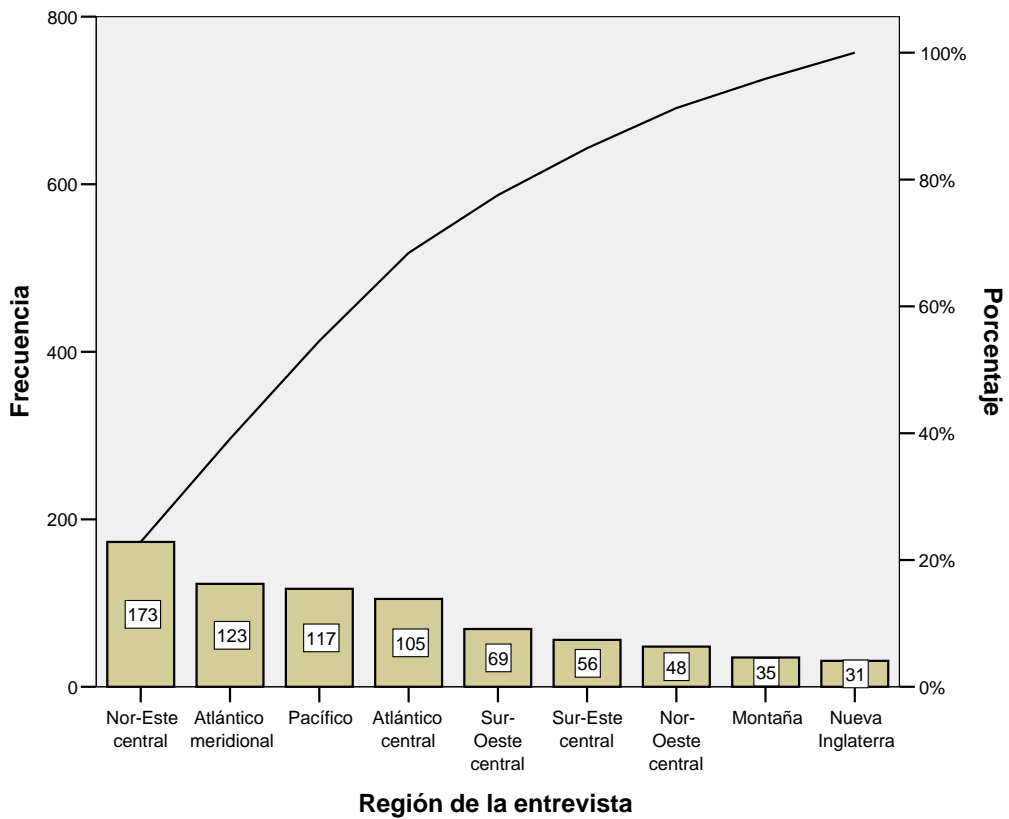
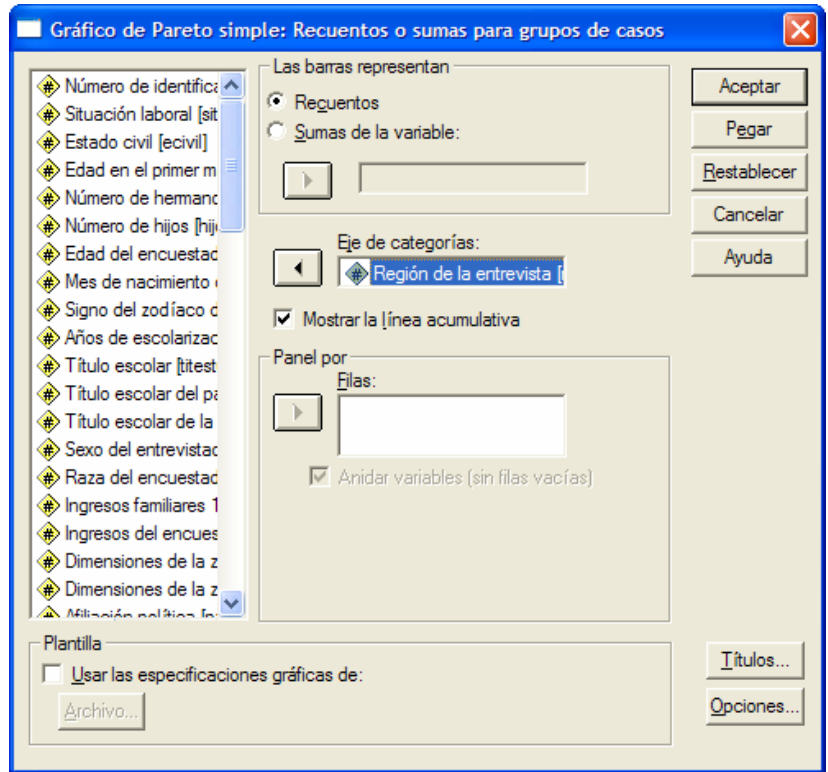
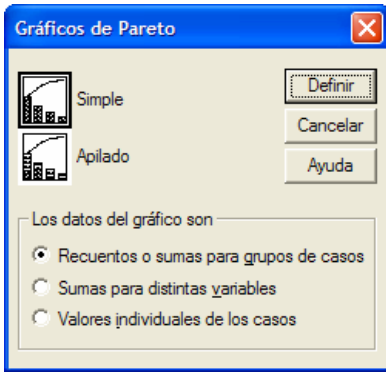


► **Exercici:**

Repetiu el mateix per a altres variables qualitatives nominals: *raza, ecivil, partido, relig, eutan,...*

► Una altre tipus de representació gràfica que es pot fer servir és l’anomenat **Gràfic de Pareto**. És un gràfic que es basa en l’anomenat principi de Pareto: “*pocs vitals, molts trivials*”, que es pot expressar dient que “el 80% dels problemes es resolen amb el 20% de les causes”. Amb el gràfic per tant es tracta de posar de manifest la concentració de la distribució de les freqüències en uns pocs valors. El gràfic de fet és un gràfic de barres que s’ordenen per ordre de freqüència dels valors de la variable. El gràfic incorpora a més una línia de les freqüències acumulades.

Demaneu el gràfic a través del menú “**Gráficos / Pareto...**” per a la variable **región**.



2. Anàlisi descriptiva de variables qualitatives ordinals

Considerarem la variable qualitativa ordinal **píldora** resultat de preguntar en el qüestionari de l'enquesta "Si està d'acord o en desacord amb l'ús de la píldora com a sistema de control de la natalitat per als adolescents (14-16)".

És una variable codificada de tipus "**numèrica**" amb valors 1 a 4, segons el grau d'acord, i amb valors 0 (per als no pertinents), 8 (per als "no sap") i 9 (per als "no contesta").

► A través del menú **Analizar / Estadísticos descriptivos / Frecuencias...** demanarem: (1) la taula de distribució de freqüències, (2) un gràfic de barres i (3) l'estadístic de la moda i la mediana.

(1) Seleccionem la variable **píldora** i la col·loquem al requadre de "Variables". D'aquesta manera obtenim la taula de distribucions de freqüències. L'opció "Mostrar tablas de frecuencias" està activada.

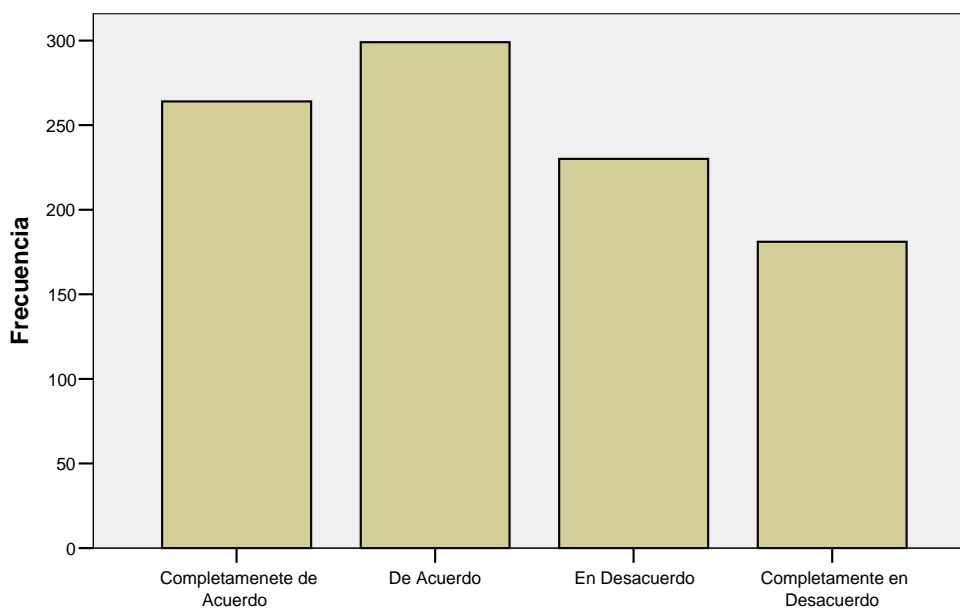
(2) Per a demanar el gràfic de barres cliquem sobre el botó "Gráficos...". Ens apareix el quadre de diàleg on seleccionarem "**Gráfico de barras**". I cliquem sobre "Continuar".

(3) Per a demanar l'estadístic cliquem sobre el botó "Estadísticos...". Ens apareix el quadre de diàleg on seleccionarem la **Moda** i la **Mediana** entre les opcions de mesures de tendència central. I cliquem sobre "Continuar". Finalment cliquem "Aceptar" al quadre de diàleg principal i observem els resultats següents:

Estadísticos		
píldora Control de natalidad para los adolescentes (14-16)		
N	Válidos	974
	Perdidos	526
Mediana		2,00
Moda		2

píldora Control de natalidad para los adolescentes (14-16)					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Completamenete de Acuerdo	264	17,6	27,1	27,1
	2 De Acuerdo	299	19,9	30,7	57,8
	3 En Desacuerdo	230	15,3	23,6	81,4
	4 Completamente en Desacuerdo	181	12,1	18,6	100,0
	Total	974	64,9	100,0	
Perdidos	0 No Procede	485	32,3		
	8 No Sabe	34	2,3		
	9 No Contesta	7	,5		
	Total	526	35,1		
Total		1500	100,0		

Control de natalidad para los adolescentes (14-16)



Control de natalidad para los adolescentes (14-16)

- Les freqüències reflecteixen la presència dels valors perduts (0, 8 i 9), són codis assignats a una manca d'informació en relació a la pregunta sobre el control de natalitat i s'han declarat com a valors perduts de l'usuari. En total són 526 casos, i les casos vàlids sobre els qual s'han de calcular els percentatges són 974.
- Fixeu-vos per tant que l'anàlisi correcte es fa sense els valors perduts (columna "Porcentaje válido"). Si considerem la columna "Porcentaje", la distribució de freqüències amb els valors perduts, llavors la variable deixa de ser ordinal, i cal considerar-la nominal.
- Veiem com el valor més freqüent és el que correspon a "De acuerdo" (299 respostes, 19,9% dels casos). Per tant, la moda correspon al valor 2.
- La mediana, el valor que acumula el 50% dels casos, és el valor 2. El 57,8% de les persones enquestades està "completament d'acord" o "d'acord" amb el control de natalitat, és a dir, més de la meitat dels enquestats està a favor de l'ús de la píldora.

► Exercici:

Podeu repetir l'anàlisi amb altre variables ordinals com: *penacap*, *cachete*, *periòdic*, *blues*, *clásica*, *jazz*,...

► Per últim, reiterarem una anàlisi com l'anterior amb la variable ordinal **ingper91** (els ingressos de l'enquestat al 1991), però a més calcularem diversos **percentils**. En aquest cas disposem de valors ordinals que identifiquen diferents intervals d'ingressos. Si bé la variable ingressos podria ser numèrica o ser tractada com a variable numèrica a partir dels intervals, aquí no obstant està codificada des de l'1 fins el 21, i l'única propietat de mesura que contenen aquests números és la de l'ordre.

Frecuencias

ingper91 Ingresos del encuestado 1991

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Menos de \$1000	26	1,7	2,6	2,6
	2 \$1000-2999	36	2,4	3,6	6,2
	3 \$3000-3999	30	2,0	3,0	9,3
	4 \$4000-4999	24	1,6	2,4	11,7
	5 \$5000-5999	23	1,5	2,3	14,0
	6 \$6000-6999	23	1,5	2,3	16,3
	7 \$7000-7999	15	1,0	1,5	17,8
	8 \$8000-9999	31	2,1	3,1	20,9
	9 \$10000-12499	55	3,7	5,5	26,5
	10 \$12500-14999	54	3,6	5,4	31,9
	11 \$15000-17499	64	4,3	6,4	38,3
	12 \$17500-19999	58	3,9	5,8	44,2
	13 \$20000-22499	55	3,7	5,5	49,7
	14 \$22500-24999	61	4,1	6,1	55,8
	15 \$25000-29999	84	5,6	8,5	64,3
	16 \$30000-34999	83	5,5	8,4	72,6
	17 \$35000-39999	54	3,6	5,4	78,1
	18 \$40000-49999	66	4,4	6,6	84,7
	19 \$50000-59999	38	2,5	3,8	88,5
	20 \$60000-74999	23	1,5	2,3	90,8
	21 \$75000+	44	2,9	4,4	95,3
	22 Rehusa	47	3,1	4,7	100,0
	Total	994	66,3	100,0	
Perdidos	0 No procede	493	32,9		
	98 No sabe	9	,6		
	99 No contesta	4	,3		
	Total	506	33,7		
Total		1500	100,0		

- Quina és la moda? _____
- Quina és la mediana? _____

Quan disposem d'un nombre elevat de valors de la variable ordinal, com en aquest cas, llavors té sentit calcular diversos percentils per tal de descriure el comportament de la distribució. Per a la variable *pildora* anterior no tenia sentit demanar aquesta informació.

Si demanem en particular els **quartils** a través del procediment de “Frecuencias” obtenim els resultats que apareixen a la taula següent:

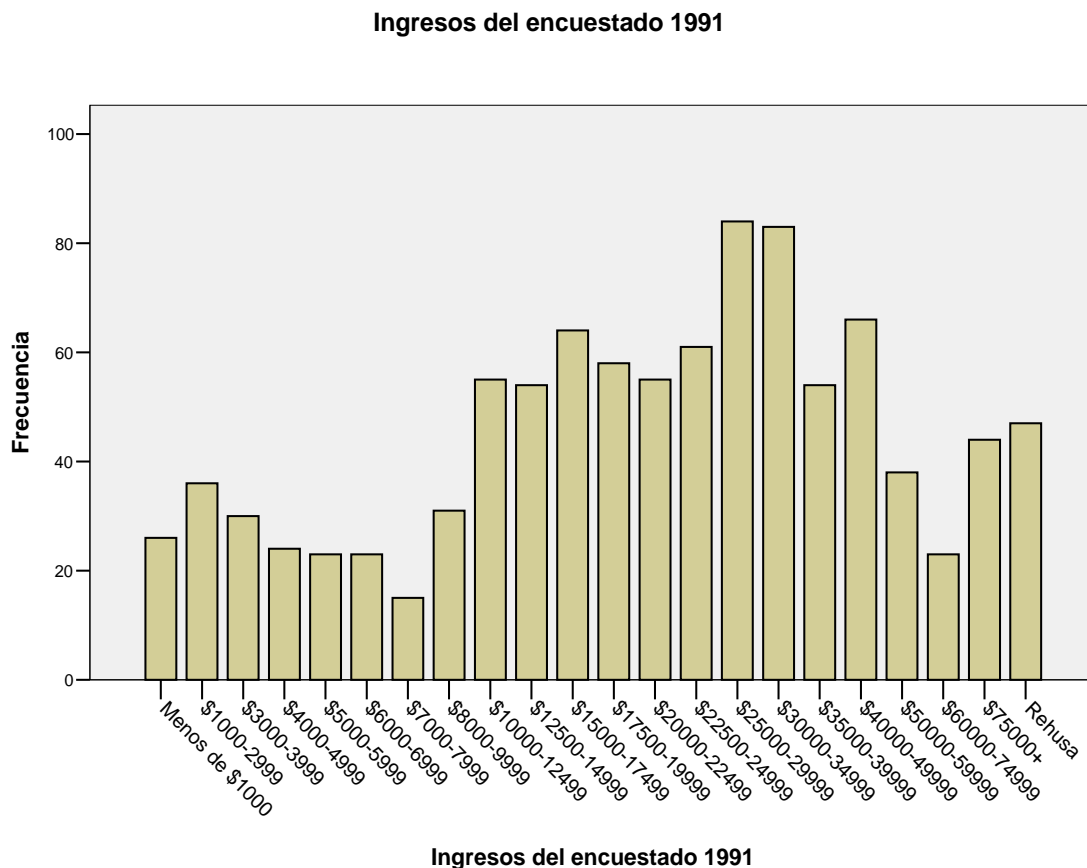
Estadísticos

ingper91 Ingresos del encuestado 1991		
N	Válidos	994
	Perdidos	506
Percentiles	25	9,00
	50	14,00
	75	17,00

Comproveu aquests resultats a la taula de freqüències.

- **Quin és el percentil 30?** _____
- **Quin és el percentil 60?** _____
- **Quin és el percentil 14?** _____

El diagrama de barres que s’obté és el següent:



► **Exercici:**

Podeu repetir l’anàlisi amb la variable **ingfam91** (els ingressos familiars totals l’any 1991).