

Pràctica 5. Dades ponderades. Transformar i recodificar variables.

A la pràctica anterior vam començar a veure possibles manipulacions d'un arxiu de dades. En aquesta pràctica en veurem algunes més: començarem veient com podem ponderar els individus si, per a l'estudi que ens interessa realitzar, cal que tinguin pesos diferents. Continuarem veient com podem canviar la codificació que tenim d'una variable. També veurem com crear intervals o classes de valors utilitzant un procediment dirigit que l'SPSS anomena "categoritzador visual". Finalment, a l'apartat de transformació de variables mitjançant funcions, veurem com podem generar noves variables calculades a partir de les que tenim. Un cas particular és la tipificació de dades.

► Recordeu activar les opcions d'edició:

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

► En aquesta pràctica treballarem amb els arxius **GSS93 reducido.sav**, **Mundo 95.sav** i **Datos de empleados.sav** del programari, ubicats tots tres a la carpeta de l'SPSS.

1. Dades ponderades

Veurem a continuació algunes de les aplicacions de l'opció **ponderar casos**. Concretament aquesta opció ens permetrà, entre altres coses, obtenir resultats ponderats i introduir dades ja tabulades a partir d'una distribució de freqüències. A la pràctica 7 veureu que aquesta aplicació també ens permet introduir taules de contingència.

Com el seu nom indica, ponderar vol dir donar pesos diferents als individus o casos d'un arxiu de dades. Els pesos vindran donats per una columna de la base de dades, és a dir, tindrem una variable que ens dirà el pes que li correspon a cada individu. Inicialment els pes de cada individu és d'1, quan ponderem el que fem és multiplicar aquest 1 de cada cas pel valor que tingui a la variable de ponderació.

1.1 Obtenir resultats ponderats

► Obriu l'arxiu **Mundo 95.sav**. Recordeu que en aquest arxiu els individus són països.

Quan treballem amb aquest arxiu, si no ponderem els casos tots els països tenen el mateix pes, independentment de la seva població, superfície, etc... De vegades ja ens interessarà treballar amb l'arxiu d'aquesta manera, però si nosaltres volem analitzar, per exemple, quina és l'**esperança de vida femenina** mundial, no podem donar el mateix pes a Bèlgica (que és l'individu 12 de la base de dades) que a la Xina (individu 25). En aquest cas, com que és una variable que fa referència a la població dels països, seria convenient donar a cada país un pes diferent segons la seva **població**. La situació ideal seria donar-li un pes proporcional al nombre de dones que habiten al país, però no disposem d'aquesta informació.

Comencem calculant la mitjana de la variable **espvidaf** (esperança de vida femenina) sense ponderar els casos. Obtenim el següent resultat:

Estadísticos descriptivos

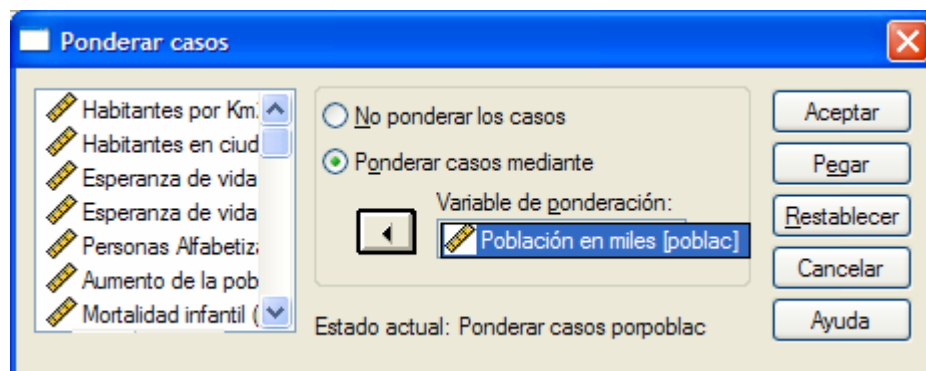
	N	Media
Esperanza de vida femenina	109	70,16
N válido (según lista)	109	

Aquesta és una mitjana on els individus són països. A partir de l'esperança de vida de cada país hem calculat la mitjana donant el mateix pes a tots els països. A Bèlgica (que té una esperança de vida femenina de 79 anys) li hem donat el mateix pes que a la Xina (que té una esperança de vida femenina de 69 anys), tot i que Bèlgica té una població de poc més de 10 milions d'habitants i la Xina 1 205 milions d'habitants.

Aquesta no és doncs l'esperança de vida femenina mundial. Per estimar-la hem de donar a cada país un pes proporcional a la seva població. El menú de l'SPSS per fer-ho és el següent:

Datos / Ponderar casos / Ponderar casos mediante

i escollim la variable que ens donarà el pes que li correspon a cada individu. En aquest cas escollim la variable **poblac** que ens dóna la població de cada país. (Si bé no és la població de dones, proporcionalment serà una bona aproximació, ja que el nombre d'homes i de dones sol ser molt similar).



Un cop fer això, l'únic canvi que observem a la finestra de l'SPSS, és que a la part inferior dreta de la finestra apareix l'etiqueta "Ponderado".

Si ara tornem a calcular la mitjana de la variable **espvidaf** obtenim els resultats següents:

Estadísticos descriptivos		
	N	Media
Esperanza de vida femenina	5201903	67,86
N válido (según lista)	5201903	

Observeu que la mitjana ara ha baixat de 70.16 a 67.86. Quina interpretació doneu a aquest resultat? Observeu també que ara el nombre d'observacions ha canviat, tenim 5 201 903 individus (ja que els individus són milers d'habitants) enlloc dels 109 individus que teníem abans (quan els individus eren països).

Aquest resultat aproxima molt millor l'esperança de vida femenina mundial.

És **molt important** que un cop haguem realitzat l'estudi que volíem fer ponderant els casos ens recordem de **desfer la ponderació**. Si no ho féssim estaríem obtenint informacions errònies. Per fer-ho, el menú és

Datos / Ponderar casos / No ponderar los casos

► Exercici 1

Feu un diagrama de sectors de la variable **relig** (religió majoritària) sense ponderar i ponderant segons la població de cada país. Compareu els resultats obtinguts.

1.2 Introducció de dades ja tabulades

Observeu la següent taula extreta del Baròmetre d'opinió política realitzat per l'IDESCAT el juny de 2005 i publicat per la Generalitat de Catalunya.

20. Suposant que demà se celebressin eleccions al Parlament de Catalunya, a quin partit o coalició votaria?
(NOMÉS UNA RESPOSTA)

	Freqüències
PP	80
CiU	320
ERC	252
PSC	403
ICV/EUIA	160
Altres partits	8
En blanc	53
No votaria	204
No té la nacionalitat per poder votar	17
No ho sap	259
No contesta	144
Total	1900

S'ha preguntat a 1900 individus quin partit votarien si se celebrassin eleccions nacionals. Imaginem que nosaltres volem treballar amb aquestes dades, per exemple volem fer un diagrama de sectors sobre la variable **intenció de vot** amb l'SPSS. Una opció seria introduir el valor de la variable **intenció de vot** dels 1900 individus un per un, però és molt més senzill introduir dues variables: una amb els diferents partits i una altra amb els pesos, és a dir, amb el nombre d'individus que han afirmat que votarien aquell partit i ponderar els casos segons aquesta variable.

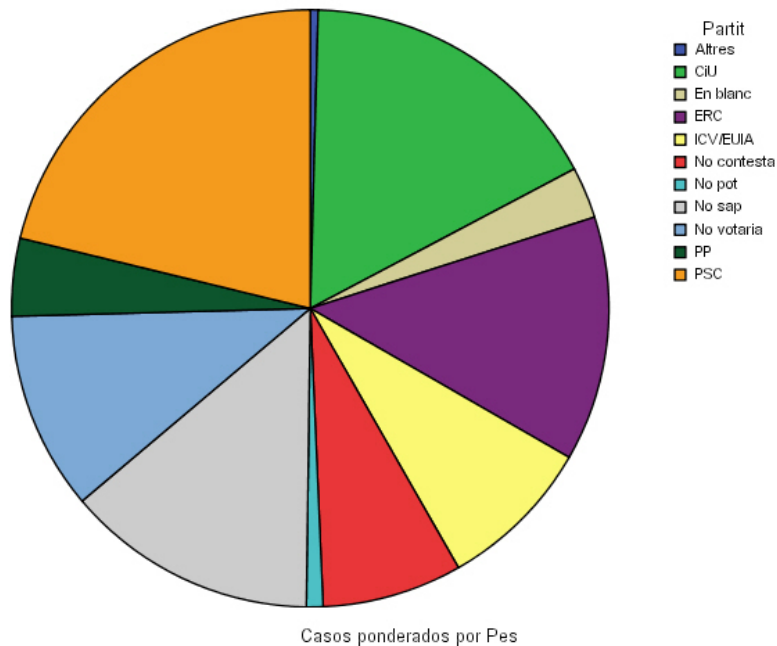
Obriu un fitxer nou amb l'SPSS i creeu dues variables:

- una variable que anomenareu **partit** amb aquestes 11 opcions (PP, CIU, ERC, PSC, ICV/EUIA, Altres, En blanc, No votaria, No pot, No sap, No contesta), i
- una altra que anomenarem **pes** amb la freqüència d'individus que correspon a cada valor de la variable **partit**. Un cop fet això pondereu casos segons la variable **pes**.

	Partit	Pes	var	var	var	var	var	var
1	PP	80						
2	CiU	320						
3	ERC	252						
4	PSC	403						
5	ICV/EUiA	160						
6	Altres	8						
7	En blanc	53						
8	No votaria	204						
9	No pot	17						
10	No sap	259						
11	No contesta	144						
12								

Construiu una taula de freqüències de la variable **partit**, observeu que hem recuperat la taula original (potser amb un ordre diferent dels valors depenent de com els haguem entrat). Observeu també que estem treballant amb un total de 1 900 individus, tot i que només hem omplert 11 files a la base de dades.

Si fem un diagrama de sectors de la variable **partit** obtenim el següent gràfic



Observeu que, per defecte, l'SPSS ens avisa que estem treballant amb dades ponderades per la variable **pes**.

2. Recodificar variables

Les opcions de recodificar i transformar ens permetran obtenir noves variables a partir de les ja existents. L'SPSS ens permet dos tipus de recodificacions: la recodificació automàtica i la no automàtica.

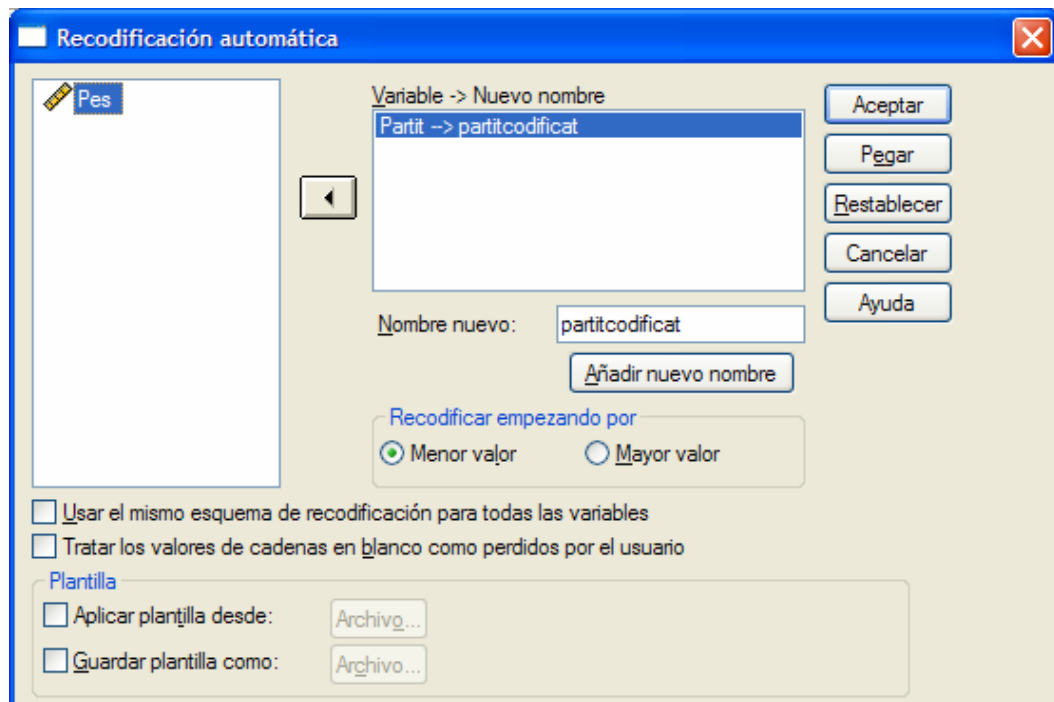
2.1 Recodificació automàtica de variables

Ens crea una nova variable amb el nom que escollim nosaltres i amb valors enters consecutius a partir d'una variable existent de tipus cadena o numèrica. A l'arxiu de resultats ens mostra la correspondència entre la codificació dels valors de la variable original (cadena o numèrics) i la dels valors numèrics de la nova variable amb les etiquetes creades automàticament. El menú és el següent:

Transformar / Recodificación automática / ...

hem d'escollir la variable que volem recodificar i el nom de la nova variable a la pestanya *Nom nou* i premem *Afegir nou nom* i *Acceptar*.

- Per exemple, volem recodificar la variable **partit** de l'arxiu que hem creat, i crear una nova variable que es dirà **partitcodificat**:



A l'arxiu de resultats hem obtingut:

```
Partit into partitcodificat
Old Value      New Value  Value Label
-----
Altres          1      Altres
CiU              2      CiU
En blanc        3      En blanc
ERC             4      ERC
ICV/EUIA        5      ICV/EUIA
No contesta     6      No contesta
No pot          7      No pot
No sap          8      No sap
No votaria      9      No votaria
PP             10     PP
PSC             11     PSC
```

Observem que ens ha creat una nova variable **partitcodificat**, que ens ha afegit al final de la base de dades, i si mirem a la *vista de variables* l'equivalència entre els valors i les etiquetes és la que ens ha mostrat a l'arxiu de resultats.

Observeu també que no li hem dit en cap moment quina era l'etiqueta que volíem donar a cada valor i l'SPSS ha decidit fer-ho per ordre alfabètic dels valors de la variable. No podem escollir la codificació ja que es fa de forma automàtica. Amb aquesta codificació numèrica és fàcil definir valors perduts: per exemple del 6 al 8. També podríem considerar perduts, a efectes de repartiment d'escons, el 3 i el 9.

► Exercici 2

Obriu l'arxiu **Mundo 95.sav** i feu una recodificació automàtica de la variable **relig** (religió majoritària) en una nova variable **relnum**.

2.2 Recodificació no automàtica de variables

Aquest tipus de recodificació permet escollir la codificació a l'usuari. L'SPSS ens permet escollir entre dues opcions: recodificar *en les mateixes variables* o bé *en variables diferents*. En general sempre és millor fer-ho en diferents variables per no perdre les dades originals.

Aquesta opció, a més a més de recodificar variables com en l'exemple anterior, ens permet també agrupar valors en classes o categories.

► Suposem que ens interessa classificar les religions en tres tipus (cristianes, musulmanes i altres) amb les etiquetes següents:

- **1= cristianes (catòlica, ortodoxa, protestant)**
- **2= musulmana**
- **3= altres**

Utilitzarem la variable **relnum** que hem generat abans. Les etiquetes que ens havia creat són:

RELIG	RELNUM	Religi3n mayoritaria
Old Value	New Value	Value Label
Animista	1	Animista
Budista	2	Budista
Cat3lica	3	Cat3lica
Ind3	4	Ind3
Jud3a	5	Jud3a
Musulma.	6	Musulma.
Ortodoxa	7	Ortodoxa
Protest.	8	Protest.
Taoista	9	Taoista
Tribal	10	Tribal

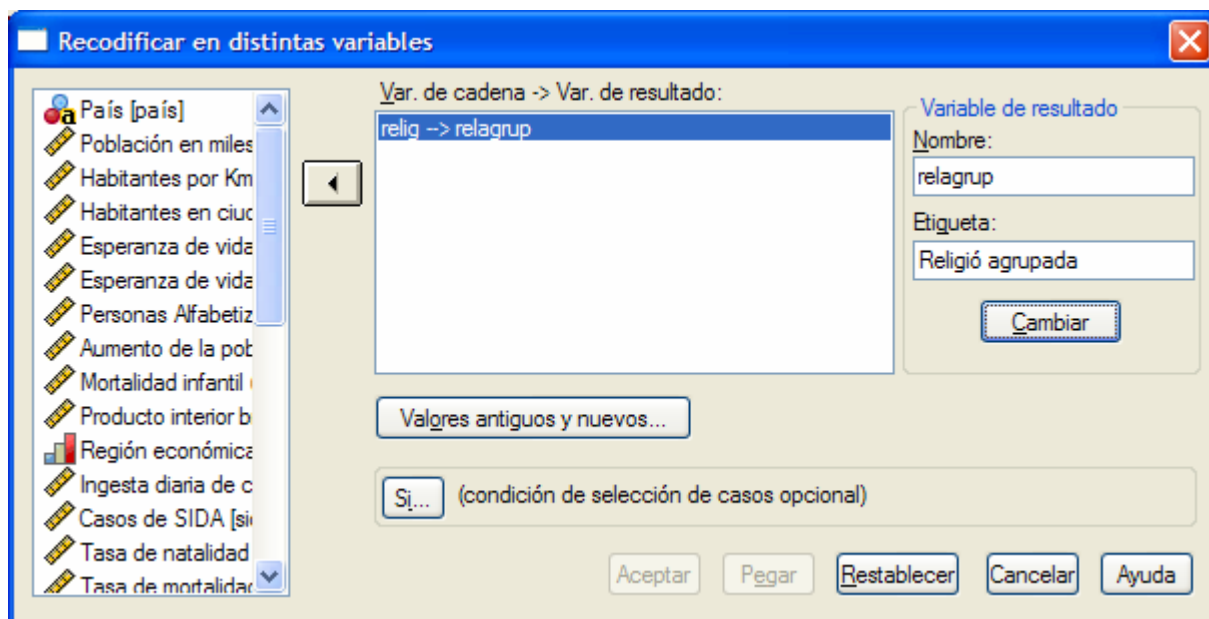
El criteri de recodificació serà per tant:

- **3, 7, 8 =1**
- **6 = 2**
- **1, 2, 4, 5, 9, 10 = 3**

Podem procedir de la següent forma:

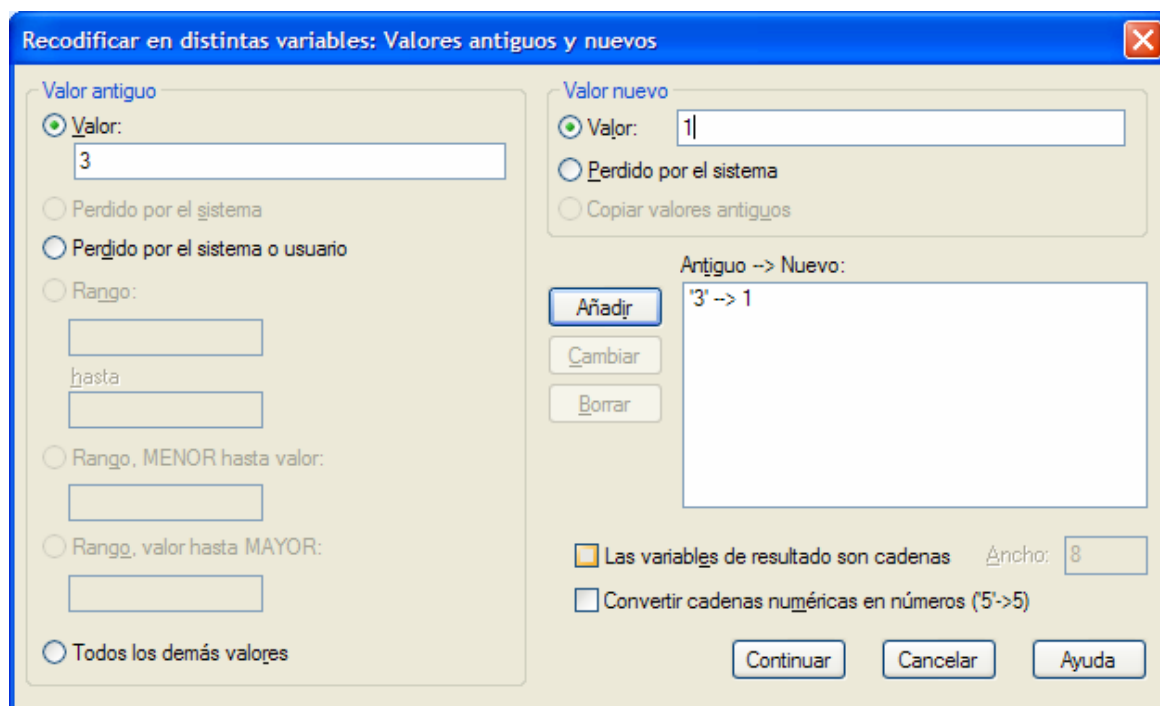
Transformar / Recodificar / En distintes variables

Escollim com a variable d'entrada la variable **relnum** i anomenem, per exemple, **relagrup** a la variable de sortida. La podem etiquetar, per exemple, amb "Religió agrupada".



Premem l'opció *Valors antics i nous* i a cada valor de la variable **relnum** (*valor antic*) li assignem el que li correspon a la nova variable **relagrup** (*valor nou*).

Per exemple, al *valor antic* 3 (religió catòlica) li correspon el *valor nou* 1 (religions cristianes) i premem "Añadir".



Continuem fent correspondre als *valors antics* 7 i 8 el *valor nou* 1, al *valor antic* 6 el *valor nou* 2, al *valor antic* Perdut pel sistema o usuari el *valor nou* Perdut pel sistema i finalment al *valor antic* Tota la resta de valors el *valor nou* 3.

Ens quedarà el quadre següent:

Recodificar en distintas variables: Valores antiguos y nuevos

Valor antiguo

☐ Valor:

☐ Perdido por el sistema

☒ Perdido por el sistema o usuario

☐ Rango:

hasta

☐ Rango, MENOR hasta valor:

☐ Rango, valor hasta MAYOR:

☐ Todos los demás valores

Valor nuevo

☐ Valor:

☒ Perdido por el sistema:

☐ Copiar valores antiguos

Antiguo -> Nuevo:

Añadir

Cambiar

Borrar

'3' -> 1
'7' -> 1
'8' -> 1
'6' -> 2
MISSING -> SYSMIS
ELSE -> 3

☐ Las variables de resultado son cadenas Ancho: 8

☐ Convertir cadenas numéricas en números ('5' -> 5)

Continuar Cancelar Ayuda

I premem *Continuar* i *Acceptar*. Ara la matriu de dades contindrà una nova columna, l'última, amb la variable que acabem de crear.

A continuació hem de posar les etiquetes a la nova variable **relagrup**. També li podem donar un format a la variable sense decimals.

Obtindrem el resultat següent:

relagrup Religió agrupada

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Cristianes	1671493	32,1	32,4	32,4
	2 Musulmana	986407	19,0	19,1	51,5
	3 Altres	2500103	48,1	48,5	100,0
	Total	5158003	99,2	100,0	
Perdidos	Sistema	43900	,8		
Total		5201903	100,0		

► Exercici 3

Podeu comprovar que l'agrupació s'ha fet correctament obtenint les freqüències d'ambdues variables i comprant-les. Feu un diagrama de barres de la nova variable **relagrup**.

► Veurem a continuació una altra utilitat de la recodificació. Imaginem ara que tenim una variable numèrica i volem convertir-la en una nova variable ordinal. Per exemple, sabem l'edat d'uns individus i volem classificar-los en joves, adults i grans.

Obriu l'arxiu **GSS93 reducido.sav**.

Classificarem la variable **edad** en 3 categories:

- **1 = joves (menys de 30)**
- **2 = adults (entre 30 i 64)**
- **3 = grans (65 o més)**

Procedim com abans:

Transformar / Recodificar / En distintas variables

Escollim com a variable d'entrada la variable **edad** i anomenem, per exemple, **edagrup** a la variable de sortida, etiquetada amb "Edat agrupada".

Premem l'opció *Valores antiguos i nous* i a cada valor de la variable **edad** (*valor antic*) li assignem el que li correspon a la nova variable **edagrup** (*valor nou*). Però ara farem servir els *Rangs*. Al *valor antic* Rang: Del menor fins a 29 li fem correspondre el *valor nou* 1:

Recodificar en distintas variables: Valores antiguos y nuevos

Valor antiguo

☐ Valor:

☐ Perdido por el sistema

☐ Perdido por el sistema o usuario

☐ Rango:

☒ Rango, MENOR hasta valor:

☐ Rango, valor hasta MAYOR:

☐ Todos los demás valores

Valor nuevo

☒ Valor:

☐ Perdido por el sistema

☐ Copiar valores antiguos

Antiguo -> Nuevo:

Lowest thru 29 -> 1

Añadir

Cambiar

Borrar

☐ Las variables de resultado son cadenas

☐ Convertir cadenas numéricas en números (5'->5)

Anchor: 8

Continuar

Cancelar

Ayuda

A continuació al *valor antic* Rang: de 30 fins a 64 li fem correspondre el *valor nou* 2, al *valor antic* Perdut pel sistema el *valor nou* Perdut pel sistema i finalment al *valor antic* Tota la resta de valors el *valor nou* 3. Ens quedarà el quadre següent:

Premem *Continuar* i *Acceptar*. Ara cal posar les etiquetes a la nova variable **edagrup** i tindrem la nova variable de tipus ordinal.

edagrup Edat agrupada

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Menys de 30	279	18,6	18,6	18,6
	2 Entre 30 i 64	931	62,1	62,1	80,7
	3 65 o més	290	19,3	19,3	100,0
	Total	1500	100,0	100,0	

► Exercici 4

Compareu les freqüències d'ambdues variables i construïu un diagrama de barres de la variable **edagrup** que hem creat.

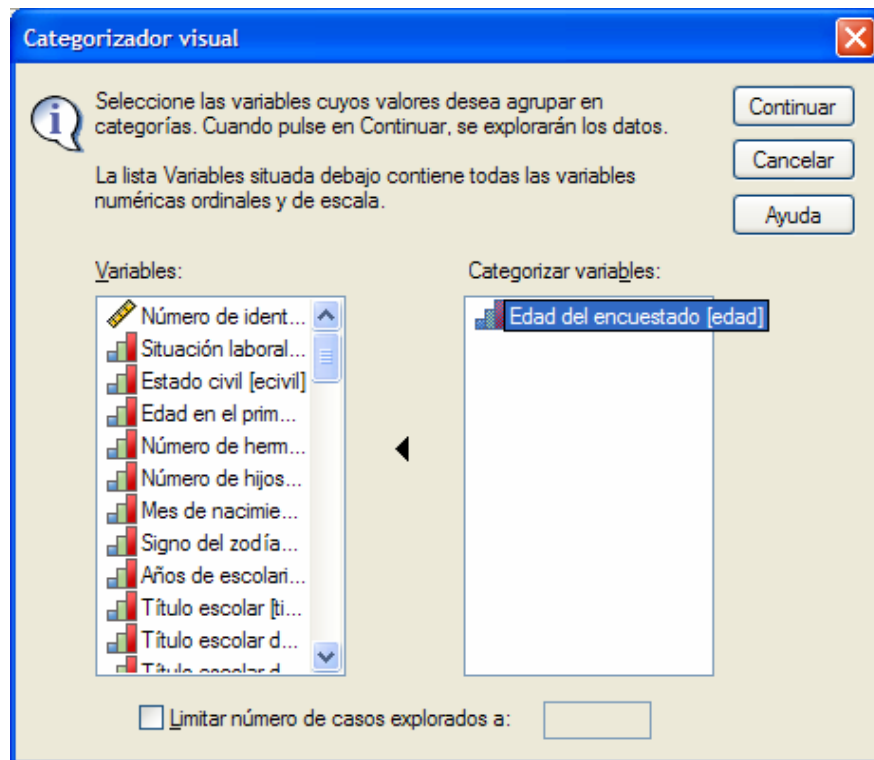
3. Creació d'interval·ls o classes de valors

Aquesta opció ens permetrà crear de manera personalitzada interval·ls o classes (o categories) de valors per a una variable quantitativa amb molts valors diferents.

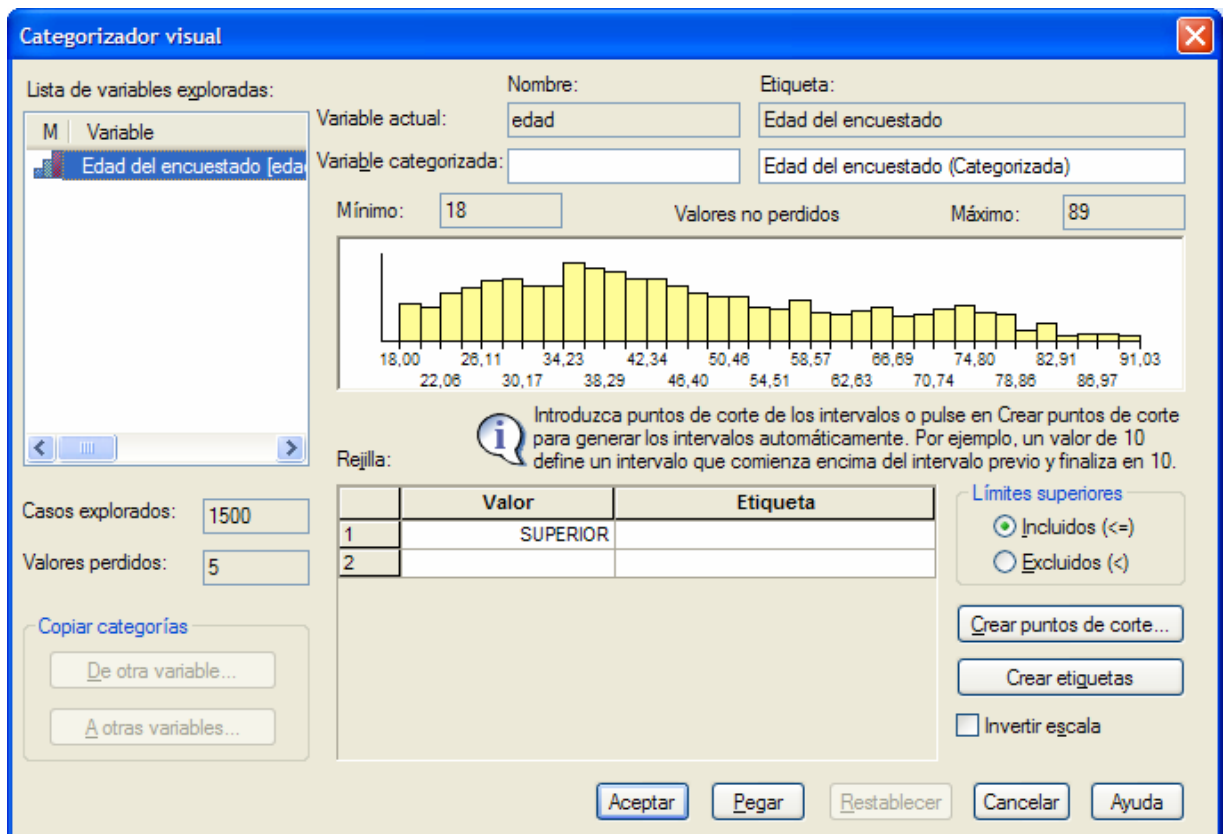
► Continuem treballant amb l'arxiu **GSS93 reducido.sav**. Utilitzarem el categoritzador visual de l'SPSS. El menú és:

Transformar / Categoritzador visual

La següent finestra ens permet escollir la variable que volem categoritzar. Escollim, per exemple, la variable **edad** (edat de l'enquestat).



Continuem i a la nova finestra, si premem la variable, se'ns mostra un histograma de les dades que tenim i se'ns ofereix, entre d'altres opcions, la possibilitat de crear els punts de tall que vulguem.



Podem escollir els punts de tall manualment, posant els valors corresponents a la taula, o podem fer-ho mitjançant la pestanya “**Crear puntos de corte**”.

Si optem per aquesta darrera opció, a la nova finestra de diàleg podem escollir tres opcions:

- Interval·s d'igual amplitud.
- Percentils igual·s.
- Punts de tall a partir de la mitjana i les desviacions típiques.

Escollim crear punts de tall en funció dels **percentils** amb **3** punts de tall. Això vol dir que crearem quatre interval·s, cadascun d'ells amb un 25% dels individus, delimitats pels 3 quartils.

The screenshot shows a dialog box titled "Crear puntos de corte" with a close button (X) in the top right corner. It contains three radio button options for creating cut points:

- ☐ Intervalos de igual amplitud: This option is currently unselected. It includes a sub-label "Intervalos: rellene al menos dos campos" and four input fields: "Posición del primer punto de corte:", "Número de puntos de corte:", "Amplitud:", and "Posición del último punto de corte:". To the right of these fields are three buttons: "Aplicar", "Cancelar", and "Ayuda".
- ☒ Percentiles iguales basados en los casos explorados: This option is selected. It includes a sub-label "Intervalos - rellene cualquiera de los dos campos" and two input fields: "Número de puntos de corte:" (containing the value 3) and "% de casos:" (containing the value 25.00).
- ☐ Puntos de corte en media y desviaciones típicas seleccionadas, basadas en casos explorados: This option is unselected. It includes three checkboxes: "+/- 1 Desv. típica", "+/- 2 Desv. típicas", and "+/- 3 Desv. típicas" (which is checked).

At the bottom left, there is an information icon (i) and a text block: "Aplicar reemplazará las definiciones de los puntos de corte actuales con esta especificación. Un intervalo final incluirá todos los valores restantes: N puntos de corte generan N+1 intervalos."

Un cop premem *Aplicar* es visualitza el resultat:

- Ens indica els valors dels punts de tall.
- Situa els punts sobre l'histograma.
- Demana el nom de la variable categoritzada: escollim, per exemple, **cat_edat**.
- Si volem podem escollir el nom que volem que tinguin les etiquetes. Si premem **Crear etiquetas** es generen automàticament.

Un cop premem *Acceptar* ens apareix la nova variable **cat_edat** a la base de dades.

► Exercici 5

Categoritzeu la mateixa variable escollint l'opció de crear els punts de tall a partir de la **mitjana** i les **desviacions típiques**. Feu-ho primer per a 1 desviació típica, després per a 2 i proveu de fer-ho per a 3 desviacions típiques: tenen sentit les categories obtingudes en aquest darrer cas per a la nostra variable?

► Exercici 6

Categoritzeu la mateixa variable escollint l'opció d'interval d'igual amplitud. Trieu 30 com a primer punt de tall i 4 punts de tall. Observeu el procés.

4. Transformació de variables mitjançant funcions

La transformació de variables és una eina que ens permet crear noves variables, a partir de les ja existents, aplicant una funció o càlcul que les relacioni.

► Per exemple, a la base de dades **mund0 95.sav**, ens pot interessar crear una nova variable **sperfi**, que ens doni la superfície de cada país, a partir de les variables **poblac** (població) i **densitat** (densitat de població).

Sabem que

$$\text{densitat} = \frac{\text{poblac}}{\text{sperfi}}$$

per tant,

$$\text{sperfi} = \frac{\text{poblac}}{\text{densitat}}.$$

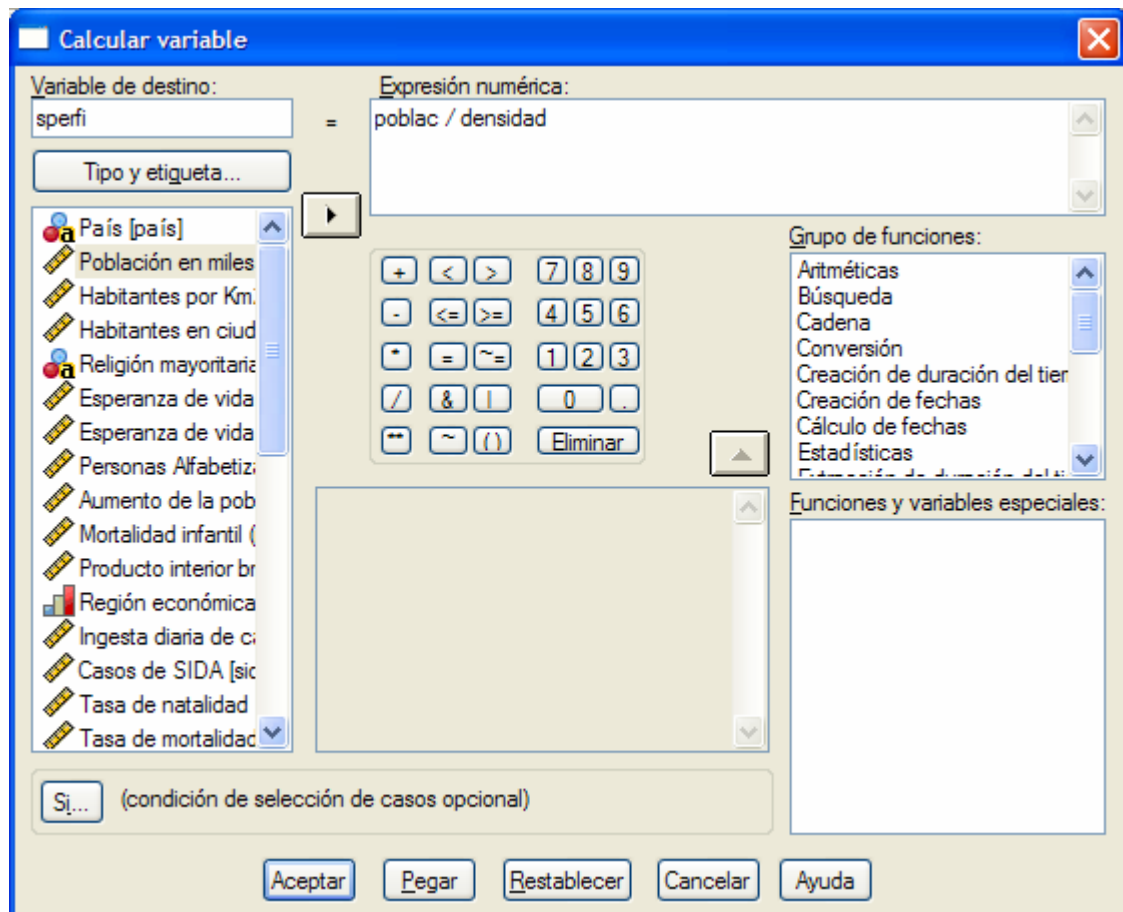
Aquesta darrera expressió ens dona la transformació que hem d'aplicar. El menú és el següent:

Transformar / Calcular / ...

Escollim com a Variable de destí **sperfi** i com a expressió numèrica

poblac / densitat

Podeu utilitzar la “calculadora” adjunta per indicar les operacions que cal realitzar:



Premem *Aceptar* i observem que s'ha creat la nova variable **sperfi**.

Observació: la superfície vindrà donada en **milers de Km²** ja que la població també venia donada en milers d'habitants. Especifiqueu-ho a l'etiqueta de la nova variable. L'etiqueta de la variable es pot crear al mateix quadr de diàleg de "Calcular variable" prenent el botó "Tipo i etiqueta".

► Exercici 7

Feu un diagrama de sectors dels climes predominants en el món. Observeu que heu de donar un pes diferent a cada país segons quina sigui la seva superfície. Utilitzeu la variable **sperfi** que heu creat.

► Exercici 8

A l'arxiu **Datos de empleados.sav**, creeu una nova variable **incrsal** que ens doni l'increment que ha tingut cada individu en el seu salari, és a dir, que calculi la diferència entre el salari inicial (**salini**) i el salari actual (**salario**).

► Exercici 9

Converteix la variable **salario** de dòlars a euros. Per fer-ho abans l'heu de convertir de format tipus "dòlar" a numèrica a la Vista de Dades. Supposeu un tipus de canvi de 1.20 dòlars per cada euro.

5. Obtenció de puntuacions tipificades

Un exemple important de transformació de variables és l'obtenció de puntuacions tipificades. Recordeu que les puntuacions tipificades ens permeten comparar les posicions relatives de dos o més individus respecte del grup.

Amb l'SPSS podem obtenir les puntuacions tipificades de dues formes:

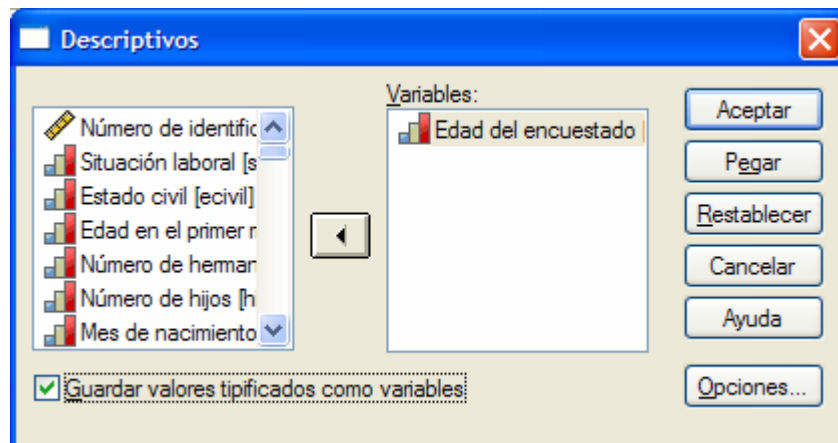
5.1 Obtenció automàtica de puntuacions tipificades

► Obriu l'arxiu **GSS93 reducido.sav**.

Volem obtenir les puntuacions tipificades de la variable **edad** (edat de l'enquestat). La forma automàtica de fer-ho és mitjançant el menú:

Analizar / Estadísticos descriptivos / Descriptivos

escollim la variable **edad** i seleccionem l'opció "Guardar valores tipificados como variables" abans de prémer *Aceptar*:



Observeu que ens crea una nova variable **Zedad** amb les puntuacions tipificades de la variable **edad**.

5.1 Obtenció manual de puntuacions tipificades

Recordem que les puntuacions tipificades s'obtenen mitjançant el càlcul següent:

$$z_i = \frac{x_i - \bar{x}}{s}.$$

És a dir, a cada valor de la variable li restem la mitjana i ho dividim tot per la desviació típica.

Per tant, una altra forma d'obtenir les puntuacions tipificades serà calculant primer el valor de la mitjana i la desviació típica de la variable **edad** mitjançant el menú:

Analizar / Estadísticos descriptivos / Descriptivos

Si a la finestra de resultats cliquem dos cops sobre el valor de la mitjana (o de la desviació típica) veurem el seu valor amb més xifres decimals:

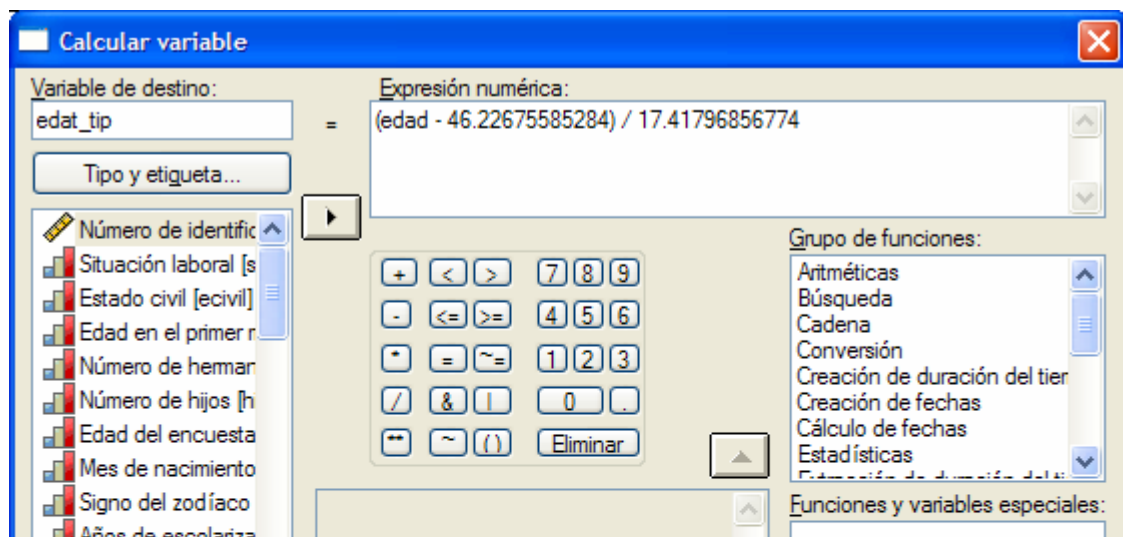
Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desv. típ.
Edad del encuestado	1495	18	89	46,2267558	17,418
N válido (según lista)	1495				

De fet, si utilitzem les opcions “Copiar” i “Enganxar” del menú contextual encara obtindrem més xifres decimals.

Un cop coneguts els valors de la mitjana i la desviació típica creem la nova variable mitjançant l'opció:

Transformar / Calcular

Escollim un nom per a la nova variable, per exemple, **edat_tip**, i apliquem la fórmula que ens dóna les puntuacions tipificades:



► **Remarca:** si utilitzeu el menú contextual per a “Copiar” i “Enganxar” el valor de la mitjana i de la desviació típica cal que canvieu la coma dels nombres decimals per un punt.

► **Exercici 10**

Construïu els histogrames de les variables **edad** i **edat_tip** i compareu-los.