

# Tratamento preliminar de dados com SPSS<sup>®</sup>

## 2. Tratamento preliminar de dados

**Nota:** A base de dados que vamos utilizar é um extracto da base de dados do *European Social Survey (round 1)*<sup>4</sup> com os resultados da aplicação do questionário em Portugal (ficheiro **ESS Portugal 2002 (TPAUB1).sav**)

### 2.1. Transformação de variáveis (comando *Transform*)

#### 2.1.1. Recodificação de variáveis

**Exemplo 1:** Pretende-se criar duas novas variáveis: a variável *idade* a partir da variável *f3* (data de nascimento) e recodificá-la, criando uma nova variável - *idade2* - com 4 escalões: “até 30 anos”; “31 – 50 anos”; “51 – 65 anos” e “> 65 anos”:

#### 1. Criação da variável *idade*:

The screenshot shows the 'Compute Variable' dialog box in SPSS. The 'Target Variable' field contains 'idade' and the 'Numeric Expression' field contains '2002 - f2'. A 'Type and Label' sub-dialog box is open, showing 'Label: idade' and 'Type: Numeric' with a width of 8. Arrows point from the sub-dialog fields to the main dialog fields.

- Nome da nova variável
- Expressão para criar a variável (2002 – o ano de nascimento)
- Label da nova variável

A variável *idade* acrescentou-se à base de dados:

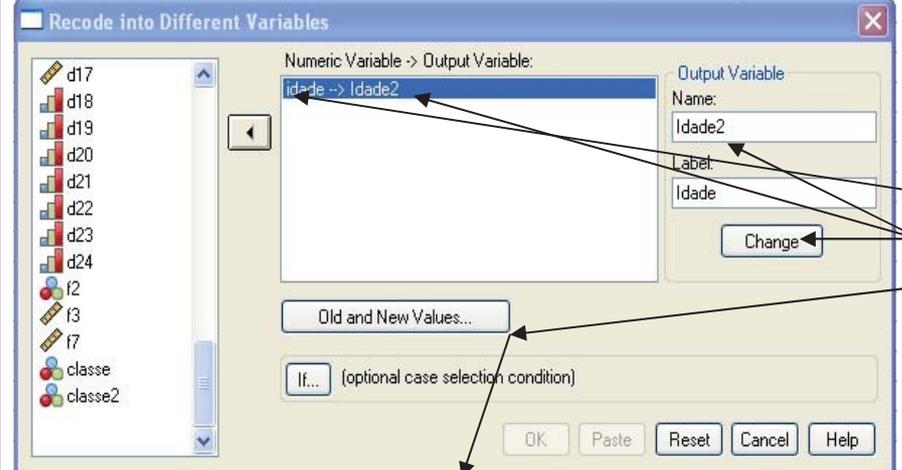
idade	Numeric	8	0	idade	None	None	1	Right	Scale
-------	---------	---	---	-------	------	------	---	-------	-------

<sup>4</sup> A base de dados original está disponível em <http://www.europeansocialsurvey.org/>.

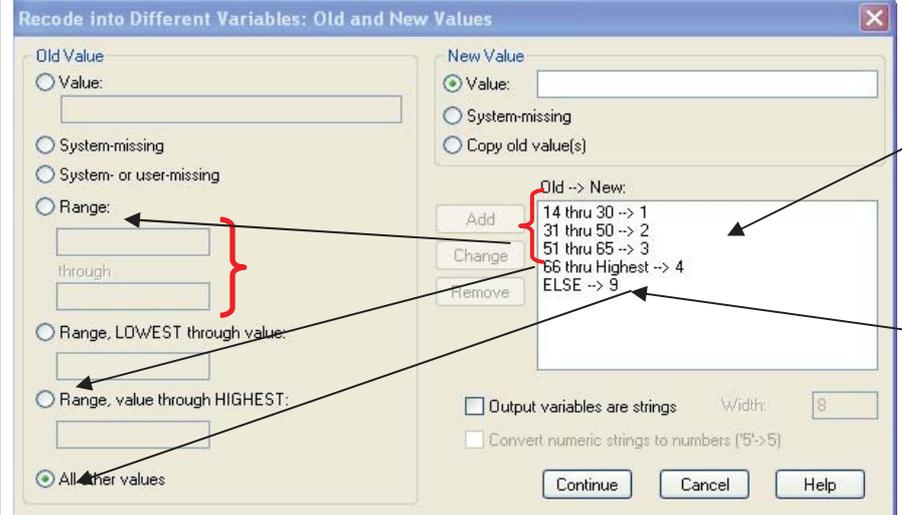
## Tratamento Preliminar de Dados

### 2. Recodificação da variável *idade* e criação da variável *idade2*:

Vamos agora proceder à sua recodificação, criando uma nova variável – *idade2* – com 4 escalões<sup>5</sup>:



- Variável original
- Nova variável
- Clicar

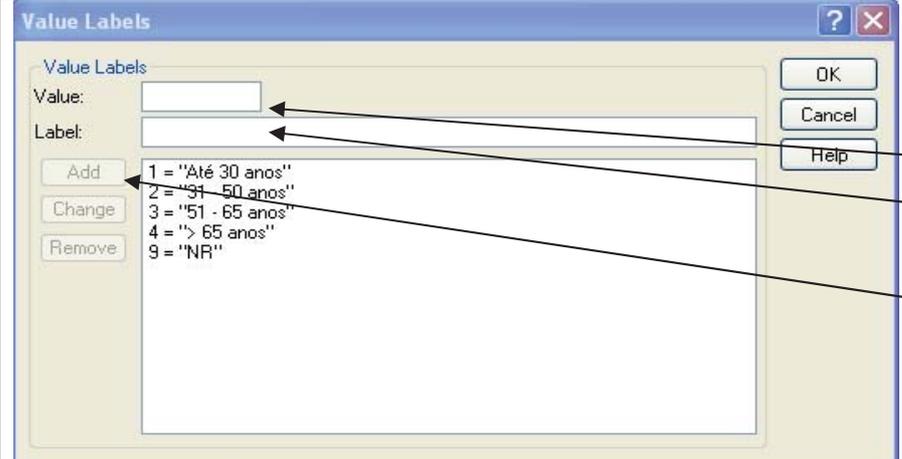


- Categorias
- Categoria para as "não respostas"

A variável *idade2* acrescentar-se-á à base e deverá ser completada com a alteração do nível de medida (*scale* para *ordinal*) e a definição dos respectivos *value labels*:

Idade2	Numeric	8	0	Idade	{1, Até 30 a ...	None	10	Right	Ordinal
--------	---------	---	---	-------	------------------	------	----	-------	---------

Clicar para criar as etiquetas das categorias (*Value labels*)



- Código da categoria
- Designação da categoria
- Clicar para adicionar

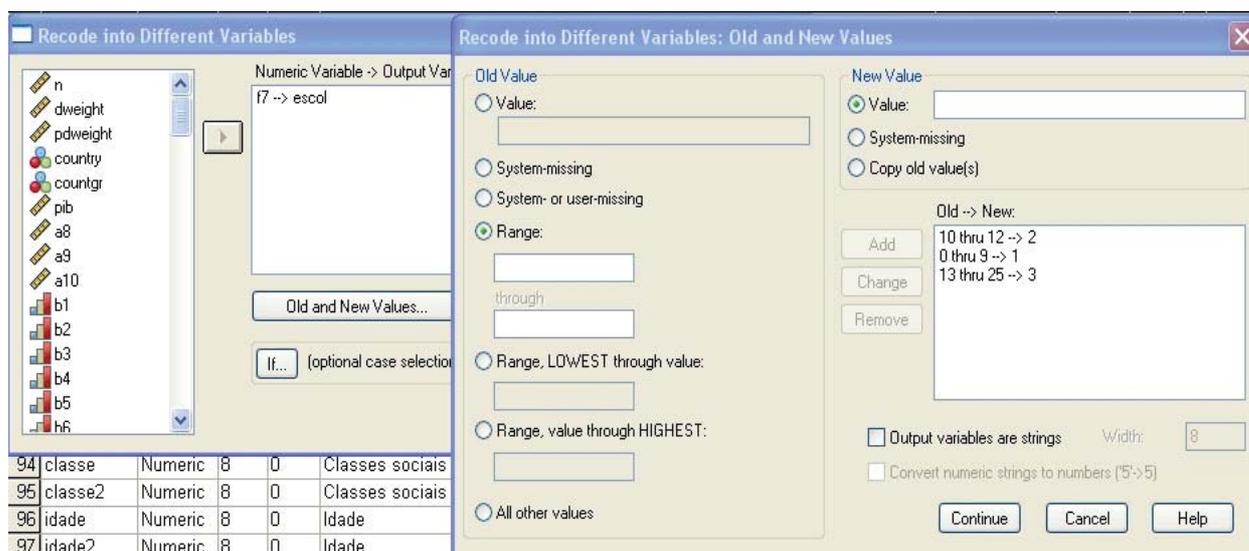
<sup>5</sup> **Nota:** na recodificação de variáveis é **recomendável** manter as variáveis originais e criar novas variáveis recodificadas, escolhendo para o efeito a opção **Into different variable**.

## Tratamento Preliminar de Dados

O resultado será o seguinte:

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Até 30 anos	341	22,6	22,6	22,6
	31 - 50 anos	505	33,4	33,4	56,0
	51 - 65 anos	315	20,8	20,8	76,8
	> 65 anos	350	23,2	23,2	100,0
	Total	1511	100,0	100,0	

**Exemplo 2:** Pretende-se recodificar a variável “escolaridade” (*f7*) criando uma nova variável (*escol*) com 3 escalões: “até 9 anos”; “10 – 12 anos” e “> 12 anos”:



A variável *escol* acrescentar-se-á à base e deverá ser completada com a alteração do nível de medida (*scale* para *ordinal*) e a definição dos respectivos *value labels*.

O resultado será o seguinte:

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Até 9 anos	1046	69,2	69,2	69,2
	10 - 12 anos	251	16,6	16,6	85,8
	> 12 anos	212	14,0	14,0	99,9
	NR	2	.1	.1	100,0
	Total	1511	100,0	100,0	

## Tratamento Preliminar de Dados

**Exemplo 3:** Pretende-se recodificar o “autoposicionamento político” (*b28*) criando uma nova variável (*b28r*) com 3 escalões: “esquerda”; “centro” e “direita”:

The screenshot shows the SPSS 'Recode into Different Variables' dialog box. The 'Numeric Variable -> Output Variable:' section shows 'b28 --> ?' with 'b28r' as the name and 'Autoposicionamento político' as the label. The 'Old and New Values' sub-dialog is open, showing the 'Range' option selected. The 'Old --> New:' list contains the following mappings: '0 thru 3 --> 1', '4 thru 6 --> 2', and '7 thru 10 --> 3'. The 'Output variables are strings' checkbox is checked, and the width is set to 8.

A variável *b28r* acrescentar-se-á à base e deverá ser completada com a alteração do nível de medida (*scale* para *ordinal*) e a definição dos respectivos *value labels*.

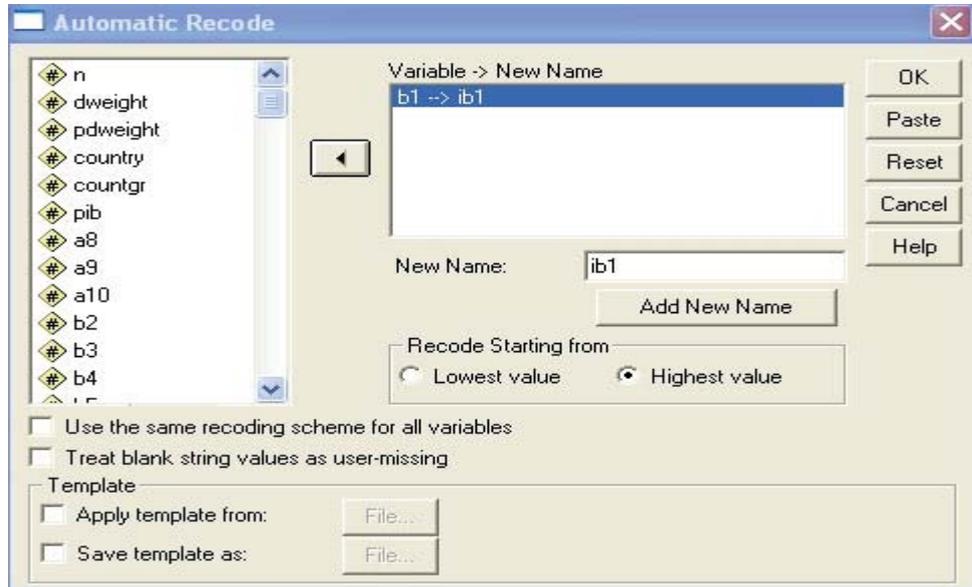
O resultado será o seguinte:

b28r	Numeric	8	0	Autoposicionamento político	{1, Esquerda}..None	1R	Nominal
<b>Autoposicionamento político</b>							
				Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Esquerda		296	19.6	24.5	24.5	
	Centro		620	41.0	51.2	75.7	
	Direita		294	19.4	24.3	100.0	
	Total		1211	79.9	100.0		
Missing	System		304	20.1			
Total			1515	100.0			

## Tratamento Preliminar de Dados

### 2.1.2. Inversão da escala de uma variável

Pretende-se criar uma nova variável (**ib1**) com a inversão da escala da variável interesse pela política (**b1**) de modo a que 1 corresponda a “nenhum interesse” e 4 a “muito interesse”:



O resultado é o seguinte:

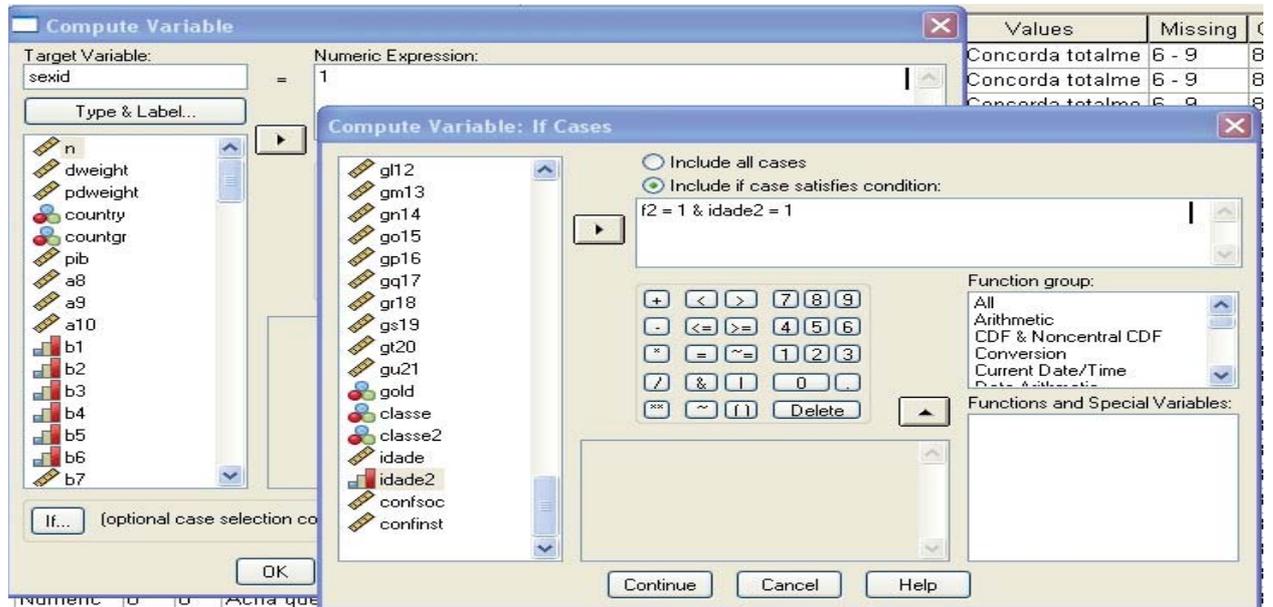
b1	Numeric	1	0	Qual o seu interesse pela política	{1, Muito interesse}...	6 - 9	8	Right	Ordinal
ib1	Numeric	1	0	Qual o seu interesse pela política	{1, Nenhum interesse}...	5 - 7	3	Right	Ordinal

<b>b1</b>						<b>ib1</b>						
Qual o seu interesse pela política						Qual o seu interesse pela política						
		Frequency	Percent	Valid Percent	Cumulative Percent			Frequency	Percent	Valid Percent	Cumulative Percent	
Valid	Muito interesse	117	7.7	7.8	7.8	Valid	Nenhum interesse	489	32.4	32.5	32.5	
	Algum interesse	456	30.2	30.3	38.1		Valid	Pouco interesse	441	29.2	29.3	61.9
	Pouco interesse	441	29.2	29.3	67.5			Algum interesse	456	30.2	30.3	92.2
	Nenhum interesse	489	32.4	32.5	100.0			Muito interesse	117	7.7	7.8	100.0
	Total		1503	99.5	100.0			Total	1503	99.5	100.0	
Missing	Recusa	6	.4			Missing	Não sabe	2	.1			
	Não sabe	2	.1				Recusa	6	.4			
	Total	8	.5				Total	8	.5			
Total		1511	100.0			Total	1511	100.0				

## Tratamento Preliminar de Dados

### 2.1.3. Transformação lógica de variáveis

**Exemplo:** pretende-se criar uma variável - *sexid* - através da transformação lógica das variáveis *f2* e *idade2*, com 4 categorias: “homens até 30 anos”, “homens com mais de 30 anos”, “mulheres até 30 anos” e “mulheres com mais de 30 anos”.



**Nota:** repetir o comando para as restantes categorias, cujas expressões numéricas são as seguintes:

- Categoria 2:  $f2 = 1 \ \& \ idade2 > 1$
- Categoria 3:  $f2 = 2 \ \& \ idade2 = 1$
- Categoria 4:  $f2 = 2 \ \& \ idade2 > 1$

A variável *sexid* acrescentar-se-á à base e deverá ser completada com a alteração do nível de medida (*scale* para *ordinal*) e a definição dos respectivos *label* e *value labels*. O resultado é o seguinte:

sexid		Nu	8	0	Sexo e Idade	{1, Homens at	None	10	Rig Nominal
<b>Sexo e Idade</b>									
						Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Homens até 30 anos					158	10.5	10.5	10.5
	Homens com mais de 30 anos					472	31.2	31.2	41.7
	Mulheres até 30 anos					183	12.1	12.1	53.8
	Mulheres com mais de 30 anos					698	46.2	46.2	100.0
	Total					1511	100.0	100.0	