

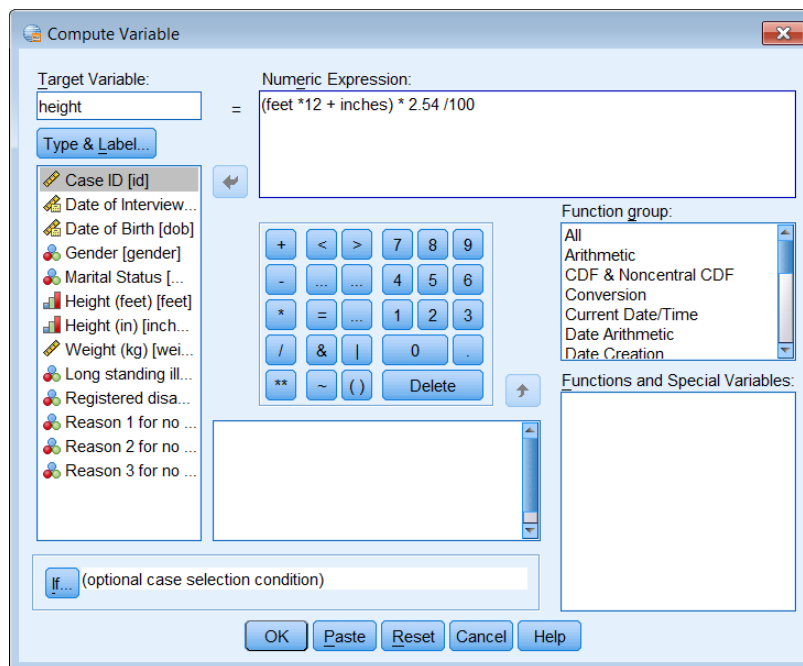
14 Creating New Variables and Modifying Data

In this section, features of SPSS that enable you to create new variables and modify existing variables will be described.

Suppose you wish to calculate the height of each respondent in metres and you have collected the data as two variables, feet and inches.

Go to the Data Editor window, Select **Transform | Compute Variable**.

The figure below shows the completed window for this.



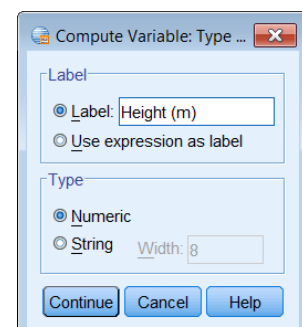
A new variable name, *height*, was typed into the *Target variable* box.

The *Numeric Expression* box is used to enter the formula used to calculate the formula for the target variable. The expression can be built either by typing it directly using the keyboard or it by clicking on the buttons on the calculator pad, the *Functions and Special Variables* list and the list of variables.

Items are added to an expression at the current cursor position, so you may need to move the cursor and click it to reset the insertion point in an expression. To delete an item, or sequence of characters, highlight the item/characters using the click and drag technique and then click on the **Delete** button on the calculator pad, or press the <Delete> key on the keyboard.

The formula for using the two variables, *feet* and *inches*, was entered as feet multiplied by 12 plus the value of inches, this total height in inches was then multiplied by 2.54 (inches to centimetres conversion) and then divided by 100 to obtain the result in metres.

Click on the **Type & Label** button to displays a dialogue box which the Variable *Label* information can be typed into.



The commands pasted into the Syntax Editor are:

```
COMPUTE height=(feet +inches*12) * 2.54/100.
VARIABLE LABELS height 'Height (m)'.
EXECUTE.
```

15 The Calculator Pad

The calculator pad has a number of arithmetic, relational and logical operators. Although these operators are presented as symbols, logical operators can be entered also as two or three character mnemonics. These are included in the table below.

Symbols and their meanings

+	Addition	<	Less than	LT	>	Greater than	GT
-	Subtraction	<=	Less than or equal to	LE	>=	Greater than or equal to	GE
*	Multiplication	=	Equal to	EQ	~=	Not equal to	NE
/	Division	&	Logical AND	AND		Logical OR	OR
**	Exponentiation	~	Logical NOT	NOT	()	Parenthesis	

Arithmetic operations have the following order of evaluation. Functions (e.g, LOG, MEAN) are evaluated first, followed by exponentiation (**), then multiplication / division and finally addition / subtraction. The order of calculation can be controlled by parentheses. Expressions inside parentheses “()” are evaluated first. So A+B/C means divide B by C and add the result to A, whereas (A+B)/C means add A and B together and divide the result by C.

The Logical and Relational operators are used when the creation of a variable is conditional on an expression. These are commonly used after the **If...** button has been clicked. This is described later.

Functions

SPSS for Windows provides many functions that can be used in expressions; these are grouped by function in the Compute dialogue window (*Function group*) and displayed in the *Function and Special Variables* list. They include:

- Arithmetic (ABS, RND, TRUNC, MOD, etc.)
- Statistical (SUM, MEAN, MIN, MAX, etc.)
- Search (RANGE and ANY, true if value is in range or list)
- Various Date and Time manipulations
- String and conversion functions

Select **Help | Command Syntax Reference**. Navigate the Command Syntax Reference to the *COMPUTE* section (page 332 -v18).

Missing Values and Functions

Functions treat missing values in a more useful way compared to simple arithmetic expressions. For example, if you wanted the mean of a set of variables, V1, V2, V3, using

```
COMPUTE X = MEAN(V1,V2,V3)
```

will calculate a mean for these three variables even if one, or more, of them has a missing value, whereas a statement such as

```
COMPUTE X = (V1+V2+V3)/3
```

will produce a missing result if any of the values of V1, V2 or V3 for a case are missing.

Rule for expressions

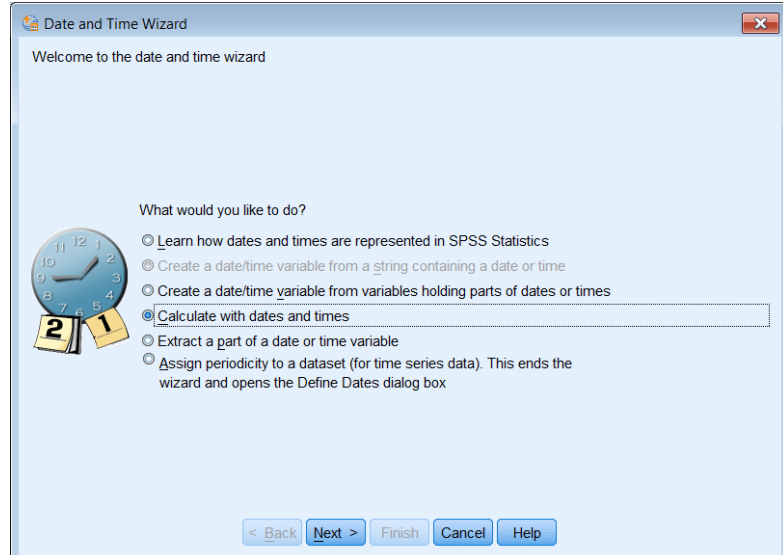
- The argument list for a function must be enclosed in parentheses.
- If a function has multiple arguments these must be separated by commas, spaces can be added for clarity.
- When comparing strings, text must be enclosed in either single ' or double " quotation marks, e.g. Sex = 'Male'. Otherwise Sex=Male would be trying to test the equality of two variables, Sex and Male.

The Date and Time Wizard

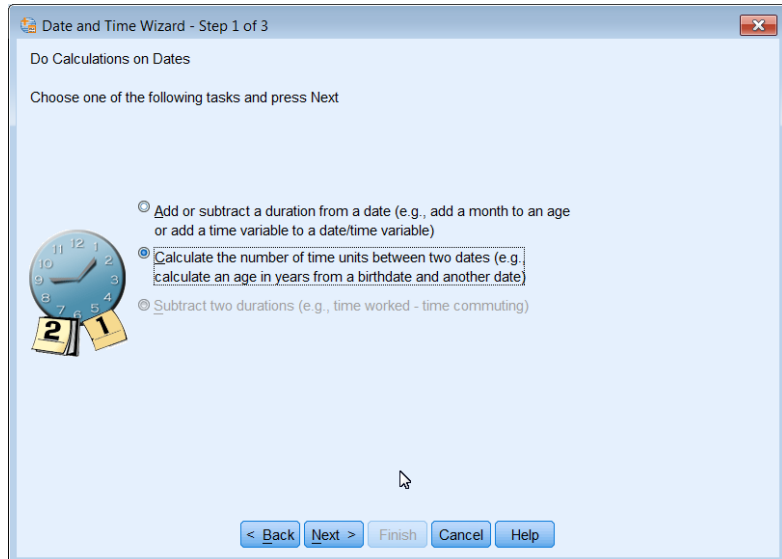
SPSS has a wizard to help you manipulate date or time variables and create new variables. It is selected from the menu, **Transform | Date and Time Wizard**. Choose options to help you generate a new variable. The sequence below shows how Age can be calculated from two dates.

First choose what type of calculation you wish to perform, here it *Calculate with dates and times*.

Click **Next**

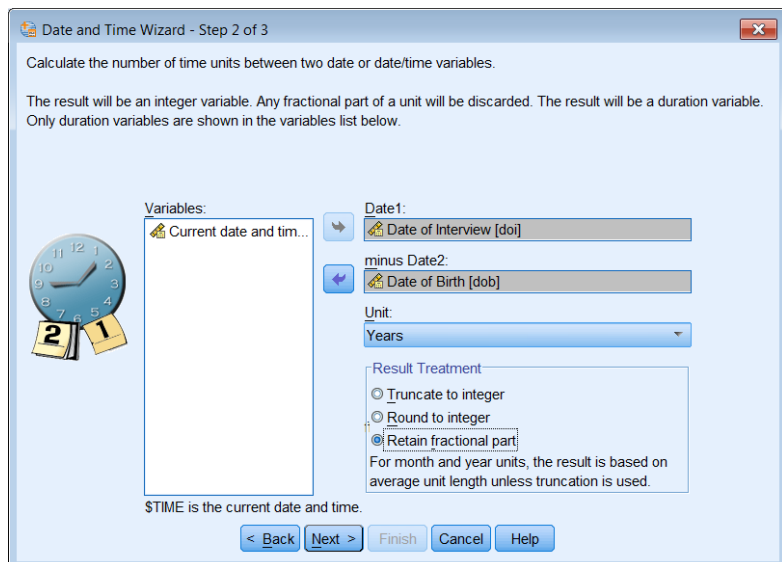


Next choose whether the calculation involves one or two date variables.



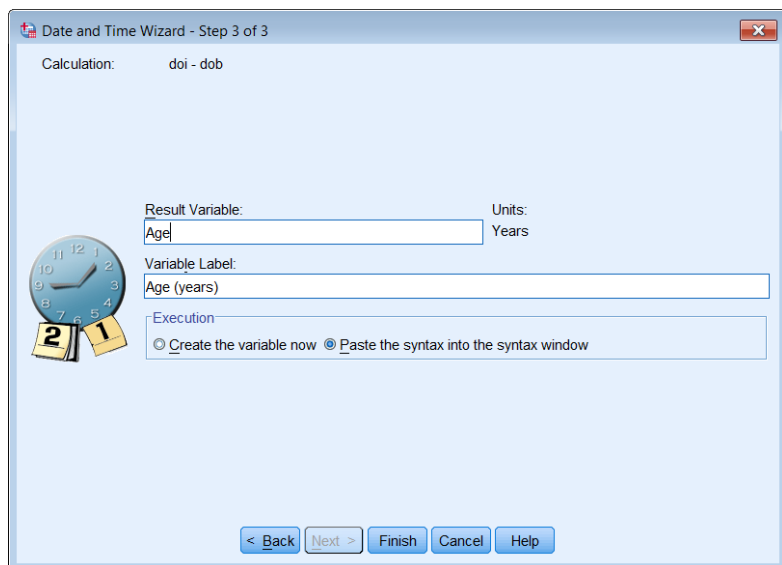
Click **Next**

Select the variables to be used for *Date1* and *Date2* (*Date of Interview doi* and *Date of Birth dob*), the *Unit* required, *years*, in this example and whether an Integer or Decimal result (*Result Treatment*) is required.



Click **Next**

Note SPSS has an in-built *Current Date and Time* variable called *\$TIME* that is available.



The syntax generated by the Date and Time Wizard is shown below. It uses a function called *time.days*.

* Date and Time Wizard: Age.

COMPUTE Age=(doi - dob) / (365.25 * time.days(1)).

VARIABLE LABELS Age "Age (yrs)".

VARIABLE LEVEL Age (SCALE).

FORMATS Age (F8.2).

VARIABLE WIDTH Age(8).

EXECUTE.

Counting

SPSS allows you to count occurrences of a response. Suppose you wish to know how many reasons respondents gave for not taking exercise to three questions (*noex1*, *noex2*, *noex3*). Did respondents give no, one, two or three reasons for not taking exercise?

Select **Transform | Count Values within Cases**.


Enter **noexcnt** in the *Target Variable* text box.

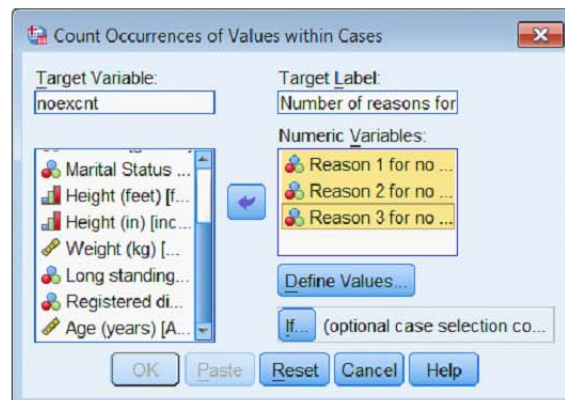
Enter '**Number of reasons for no exercise**' in the *Target Label* text box.

Highlight **Reason 1 for no exercise [noex1]**.

Press <Shift> and Click on **Reason 3 for no exercise [noex3]**.

(This highlights all three variables sequentially.)

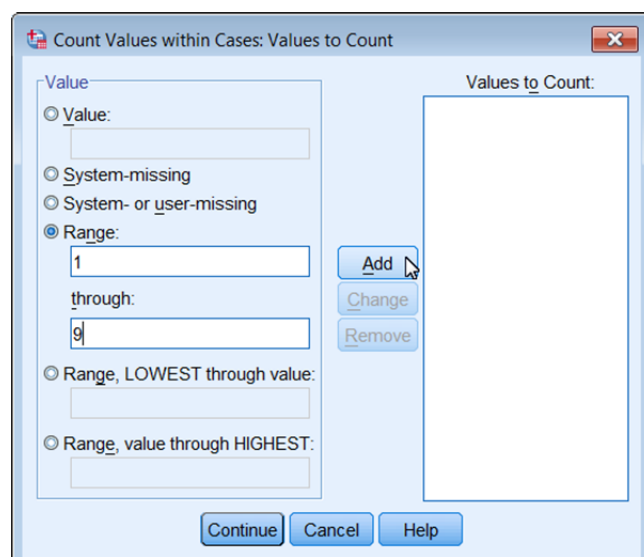
Press  to add these variables to the Variables list box.



The **Define Values...** button is used to enter the values of each variable that are valid, that is the values you want to Count.

For this example a valid response to the three questions (*noex1*, *noex2* and *noex3*) is recorded by values between one and nine.

Click on **Range** in the *Value* panel and enter **1** to **9** and **Add** this to the list as shown.



Count can create a new variable in several ways. It depends on the option selected.

Value	Counts the occurrence of the value given. This is the only option for string values.
System-Missing	Counts the values of the system-missing value.
System or user-missing	Counts any all missing values, system-missing or user define missing.
Range	Counts values in the given range.
Range, LOWEST through value	Counts from the lowest value encountered to given value.
Range, value though HIGHEST	Counts from a given value to the highest value encountered.

Tips and Hints

Count is more flexible than the Dialogue boxes would suggest. Suppose that you wanted to build a variable, called total, from correct answers to various questions (q1, q2, q3 etc.) and that some answers were given more marks than others. Count can be used for this but only by directly typing the command into the Syntax window. For example:

COUNT total = q1 q1 q2 (1).
VARIABLE LABELS total 'Total Marks'.

The Count command would give two marks for answering 1 to question 1 (q1) correctly and one mark for answering 2 correctly, where a value of 1 is a correct answer and any other value is considered to be wrong. The alternative way of doing this, but involving more typing would be:

COMPUTE total = 0.
IF (q1 = 1) total = total + 2.
IF (q2 = 1) total = total + 1.
VARIABLE LABELS total 'Total Marks'.

IF - Combining Logical and Relational operators

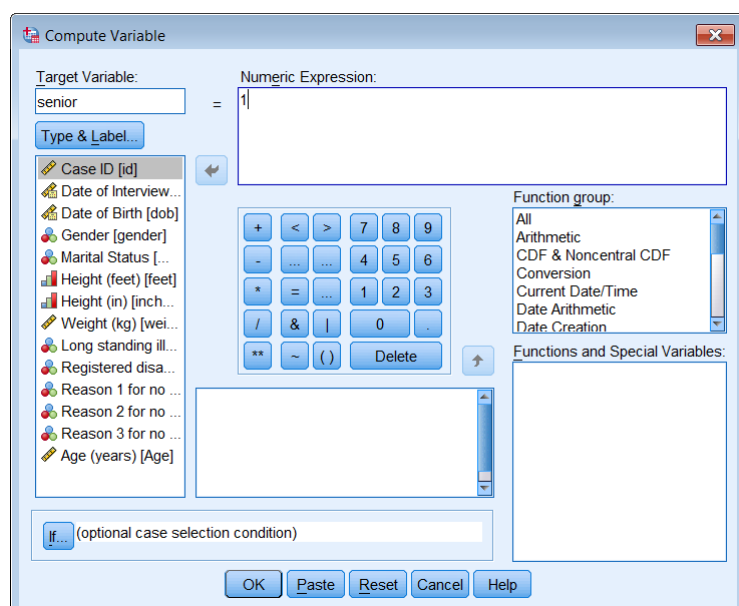
Suppose that you wished to calculate a new variable to identify Senior Citizens in a study. Create a variable *senior* that will equal 1 if the respondent is over 65.

Select **Transform | Compute**.

Click on Reset to clear any existing contents of this dialogue box before starting.

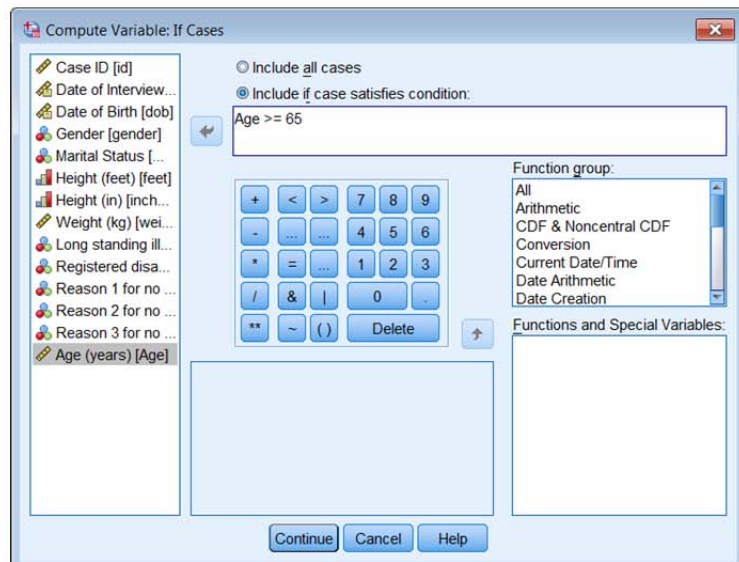
Use the **Compute Variable** dialogue box to create the Target Variable to *senior*, and enter 1 in the Numeric Expression panel.

Click the If... button to display the **Compute Variable: If Cases** dialogue box.



The ***Include if case satisfies condition***: radio button must be checked for the logical condition within this window to apply. When ***Include all cases*** is chosen any conditional expression will be ignored.

Build or enter the expression:
age >= 65



The SPSS syntax for this is:

IF (Age >= 65) senior=1.
EXECUTE.

Cases where age is less than 65 will be given the system missing value.


Recode

Suppose that you want to recode Age of Respondents into Age groups. This can be done using the Recode command. The Recode command is comparable to saying, "If the value is now this, make it equal to that".

Select:

Transform | Recode into Different Variables

Click on ***age*** in the list of variables.

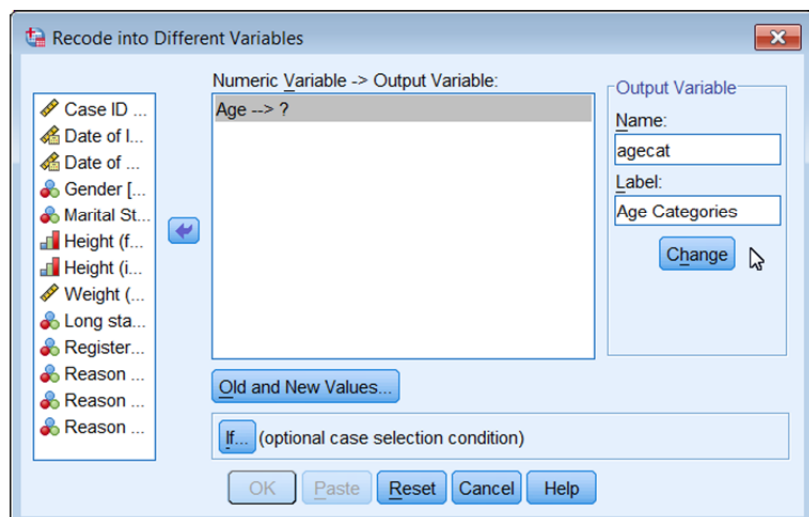
Click on .

Click on the ***Output Variable- Name*** text box and enter **agecat**.

Click on the ***Label*** text box and enter **Age Categories**.

Click on **Change**.

Click on the **Old and New Values** button.



Always take care of missing values first.

Click on *System- or user-missing* in the *Old Value* panel.

Click on *System-missing* in the *New Value* panel.

Click on the **Add** button.

Click on *Range, LOWEST through value* button (Old Value panel), enter 25 in the text box.

Click on *Value* (New Value panel) enter 1.

Click on **Add**.

Click on *Range*:(Old Value panel), enter 25.01 in the first text box and 65 in the second text box.

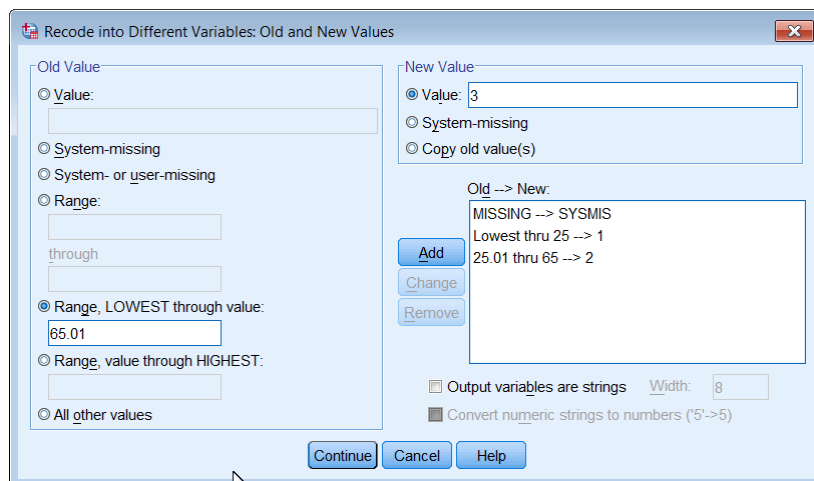
Click on *Value* (New Value panel) enter 2.

Click on **Add**.

Click on *Range, value through HIGHEST* button (Old Value panel), enter 65.01.

Click on *Value* (New Value panel) enter 3.

Click on **Add**.



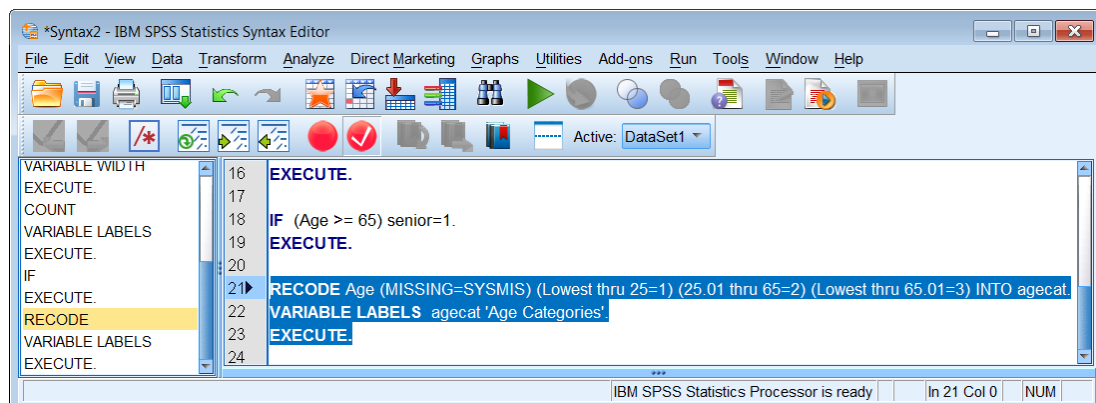
Click on *All other values* button (Old Value panel).

Click on *Copy old value(s)* (New Value panel).


Click on **Add**.

Click on **Continue**.


Click on **Paste**.



Switch to the syntax window and highlight the text:

Click on the Run highlighted command syntax button .

Switch to the Data Editor window by clicking in it, select the *Variable View* sheet.

Select the *Agecat* Values cell and click on .

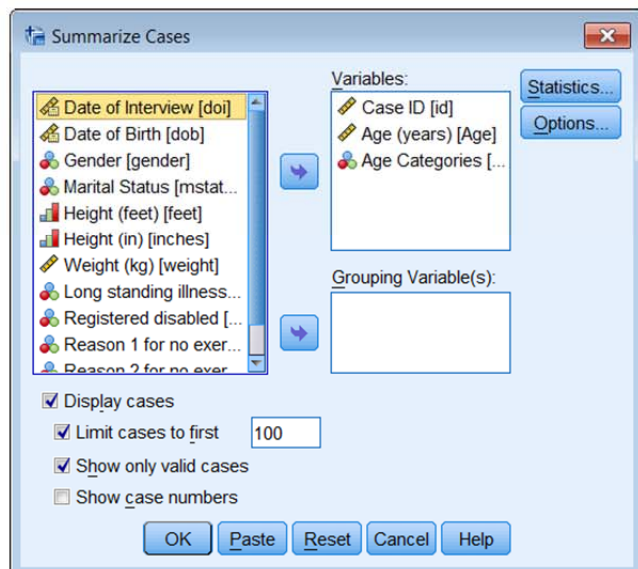
Add suitable *Value Labels* for the *agecat* variable (look at the recode command).

16 Listing Data Values

The Summarize command displays the values of selected variables for selected cases or all cases.

Use **Analyze | Reports | Case Summaries** to list data when creating new variables so that you can confirm SPSS has calculated what you expect.

In this example *id*, *age* and *agecat* are summarised to check the recoding as shown in the section of Output below.



Case Summaries^a


	Case ID	Age (years)	Age Categories
1	26	59.05	2.00
2	29	36.15	2.00
3	41	64.04	2.00
4	59	17.47	1.00
5	69	29.37	2.00

17 Saving modified data

After modifying data/variables (by recoding, etc.) and/or creating new variables, you should always save the updated SPSS system file. It is always recommended that you create a new file, for example, Health4.sav. If you save the SPSS data to an existing file (the one you opened initially) and something had gone wrong during your SPSS session, perhaps you accidentally deleted some important variables, you will have to recreate, or re-enter this information.

To save the file, make the SPSS Data Editor the active window (click on it) and then select: **File | Save As** and give the file a new name, e.g. Health4.sav, click the **Paste** button, switch to the syntax window and run the command:

```
SAVE OUTFILE='C:\Users\tnb\Desktop\health4.sav'  
/COMPRESSED.
```

DO NOT Click on the save button  in the Data Editor window as this will try to replace the original data file *soton11.sav*, which you started with.