Guideline 5_SPSS: Data manipulation_ Data Transformation

Data transformation allows us to edit certain aspects of the data, be that making new variables or editing current variables to suit our needs, e.g. regrouping values into categories or converting (character) string to numbers.
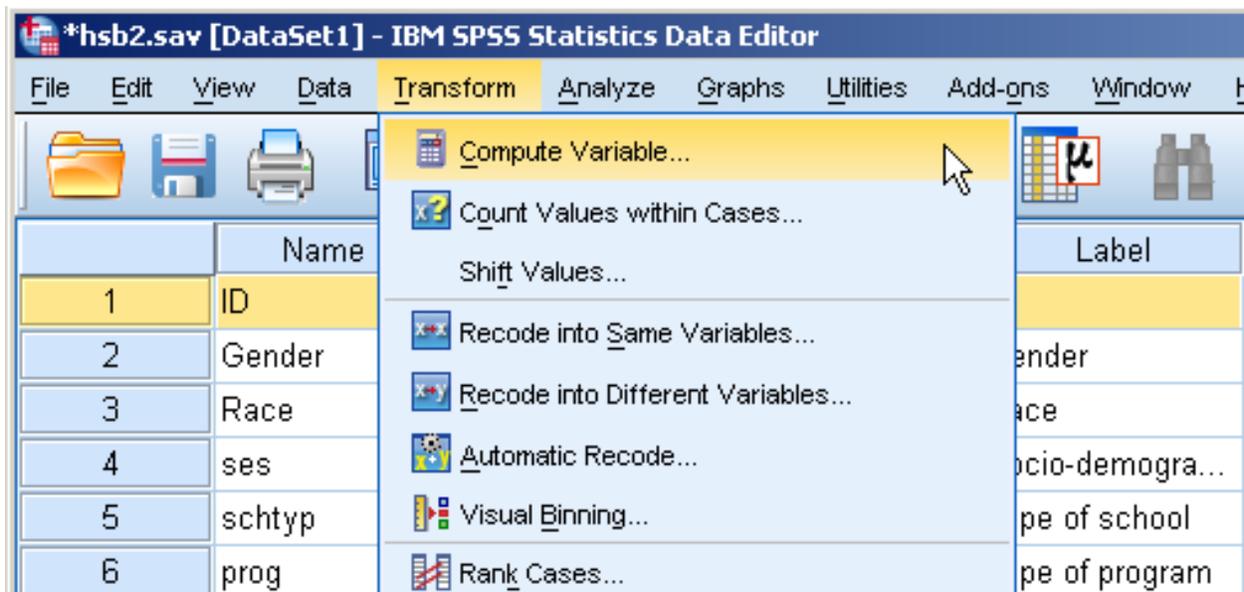
Most commonly used functions include:

- **Compute**: Create a new variable through some calculation
- **Recode**: Recodes certain values into another value
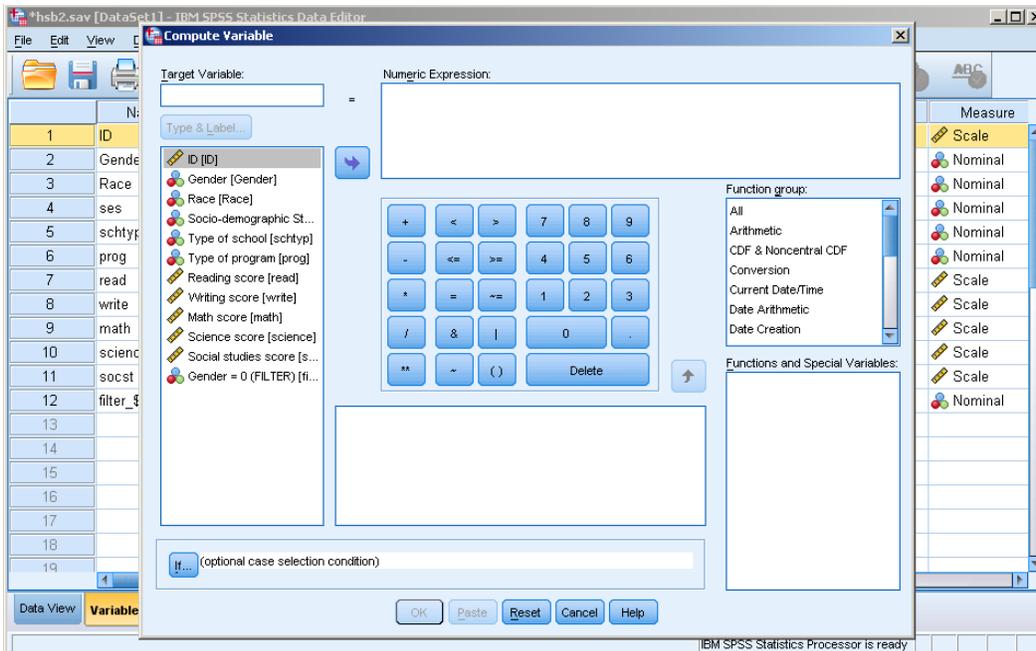
How to compute?

Let's say that we want to determine the students overall "intelligence" by combining the scores for reading, writing, math, and science (e.g. SAT, GRE, etc). How do we do this?
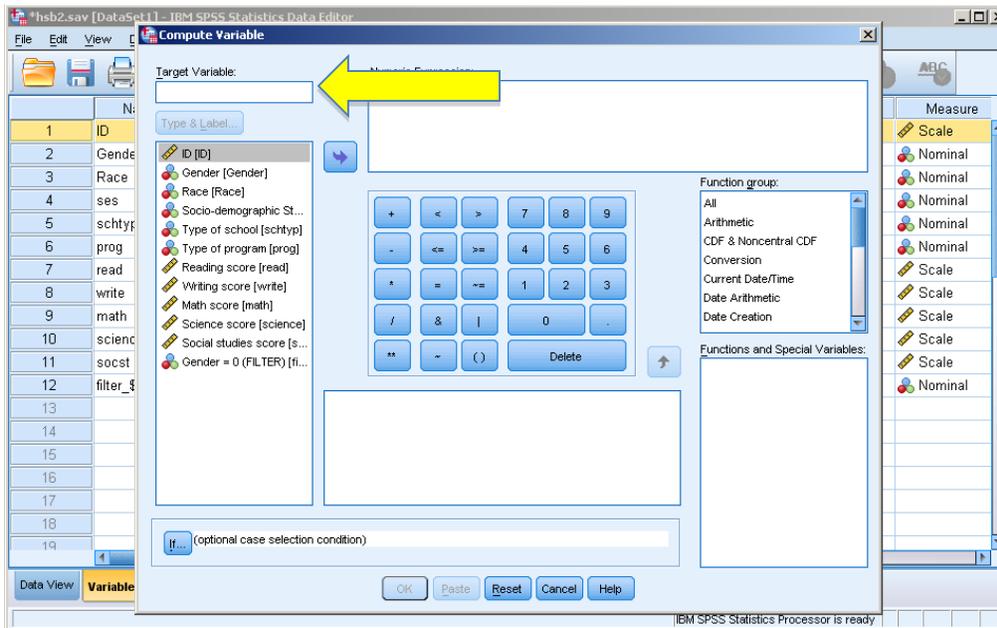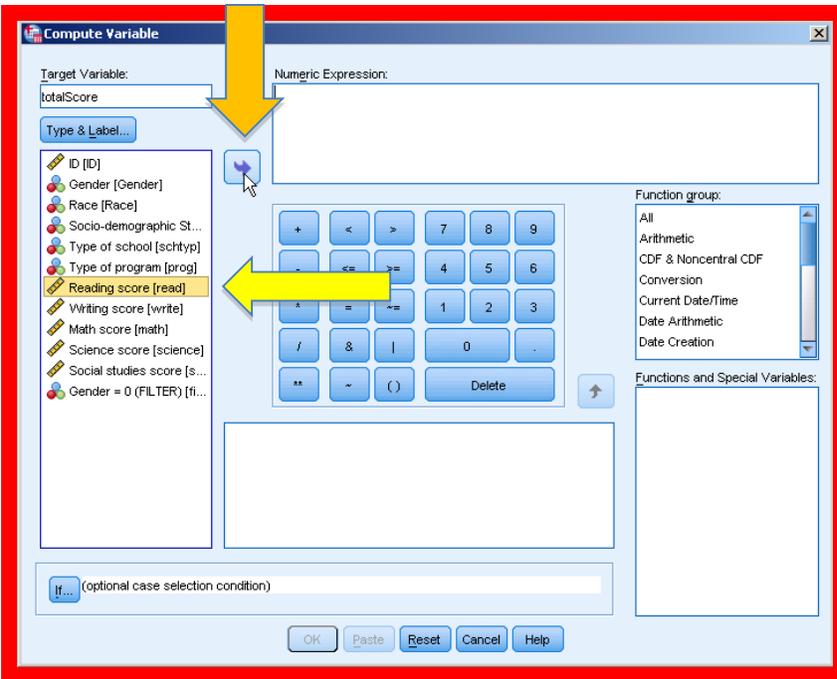
Go to Transform<compute

Click on Compute and a new menu will be opened as shown below:

You should add up scores of reading, writing, math and science. Target value is the name of combined score. In that box type: total score


Then add the varaibles (reading, writing, etc) one at a time with a plus sign in between

Move each variable to the "numeric expression" box one at a time. Use the arrow to make the shift. See how the variable "read" is added to the numeric expression box. Repeat the process for other variables, until all variables are shifted. Use the plus sign in between varaibles.
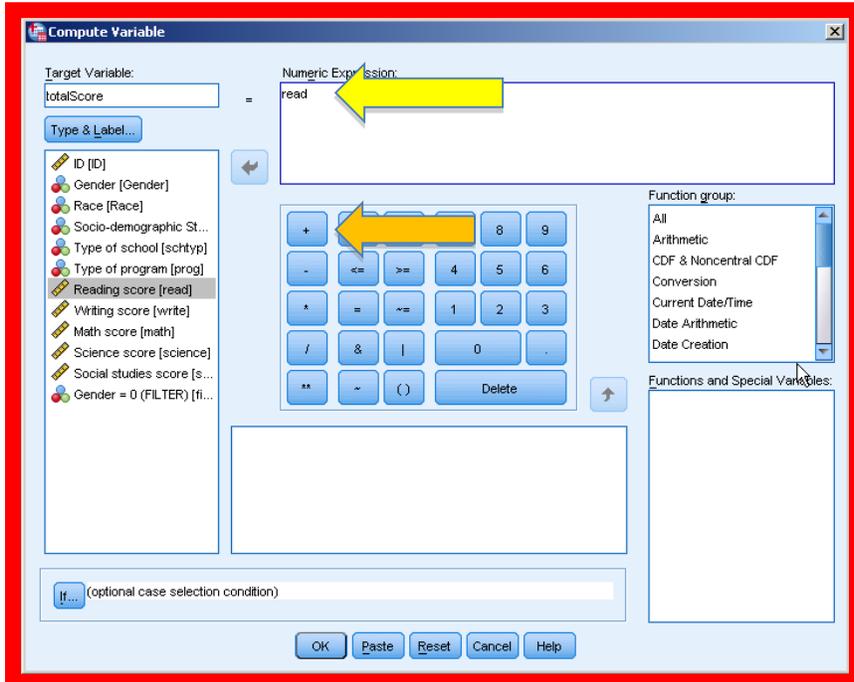
Here is the final line: Now all the four variables in the numeric expression box. You can now, click on "Type &Label" and give a label to the new variable.
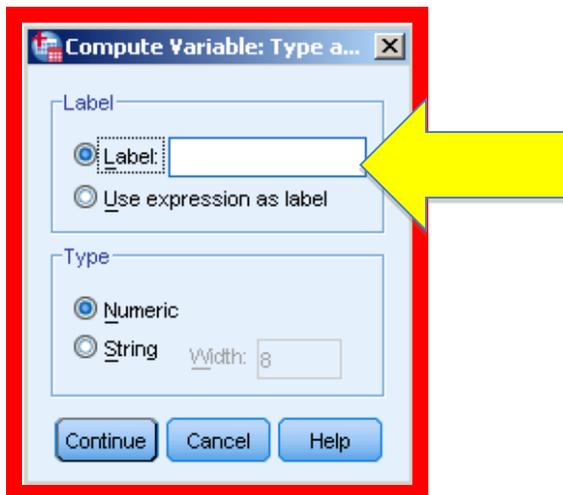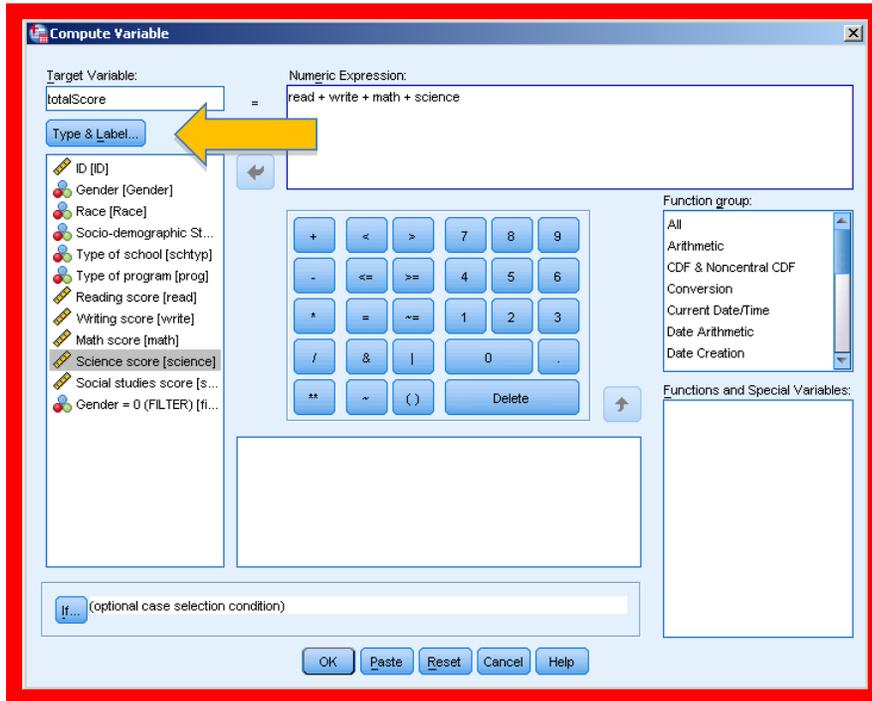
Fill up this box with a name of your choice.

Question 1: Can you find the mean of "total score"?

Question 2: Can you find the mean of total score for each gender?

Answers:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
totalScore	200	139.00	277.00	209.5000	32.94933
Valid N (listwise)	200				

```
SORT CASES  BY Gender.
SPLIT FILE LAYERED BY Gender.
DESCRIPTIVES VARIABLES=totalScore
  /STATISTICS=MEAN STDDEV MIN MAX.
```

Descriptive Statistics

| Gender		N	Minimum	Maximum	Mean	Std. Deviation
male	totalScore	91	144.00	277.00	209.1209	35.03897
	Valid N (listwise)	91				
female	totalScore	109	139.00	274.00	209.8165	31.25853
	Valid N (listwise)	109				

Recode

Let's say math scores are a very important marker for problem solving skills. Those that have a score of less than 40 are considered poor math skills, 40 to 60 is average, and +60 is considered above average. How many of the students are above average?

To use recode function, go to Transform<Recode into different variable.

Note: If you use "recode into same variables" option, you will overwrite the existing variable. You do not want to do that!
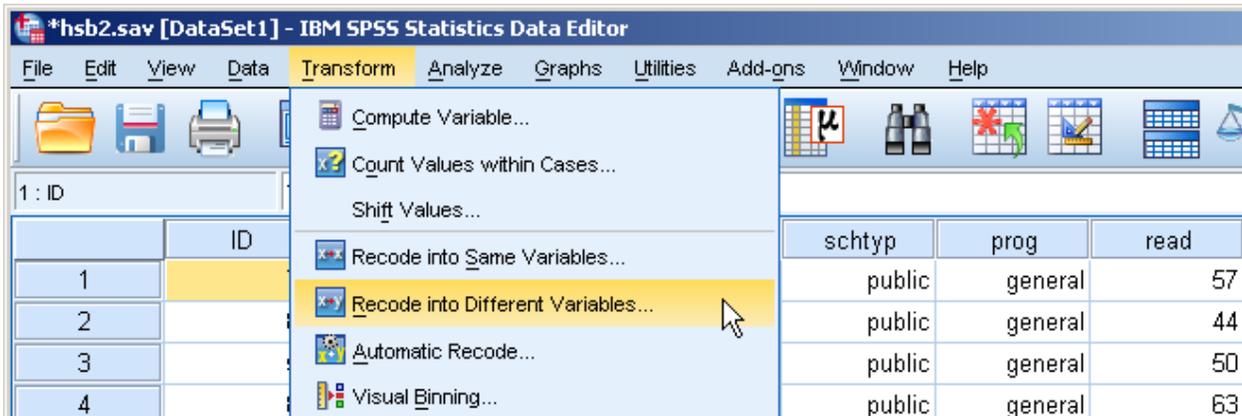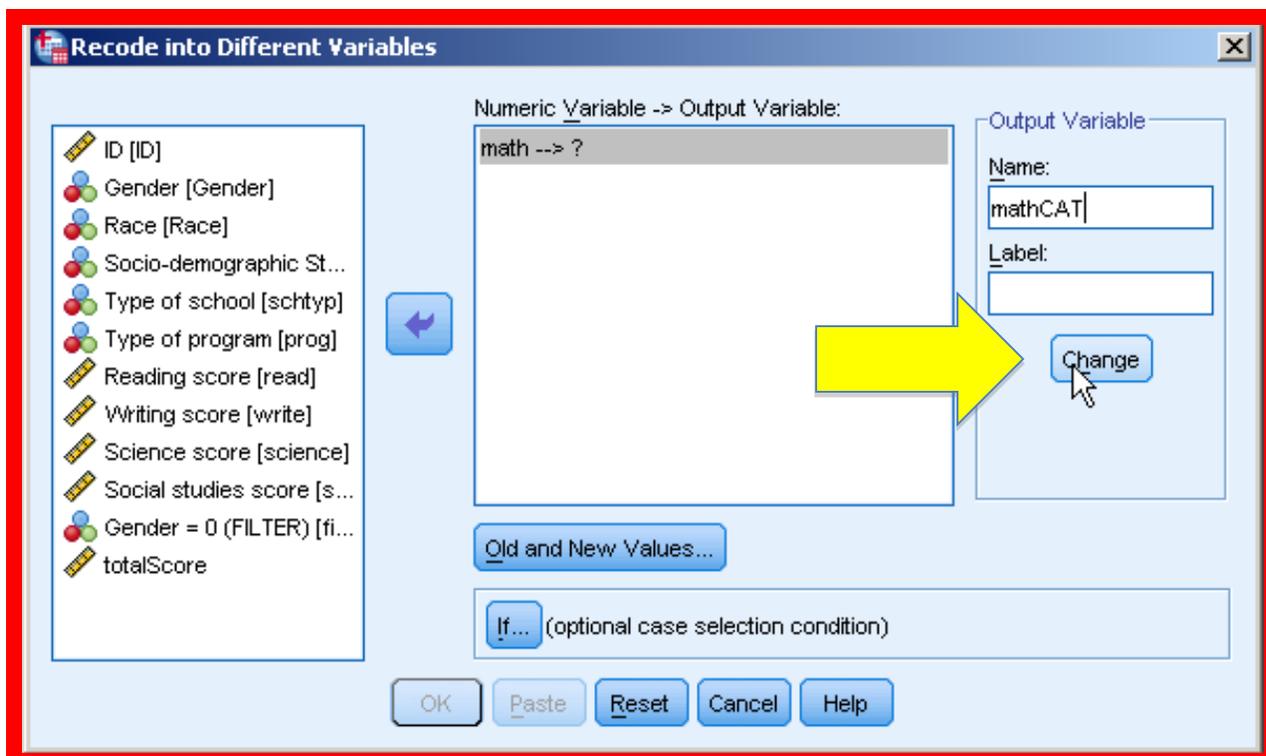

Click on "Recode into different variable" and a new menu will be opened as shown below. Shift the variable math to the empty box.
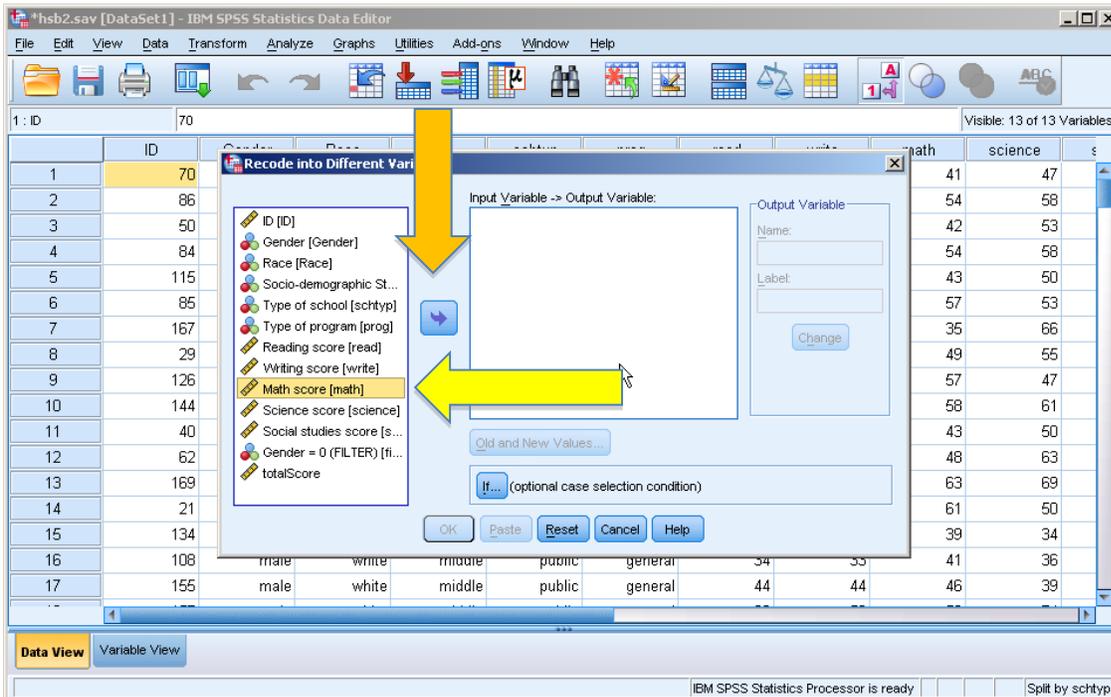

Immediately after you send a variable to the empty box, the output variable section will be highlighted.
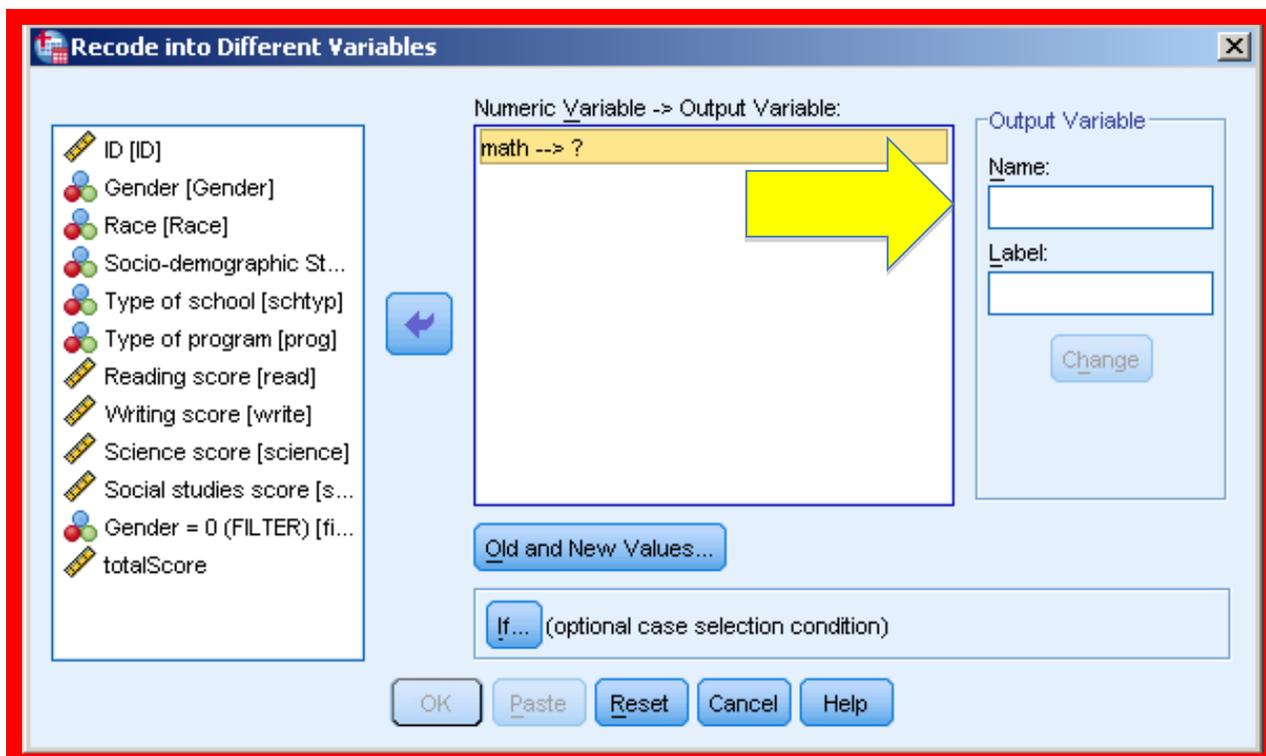
You can now fill up that box using a new name, say "mathCAT"

Click change. You will see that the question mark (?) in front of math in the numeric variable box will change to the new variable name (mathCAT). Here you are telling SPSS that I am creating a new variable which will be a categorised version of the continuous variable, the math.
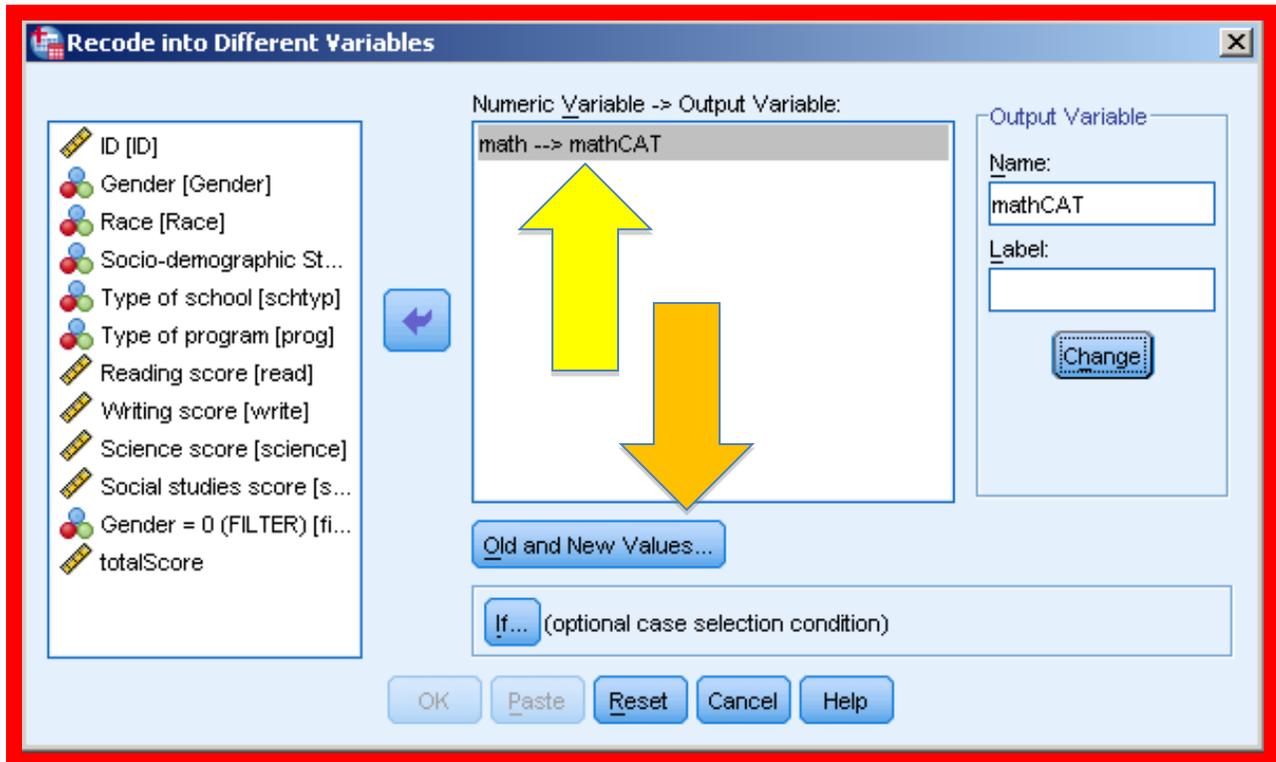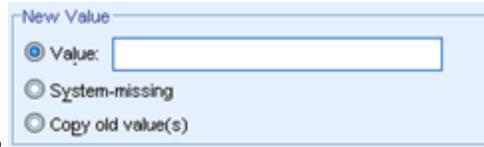

Then click on "Old and New Values" key as shown with orange arrow. A new menu will be opened: You will see three areas:
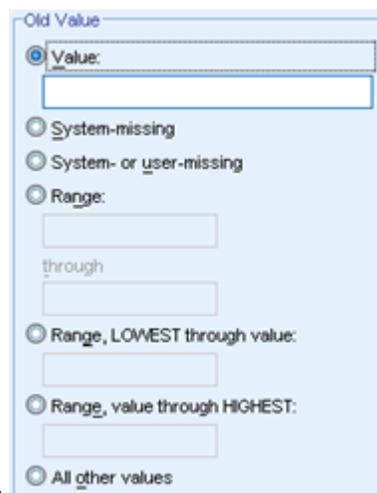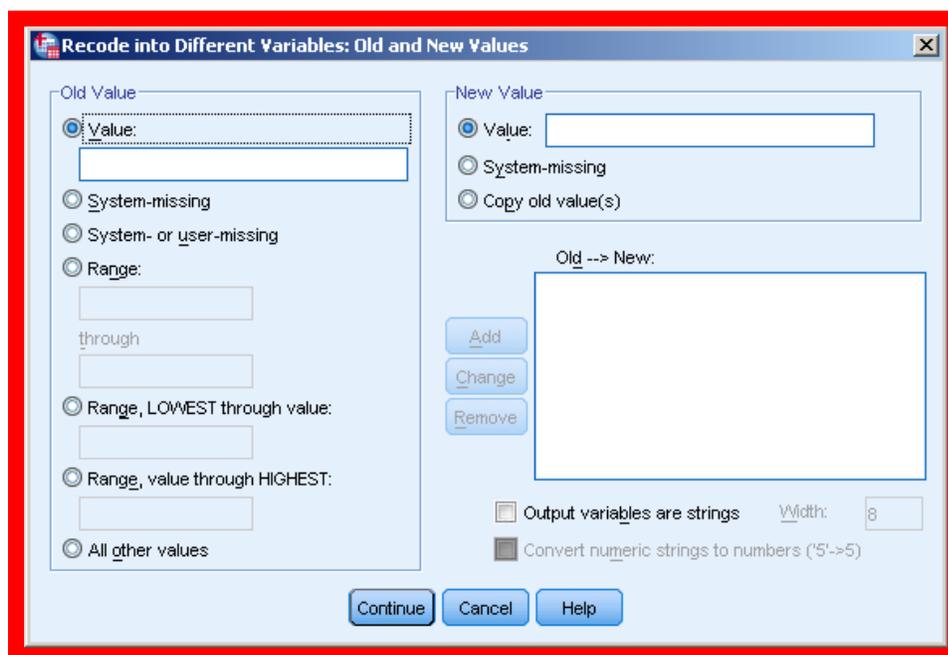

1. Old Value

2. New value

3. Old>New

You will enter your condition on the old value section, your new code under Old value section and transfer Old to new code under Old>New section.


Let us say that you want to define three categories for math score

- Below 39 and code it as 0
- Between 40 to 50 and code it as 1
- Above 50 and code it as 2

To start coding the first one (Below 39 and code it as 0), type 39 in the box titled as "Range, LOWEST through value, type 0 in the box titled "value" under New Value section and click add as shown by the orange arrow:
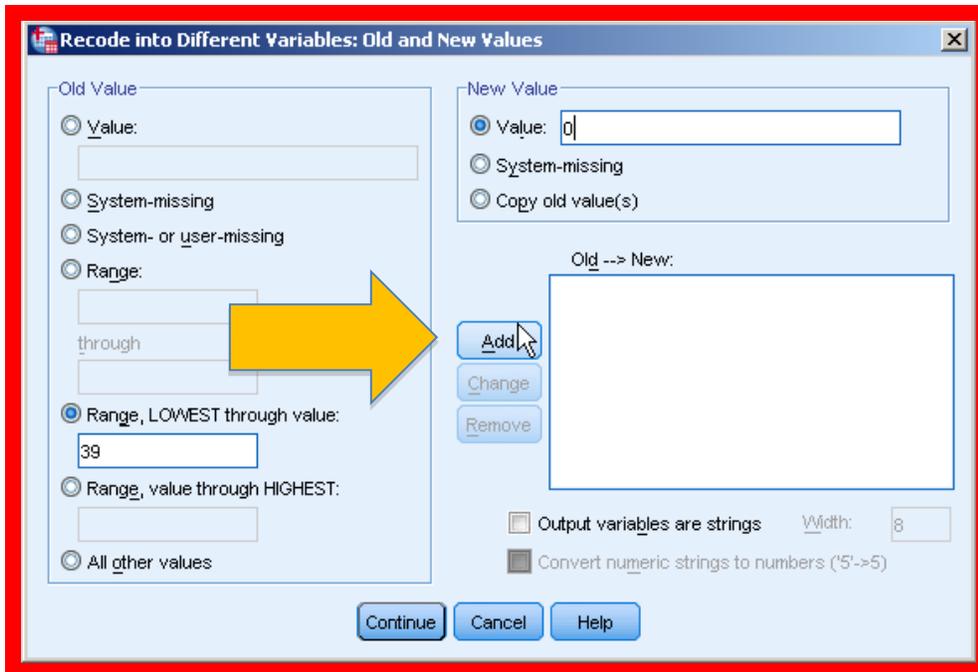
Now start another transformation or recoding for the next level. This time, click on Range
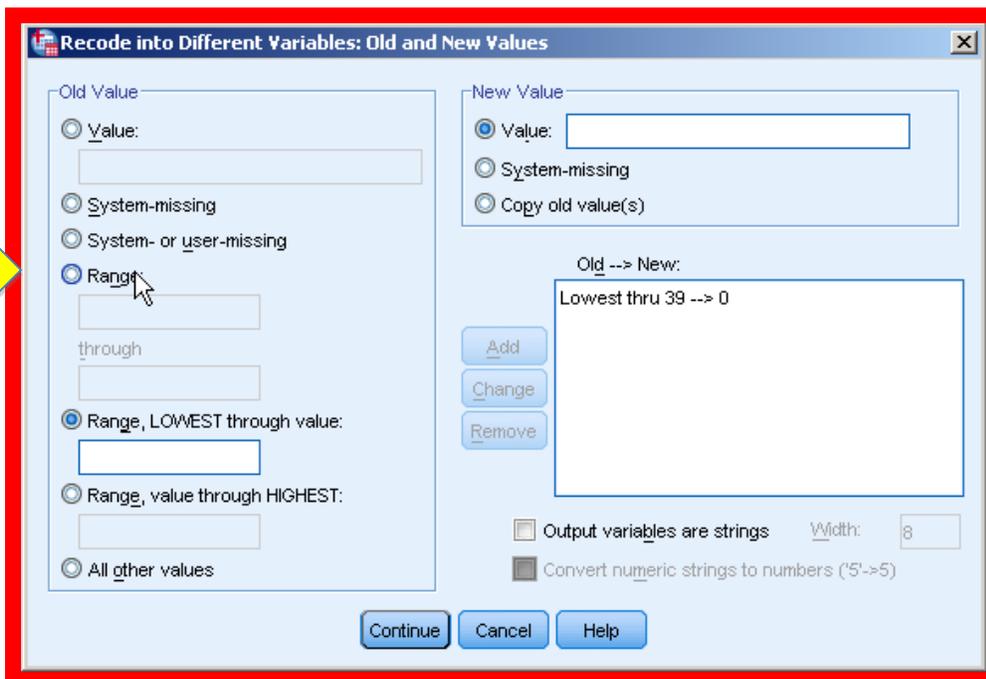
Range section will be highlighted. Add 40 and 60 under old section, and code 1 under the new section