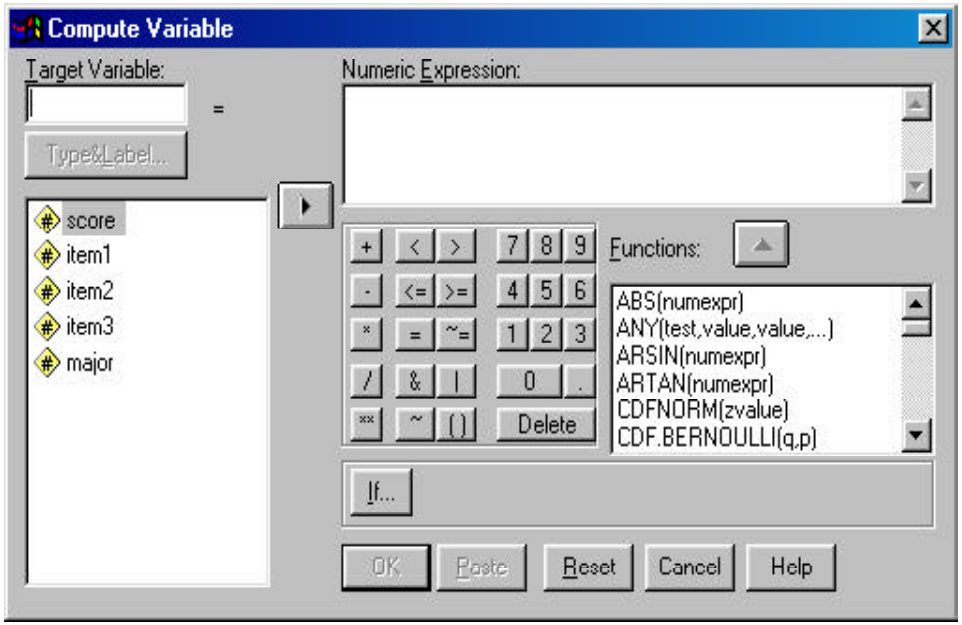
Common Data Transformations & Case Selection

Often we need to "modify" or "transform" the data that we enter before we can analyze it. Here are some of the most common transformations...



Compute Transform ➔ Compute

- compute statements can be used to create new variables
- the new variable will "appear" in the rightmost column in the data set
- be sure to put in labels, values and missing values for the new varible
- if a compute statment includes a variable with a missing value for a case, then the resuting variable also has a missing value

The name of the new variable goes in the "Target Variable" window

- be sure the variable name is no more than 8 characters

Put the formula for the new variable in the "Numeric Expression" window

- you can either type in the formula or use the point-n-click buttons
- the "Functions" do a variety of common transformations (dig around a bit)

Examples:

Transforming a score into a percentage

- the variable **score** holds the # correct out of 20
- we want to make a new variable called **perscore** that is a percentage score
- type **perscore** in the "Target Variable" window
- type **(score / 20) * 100** in the "Numeric Expression" window
- Click "OK"

Computing an average of several variables

- variables **item1**, **item2** & **item3** hold scores on those items
- we want to make a new variable called **avescore** that is the average of these
- type **avescore** in the "Target Variable" window
- type **(item1 + item2 + item3) / 3** in the "Numeric Expression" window
- Click "OK"

Recode

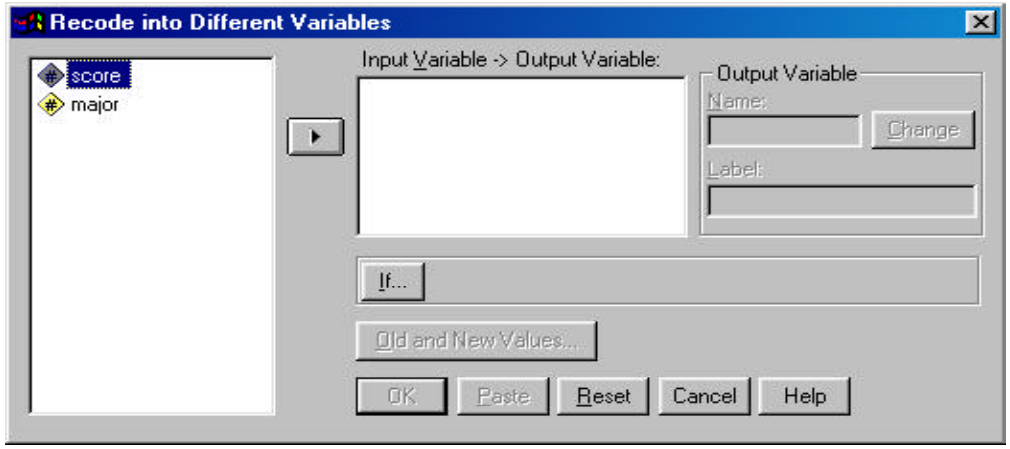
Transform ➔ Recode ➔ Into Different Variables

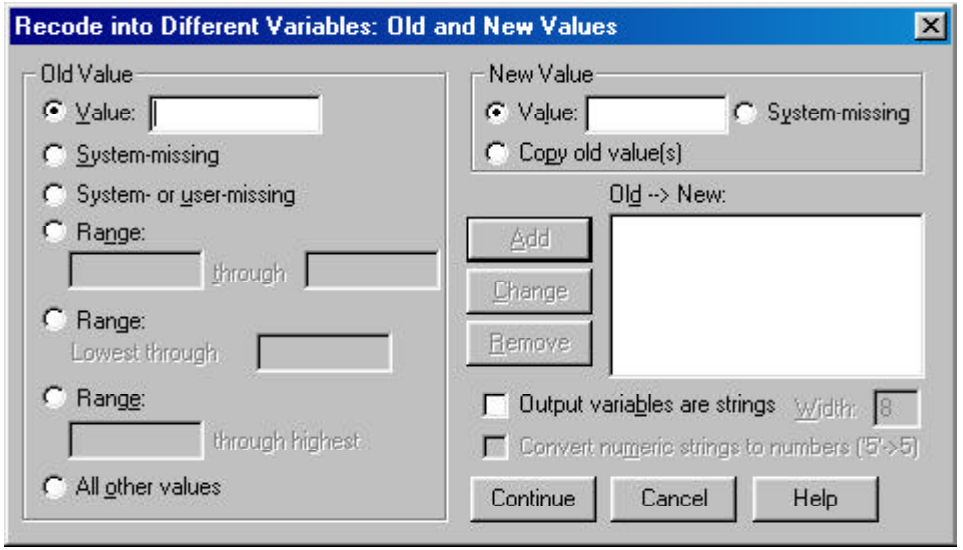
- recode statements can be used to create new "versions" of variables
- the new variable will "appear" in the rightmost column in the data set
- be sure to put in labels, values & missing values for the new varible
- if a recode statment includes a variable with a "systen missing" value (".") for a case, then the resuting variable will automatically have a missing value (.)
- if a recode statement includes a variable with a "user-missing" value for a case, they you must specify that "System- or user-missing values" should become a "System-missing value"

General procedure for using Recode into Differnt Variables

- highlight the "old" variable and move it into the "Input Variable ->Output Variable" window
- type the name of the new variable itnto the "Output Variable" window. You can add a variable label using the "Label" window.
- Click the "Change" button -- the new variable will appear in the "Input Variable ->Output Variable" window
- Click the "Old and New Variable" button -- the "Recode Into Different Variables Old and New Values" window will appear
- for each recoding, specify the old value and the new value

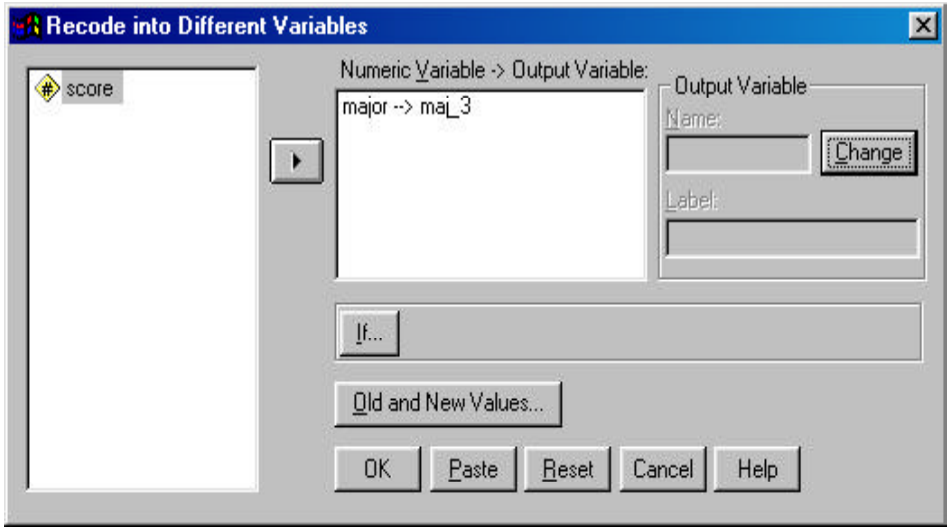
Examples of the most common types of recodes are on the next two pages.

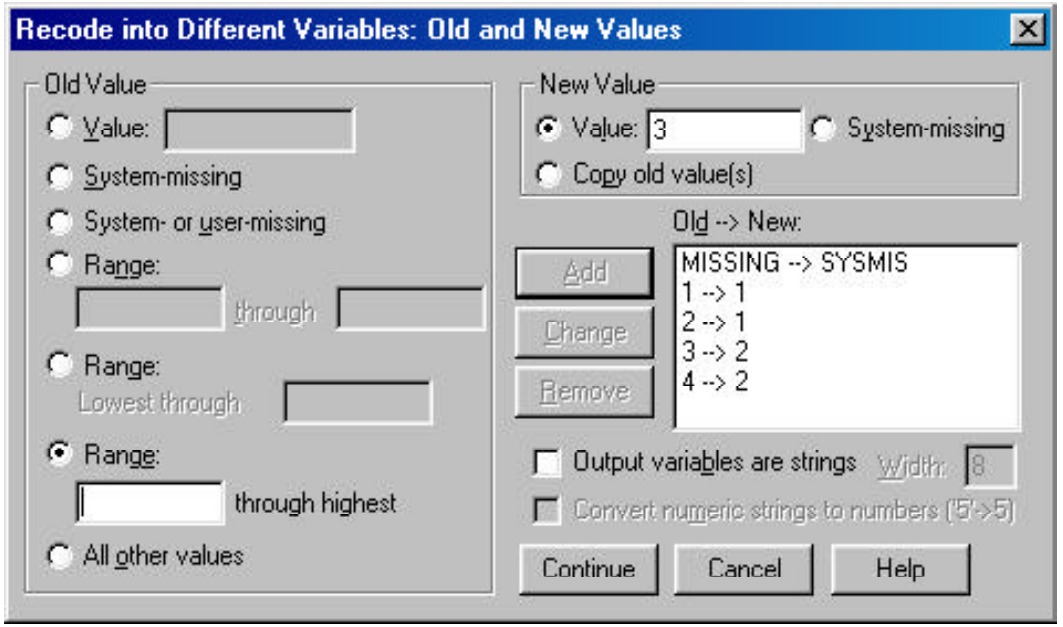




Combining categories of a qualitative variable

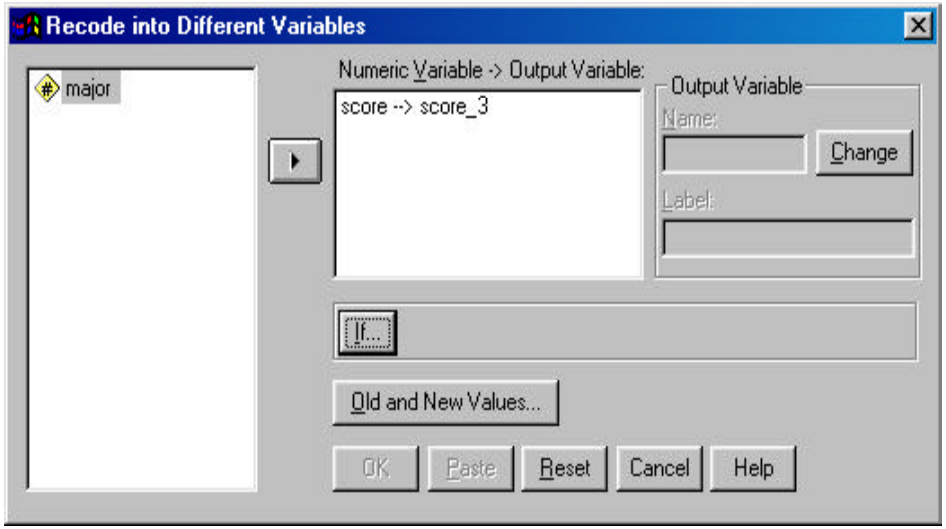
- The variable **major** has 5 categories 1 = psych, 2 = socio, 3 = business 4 = marketing and 5 = English (-9 means a missing value)
- We want a new variable **maj_3** that has three categories 1 = psych & socio, 2 = business and marketing and 3 = English
- How to do it...
  - In the "Recode into Different Variables" window
    - Highlight **major** & move to "Input Variable -> Output Variable" window
    - Type maj_3 in the "Name" box under "Output Variable" & click "Change"
    - Click the "Old and New Values" button
  - In the "Recode into different Variables: Old and New Values" window
    - Taking care of missing values
      - On the "Old Value" side -- click the radio button beside "System- or User-Missing"
        - On the "New Value" side -- click the radio button beside "System Missing"
        - "Add" can now be clicked - click it to move the transformation into the "Old --> New" window
    - Specifying recode values
      - On the "Old Value" side -- click the radio button beside "Value" and type a "1" into the window
        - On the "New Value" side -- click the radio button beside "Value" & type a "1" into the window
        - "Add" can now be clicked - click it to move the transformation into the "Old --> New" window
      - Repeatwith Old--> New values of 2=1, 3=2, 4=2 and 5=3 untill all are visiible in the window, then click "Continue"
- The windows below show the last step and the results (in the "Old--> New" window) of the previous steps.

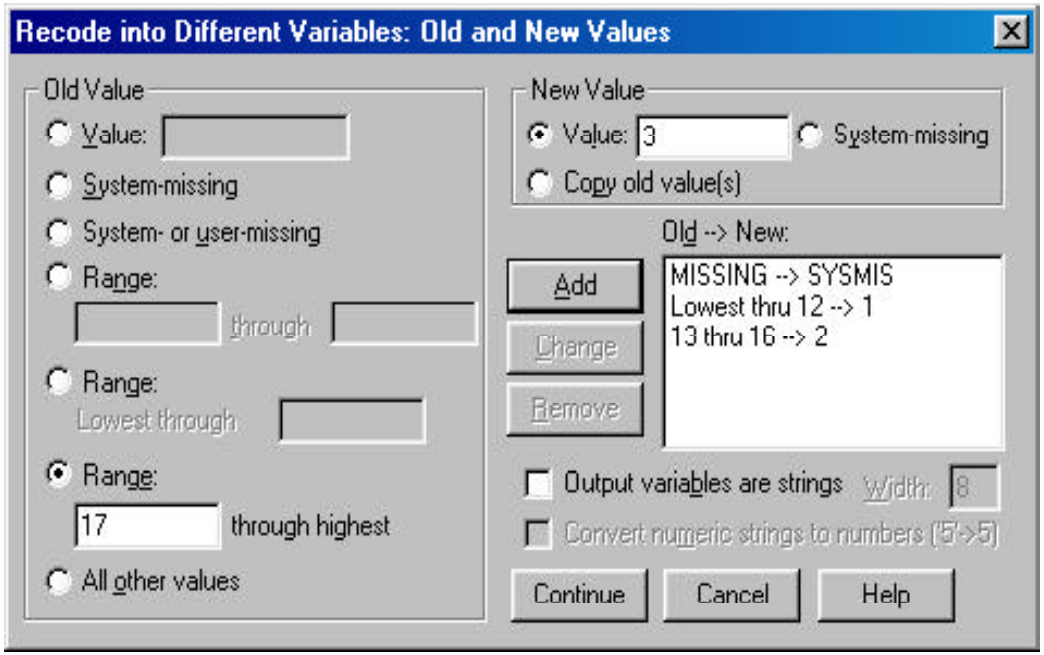




Making a quantitiative variable into a set of ordered categories

- the variable **score** holds the # correct out of 20
- we want to make a new variable score_3 that puts cases into one of three categories, based on their score value 0-12 = 1, 13-16 = 2 and 17-20 = 3
- For each value of the original variable we will enter that value and the corresponding value for the new variable
- For example
  - In the “Recode into different Variables” window -- Highlight **score** and move to “Input Variable -> Output Variable” window
  - Specify score_3 as the “Output Variable” and click “Change”
    - In the “Recode into different Variables: Old and New Values” window
      - On the “Old Value” side -- click the radio button beside “System- or User-Missing”
        - On the “New Value” side -- click the radio button beside “System-Missing”
      - On the “Old Value” side -- click the “Range” button just above “Lowest through” and type “12” into the window
        - On the “New Value” side -- click the radio button beside “Value” & type a “1” into the window -- then click “Add”
      - On the “Old Value” side -- clidk the “Range” button just above the two windows , type “13” into the left window and “16” into the right window
        - On the “New Value” side -- click the radio button beside “Value” & type a “2” into the window -- then click “Add”
      - On the “Old Value” side -- clidk the “Range” button just above “ through highest” and type “17” into the window
        - On the “New Value” side -- click the radio button beside “Value” & type a “3” into the window -- then click “Add”

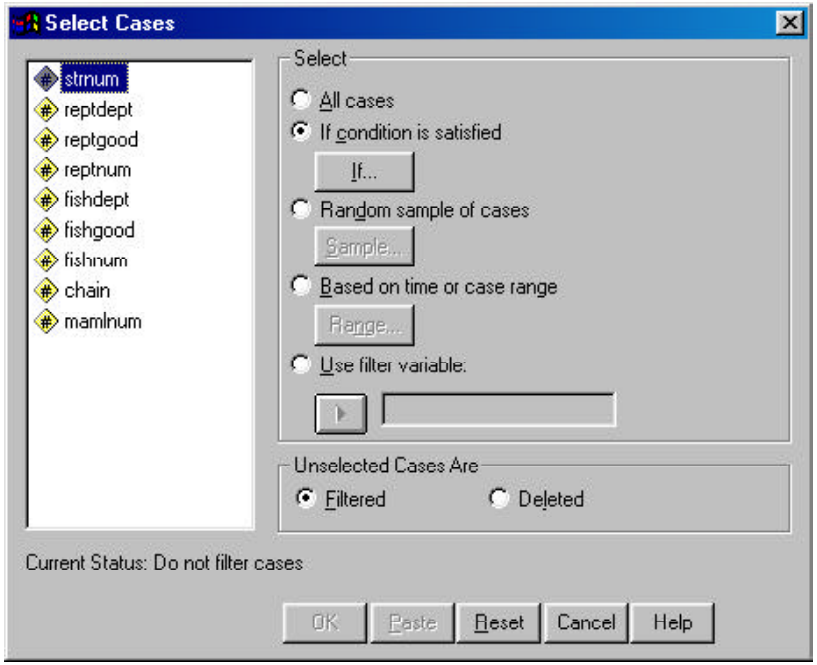




Data Selection

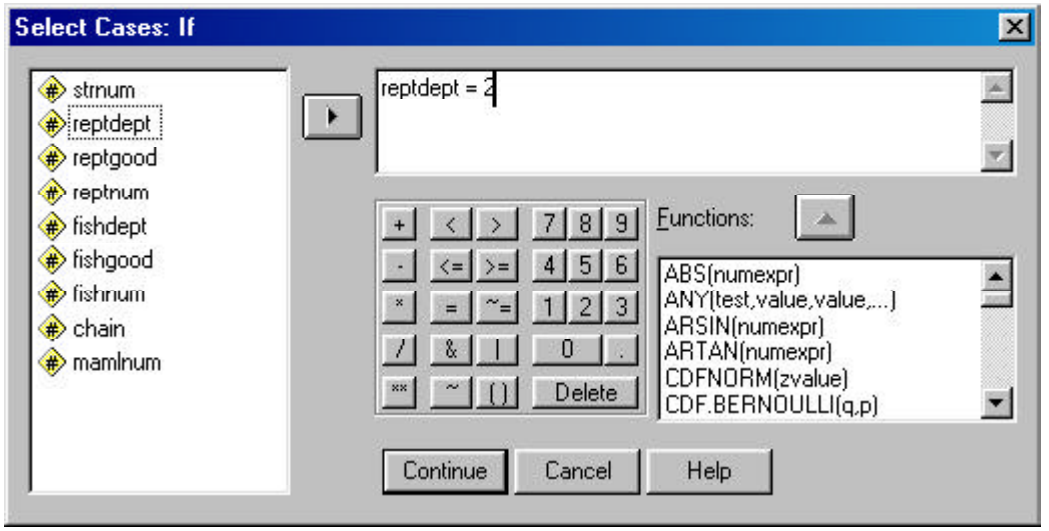
Sometimes we want to analyze only part of the data in the data set. For example, we might want to look only at those stores that had separate reptile departments. If so, we'd do the following...

Data ➔ Select Cases



0

Click the radio button beside "If condtion is satisfied". Thenclick the "IF" key.



Enter the variable name and the value of the condition you want.

SPSS will make a new colum that indicates which cases have that value of that variable -- **only those cases will be used for further analyses** (untill you turn it off by clicking the "All Cases" radio button in the "Select Cases" window.