Statistical Package for Social Science (SPSS)

Workshop Session 2 Advance

SPSS 19
Workshop October 12th 2011

The American University in Cairo
Department of University Academic Computing Technologies (UACT)

Table of contents

1 - SPSS Utilities Menu

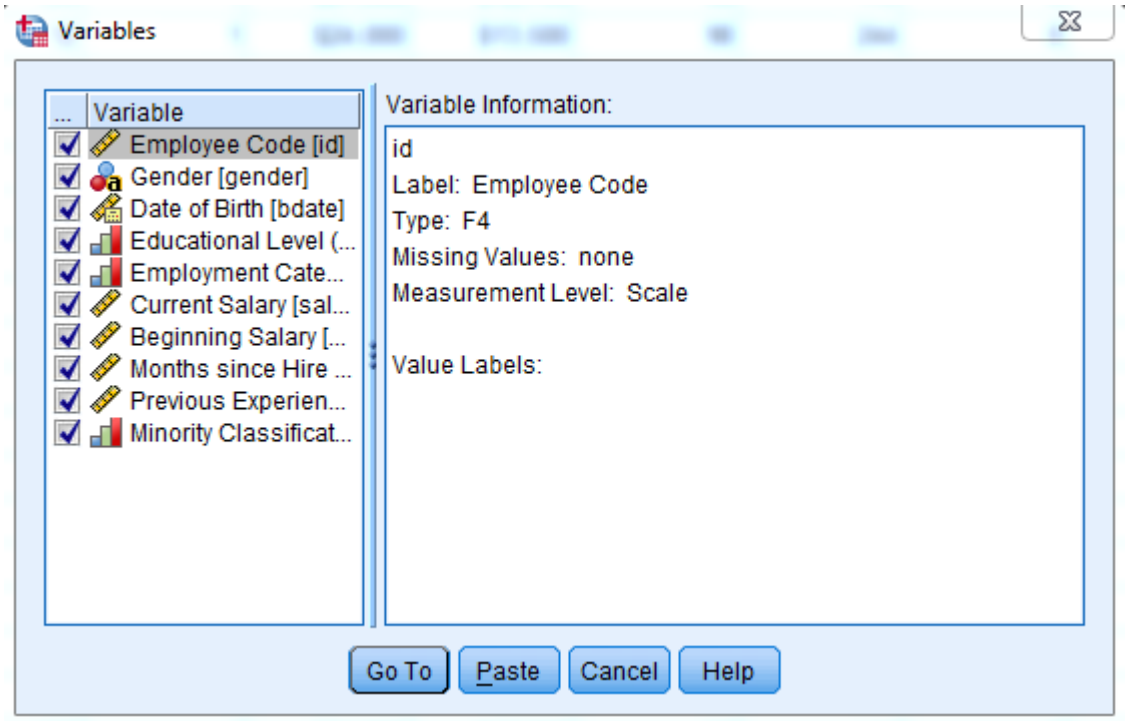
In the first session, we explored the SPSS interface, some basic data management and recodes, and some basic charting. In this second session, we'll explore on how to work with more complex data transformations such as combining variables and sub setting populations and work with some of the primary statistical functions.

1.0 - Utilities: Variable Information

Tucked quietly under the Utilities menu is an especially useful functions: Variables. You can use this function to get a snapshot of each variable in the file (Variables).

The Variables function provides all the information about each variable in your data file, including any categorical codes and their value labels.

From the menu, select Utilities > Variables. (Figure 1)



(Figure 1 from "Employee data Sample")

-In the variable list, click jobcat. Notice that the Variable Information pane displays the variable name, label, defined missing value, measurement level, values, and value labels.

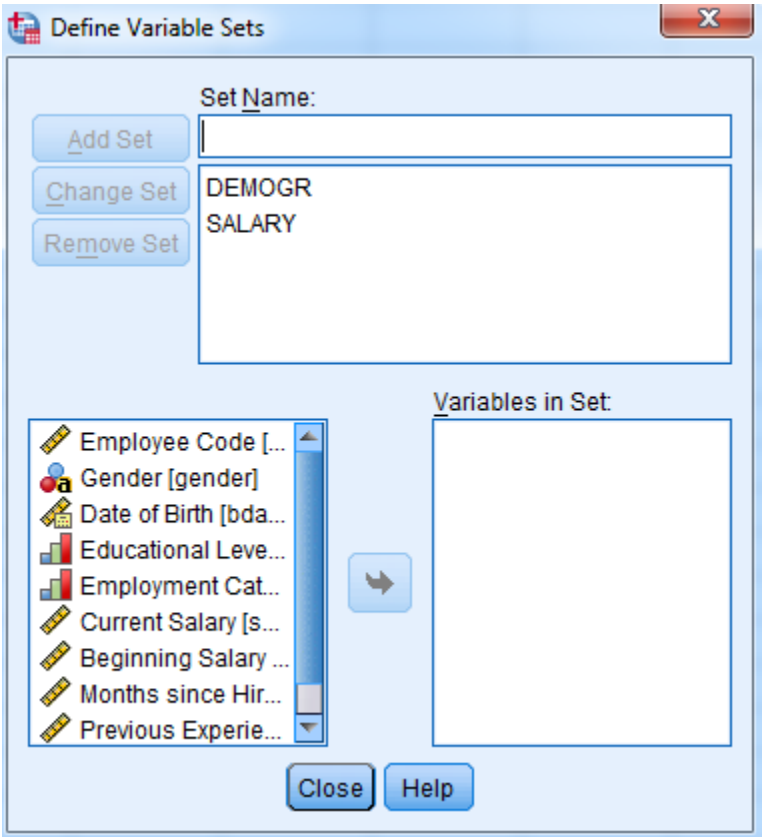
- Click salary. This value is a scale variable (continuous) and so has no value labels.

- The Go To button takes you to the specific variable within a selected case or to the variable in the first case if no case is selected.

1.1- Defining variable sets

Define Variable Sets creates subsets of variables to display in the Data Editor and in dialog box variable lists. Defined variable sets are saved with SPSS Statistics data files.

Variables in Set. Any combination of numeric and string variables can be included in a set. The order of variables in the set has no effect on the display order of the variables in the Data Editor or in dialog box variable lists. A variable can belong to multiple sets.

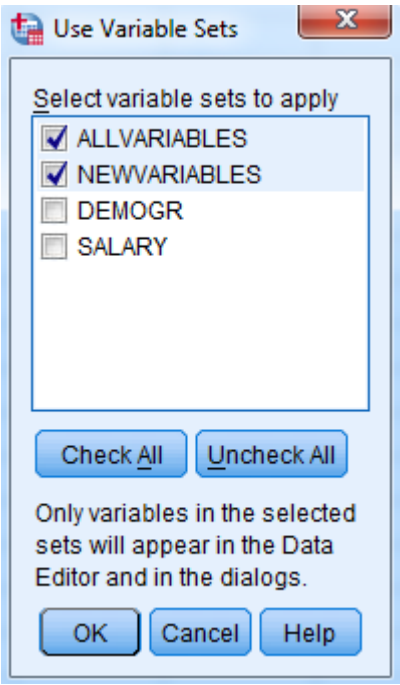


Steps

1) To define variable sets, from the menus choose:
2) Utilities > Define Variable Sets...
3) Select the variables that you want to include in the set.
4) Enter a name for the set.
5) Click Add Set.

1.1.1- To use a variable set

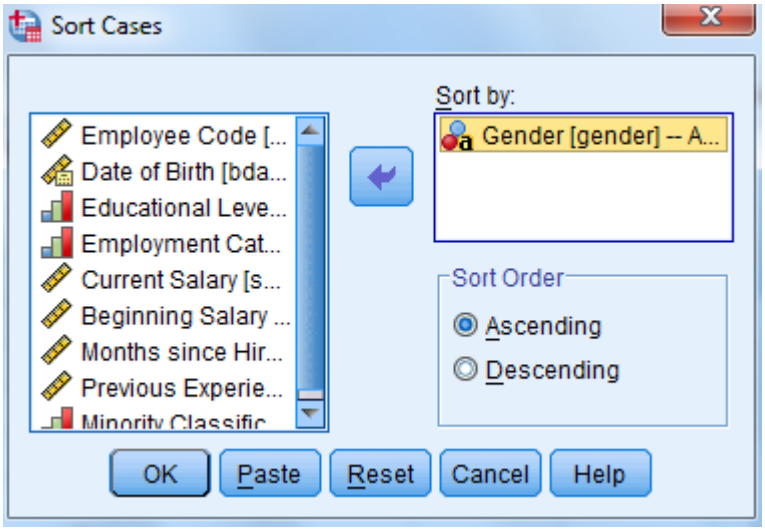
1) From the menus choose:
2) Utilities > Use Variable Sets...



2- SPSS Data Arrangement

2.0- Sort Cases

This dialog box sorts cases (rows) of the data file based on the values of one or more sorting variables. You can sort cases in ascending or descending order.



Data Menu > Sort Cases

• If you select multiple sort variables, cases are sorted by each variable within categories of the preceding variable on the Sort list. For example, if you select gender as the first sorting variable and minority as the second sorting variable, cases will be sorted by minority classification within each gender category.

• The sort sequence is based on the locale-defined order (and is not necessarily the same as the numerical order of the character codes). The default locale is the operating system locale. You can control the locale with the Language setting on the General tab of the Options dialog box (Edit menu).

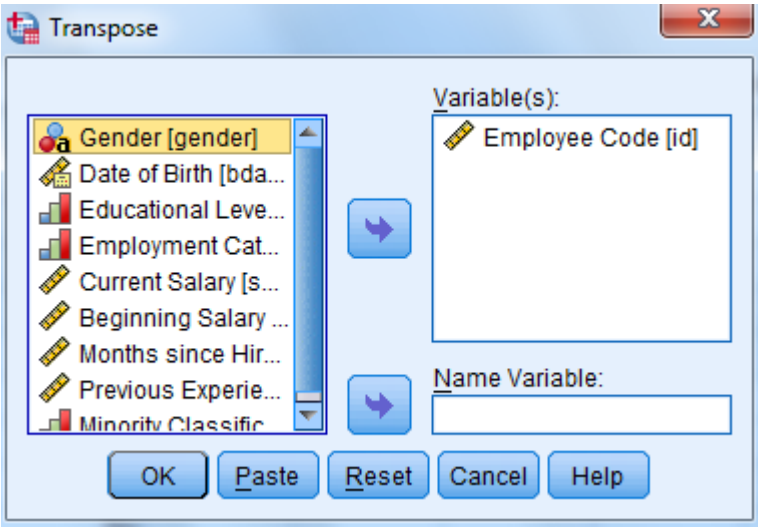
2.1- Sort Variables

You can sort the variables in the active dataset based on the values of any of the variable attributes (e.g., variable name, data type, measurement level)

Data Menu > Sort Variables

2.2- Transpose

Transpose creates a new data file in which the rows and columns in the original data file are transposed so that cases (rows) become variables and variables (columns) become cases. Transpose automatically creates new variable names and displays a list of the new variable names.



Data Menu > Transpose

2.3- Split File

Split File splits the data file into separate groups for analysis based on the values of one or more grouping variables. If you select multiple grouping variables, cases are grouped by each variable within categories of the preceding variable on the Groups Based On list.

Note: Any changes to the arrangement of the cases may lead to the cancelation of the split file.

2.4- Merge

You can merge data from two files in two different ways. You can:

• Merge the active dataset with another open dataset or SPSS Statistics data file containing the same variables but different cases.

• Merge the active dataset with another open dataset or SPSS Statistics data file containing the same cases but different variables.

To Merge Files, from the menus choose:

Data > Merge Files > Select Add Cases or Add Variables.

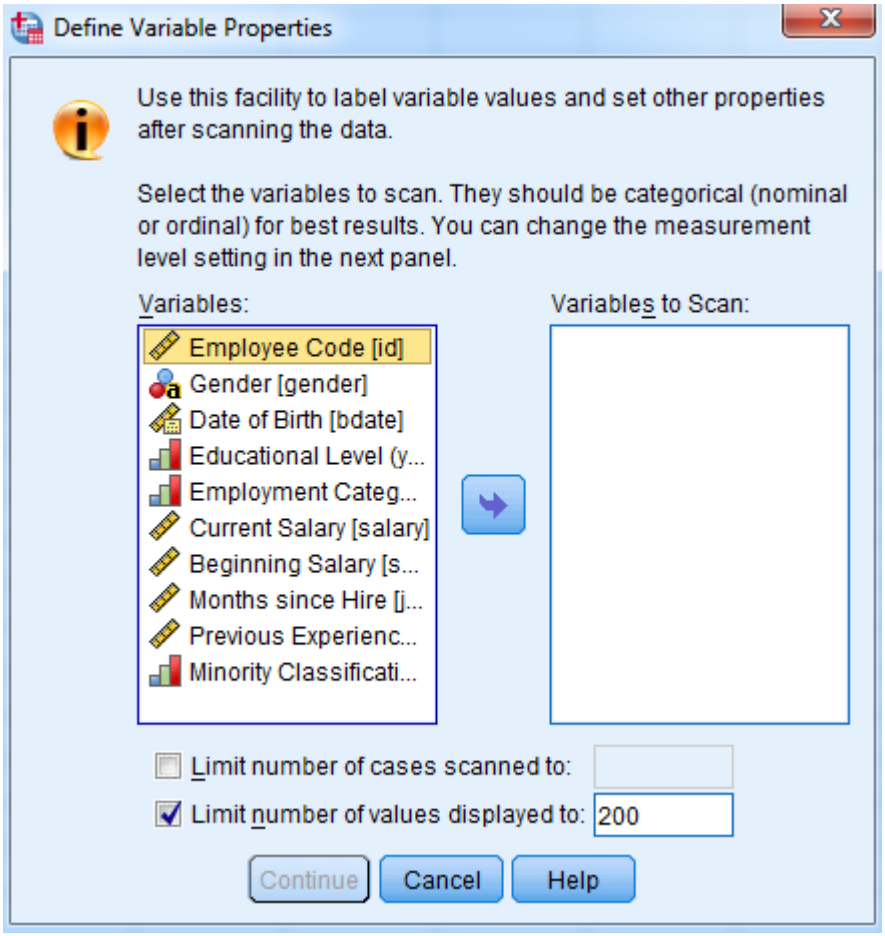
2.5 - Data Preparation

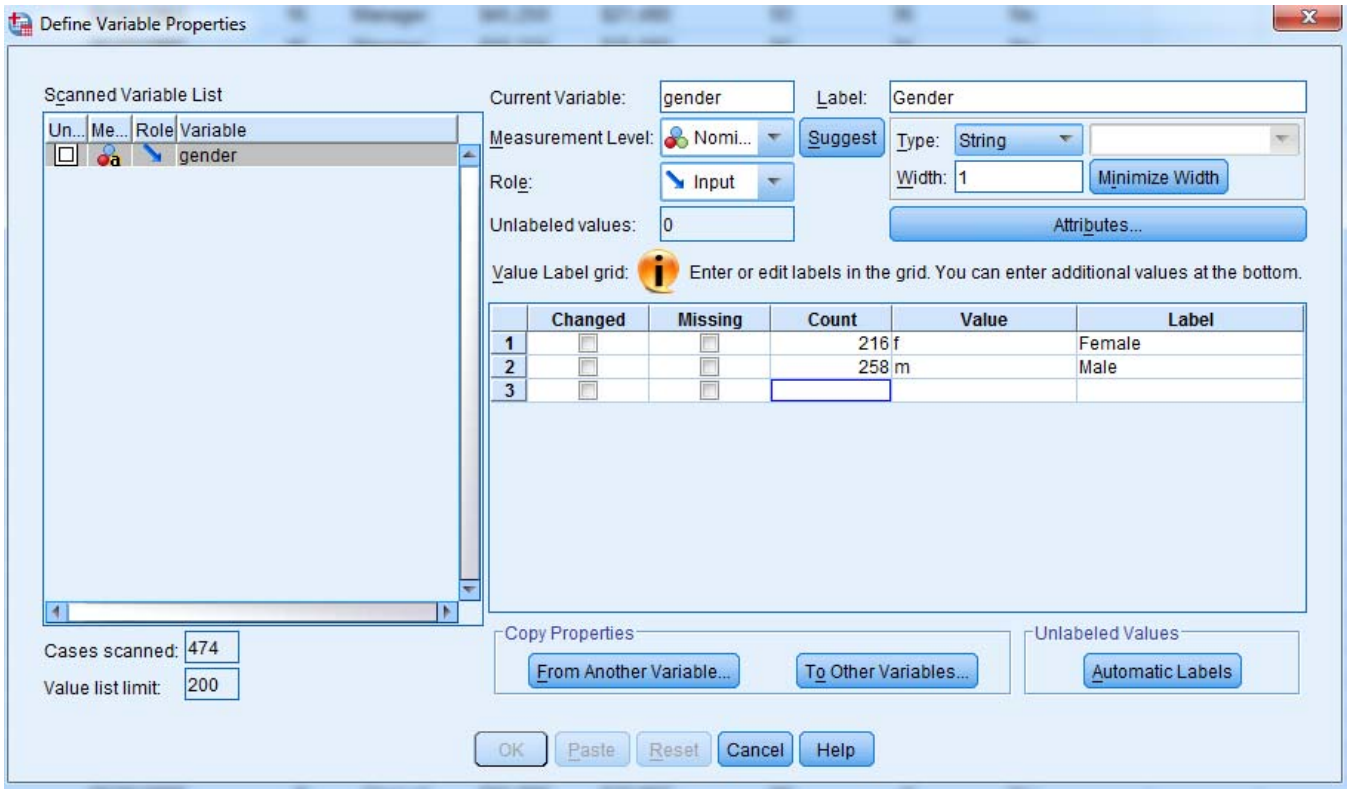
Define Variable Properties is designed to assist you in the process of assigning attributes to variables, including creating descriptive value labels for categorical (nominal, ordinal) variables.

• Scans the actual data values and lists all unique data values for each selected variable.

• Identifies unlabeled values and provides an "auto-label" feature.

• Provides the ability to copy defined value labels and other attributes from another variable to the selected variable or from the selected variable to multiple additional variables.




Define Variable Properties
Scanned Variable List
Un... Me... Role Variable
gender
Current Variable: gender
Label: Gender
Measurement Level: Nomi...
Suggest
Type: String
Width: 1
Minimize Width
Role: Input
Unlabeled values: 0
Attributes...
Value Label grid:
Enter or edit labels in the grid. You can enter additional values at the bottom.
Changed Missing Count Value Label
1 216 f Female
2 258 m Male
3
Cases scanned: 474
Value list limit: 200
Copy Properties
From Another Variable...
To Other Variables...
Unlabeled Values
Automatic Labels
OK Paste Reset Cancel Help

3- SPSS Data Transformations

3.1- Compute Data

SPSS provides a number of functions you can use in computing new variables. In some cases, you might want to calculate new variables based on values in existing variables and some arithmetic function like multiplying or dividing. For example, if you have a variable that contains an annual salary, you might want to calculate a monthly salary. To create the new variable, you use the Compute function.

1) In the Data window, select from the menu Transform -> Compute

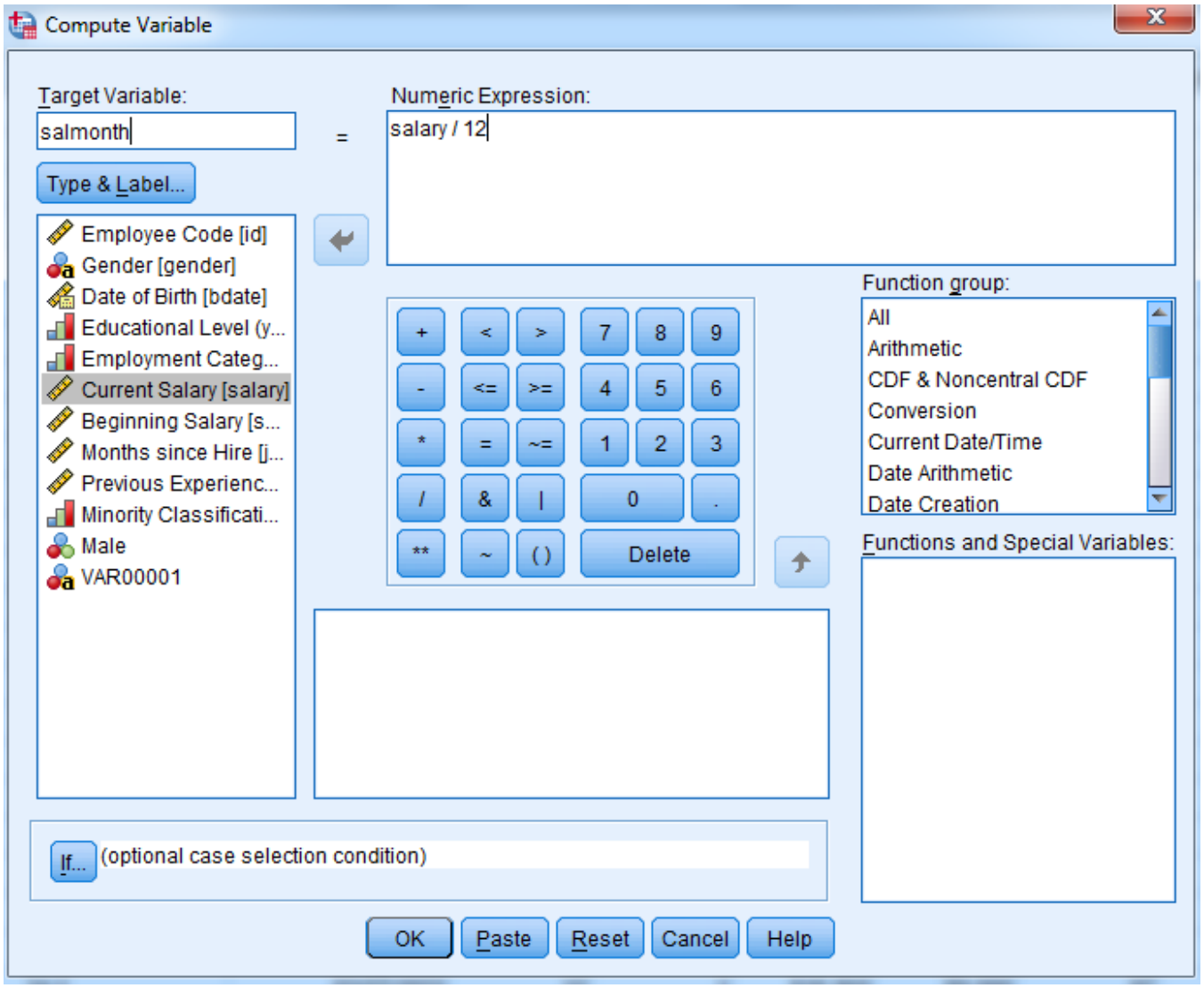
2) In the Target Variable field, type: “salmonth”

3) Click Type & Label.

4) In the Label field, type: Average monthly salary

5) In the Compute Variable window, select Current Salary and move it to the Numeric Expression pane by clicking the right arrow.

6) In the Numeric Expression pane, click the cursor after salary and type:



Click OK. The Compute Variable window closes and the new variable is displayed in the Data window.

3.2- Conditional expressions

In some cases, you might want to look at only a specific subset of your data. Say you want to increase the salary to only female clerical staff. To identify these staff, you'll calculate a new binary variable (one that has only two values) using the IF statement to set the condition.

1) From the menu, select Transform > Compute.

2) In the Target Variable field type: femclerknewSal

3) Click Type & Label.

4) In the Label field type: Female Clerical New Salary

5) Click Continue.

6) Select all the text in the Numeric Expression field and delete it.

7) Click if to open the Compute Variable: If Cases window

8) Select “Include if case satisfies condition”.

9) Double-click Gender to move it to conditions field.

10). Click after Gender in the conditions field and type: = ‘f’

Note: Whenever you create a condition, you must use the actual values in the variable, not their labels. Thus, setting a condition to gender = “Female” would not select any cases.

11) Click after ‘f’ and type a space.

13. Using the keypad in the Compute Variable window, click &. You use the ampersand to add a second condition.

14. From the field list, double-click Employment category to move it to the calculation pane.

15. In the calculation pane, type: = 1

Note that you don’t use quotation marks this time because is a numeric variable.

16. Click Continue.

17. Click OK. The new variable appears in the Data window. Scroll through the records to see how the values in the new variable. Notice that cases where gender is not female and job category is not manager have only a period, indicating a missing value. Only those cases where gender is female and jobcat is manager contain value.

3.3- Arithmetic Operations

+ (Plus)

< (Less Than)

> (Greater Than)

- (Minus)

<= (Less than or equal)

>= (Greater than or equal)

* (Multiple)

= (Equal)

~= (Not Equal)

/ (Division)

& (And)

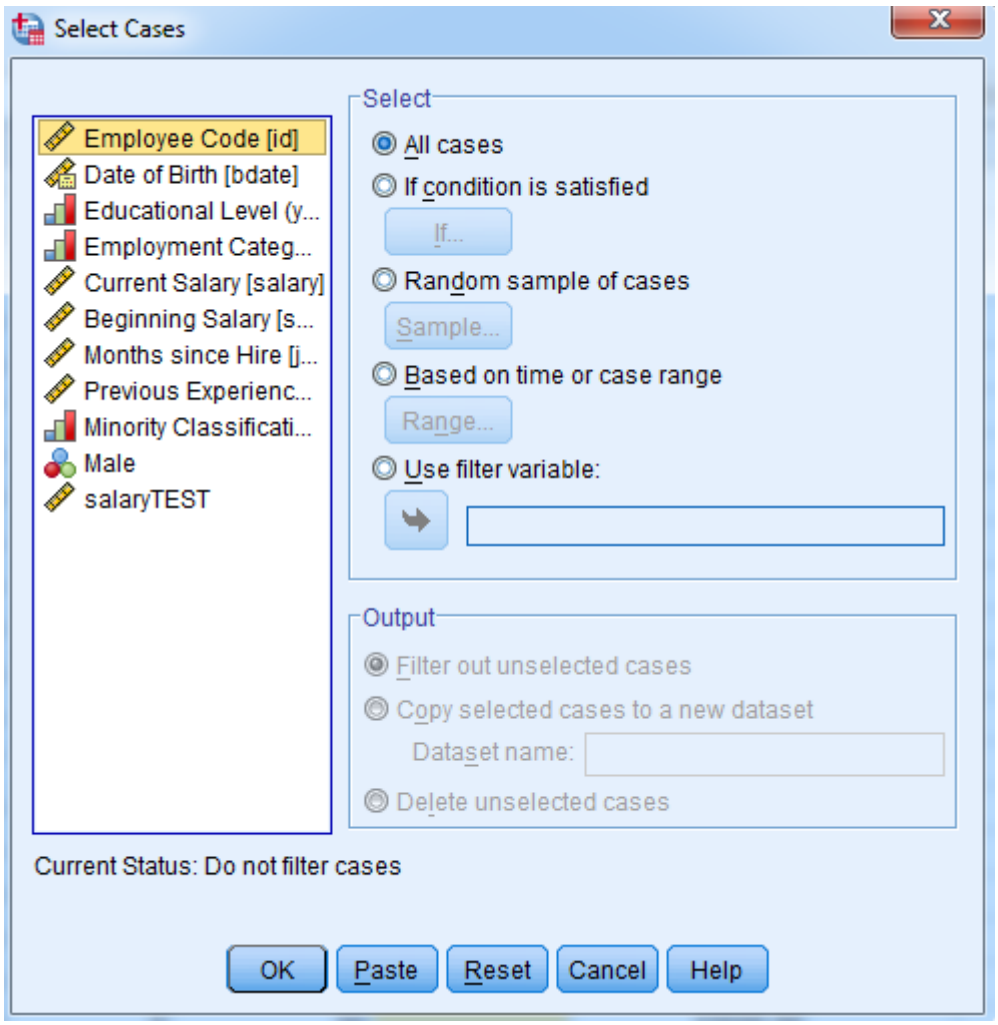
| (Or)

** (Exponentiation)

3.4- Creating Subsets

In some instances, you might want to use only part of the file in an analysis. For example, you might want to look at changes in income among single working mothers. Or you might want to consider only staff born before a specific date. To select a subset of the cases in your file,

1) From the menu, select Data > Select Cases



2) Select “If condition is satisfied” by clicking its radio button.

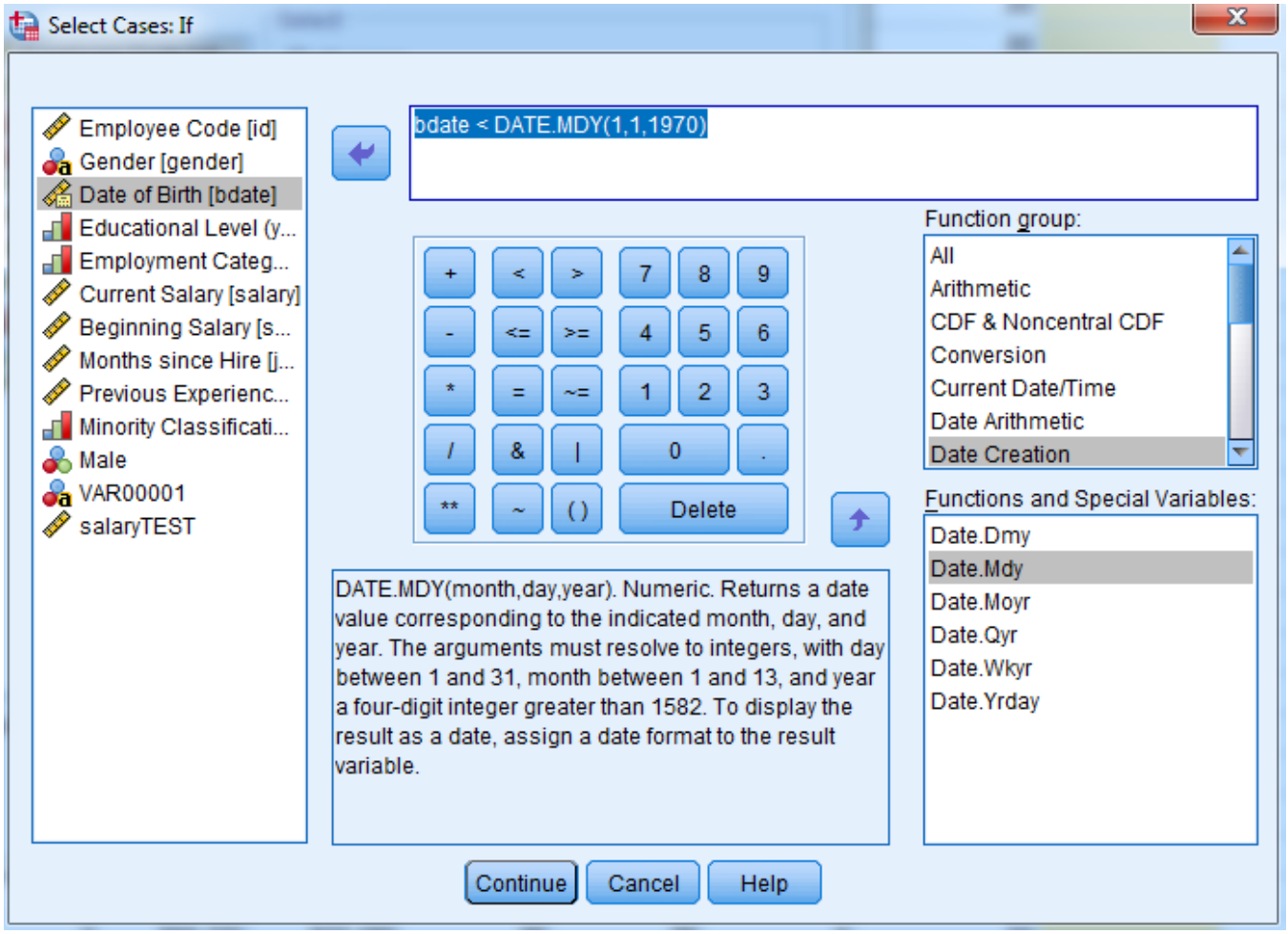
3) Click “If....” Notice that the Select Cases: If window looks exactly like the If window you used in the earlier compute procedures.

4) From the variable list, double-click Date of Birth.

5) Click the cursor anywhere after bdate in the calculation pane.

6) Type (or select from the keypad): <

7) Scroll through the Function menu and double-click DATE.MDY(month,day,year).



In the next step, you'll set the date criterion. SPSS adds the function to the calculation pane, substituting question marks to indicate that you need to specify the values.

8) Select the first question mark and type: 1

9) Select the second question mark and type: 1

10) Select the third question mark and type: 1970

11) Click Continue.

12) Click OK. Notice that many of the records are marked with a diagonal line through the record number. These cases are excluded from any further calculations until you specifically include them again.

13) To see the effect of the subset selection, right click the heading for bdate.

14) From the pop-up menu, select Sort Ascending. Notice that all employees born before 1970 are selected, except for the person with the missing date of birth. In the next step, you'll instruct SPSS to include all cases.

15) From the menu select Data > Select Cases.

16) Select All Cases by clicking its radio button

3.5- Recode into Same Variables

The Recode into Same Variables dialog box allows you to reassign the values of existing variables or collapse ranges of existing values into new values. For example, you could collapse salaries into salary range categories.

You can recode numeric and string variables. If you select multiple variables, they must all be the same type. You cannot recode numeric and string variables together.

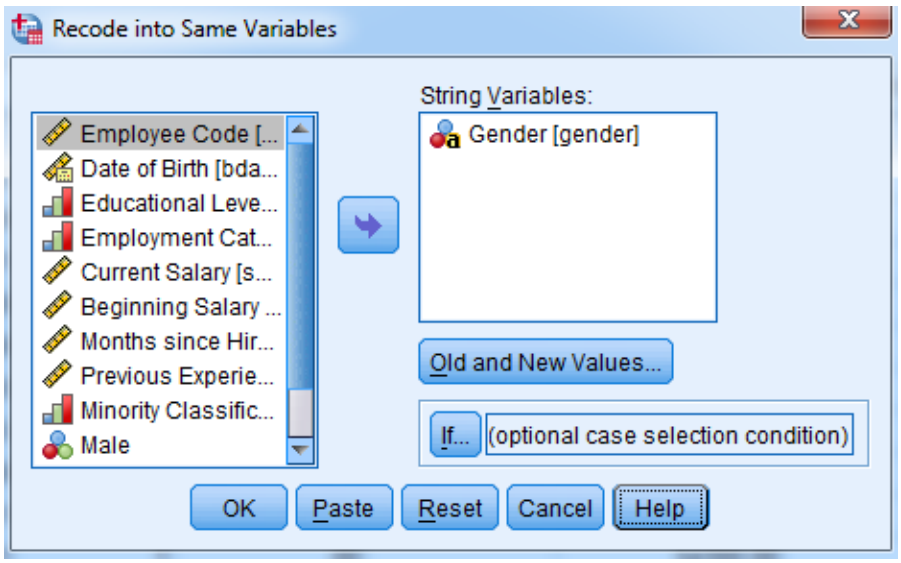
To Recode Values of a Variable

1) From the menus choose:

2) Transform > Recode into Same Variables...

3) Select the variables you want to recode. If you select multiple variables, they must be the same type (numeric or string).

4) Click Old and New Values and specify how to recode values. Optionally, you can define a subset of cases to recode.



Recode into Different Variables

The Recode into Different Variables dialog box allows you to reassign the values of existing variables or collapse ranges of existing values into new values for a new variable. For example, you could collapse salaries into a new variable containing salary-range categories.

• You can recode numeric and string variables.

• You can recode numeric variables into string variables and vice versa.

• If you select multiple variables, they must all be the same type. You cannot recode numeric and string variables together.

To Recode Values of a Variable into a New Variable

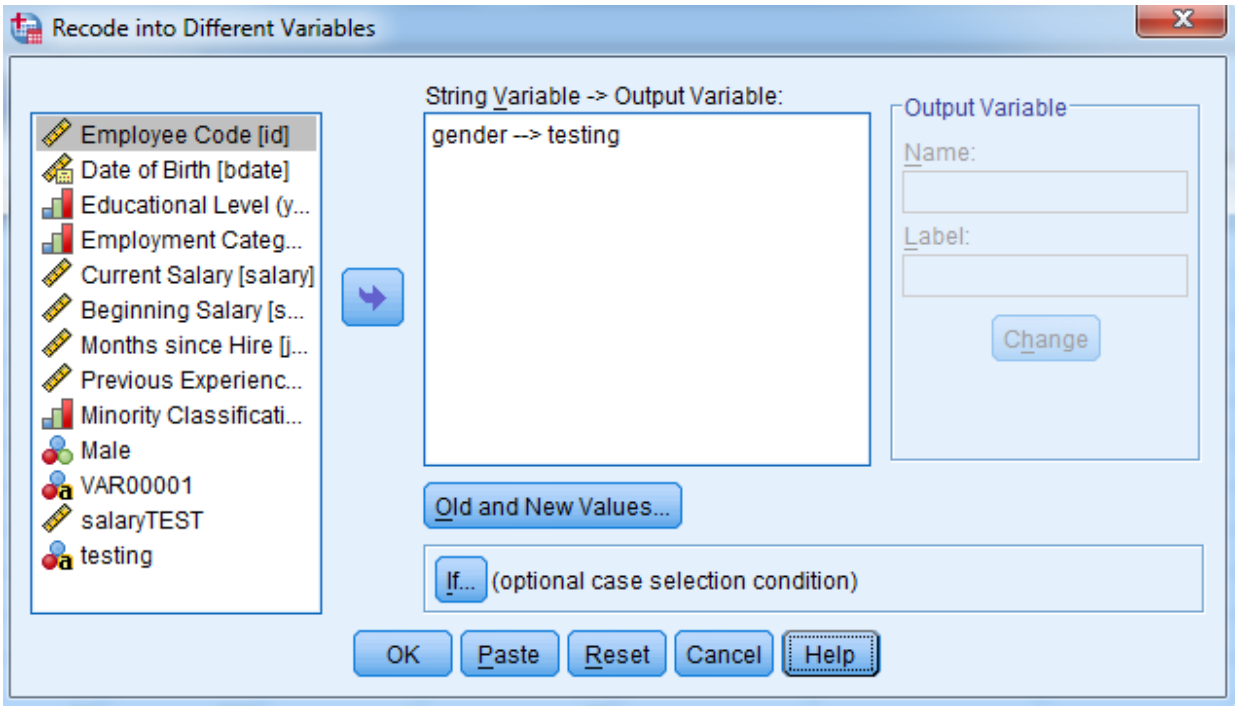
1) From the menus choose:

2) Transform > Recode into Different Variables...

3) Select the variables you want to recode. If you select multiple variables, they must be the same type (numeric or string).

4) Enter an output (new) variable name for each new variable and click Change.

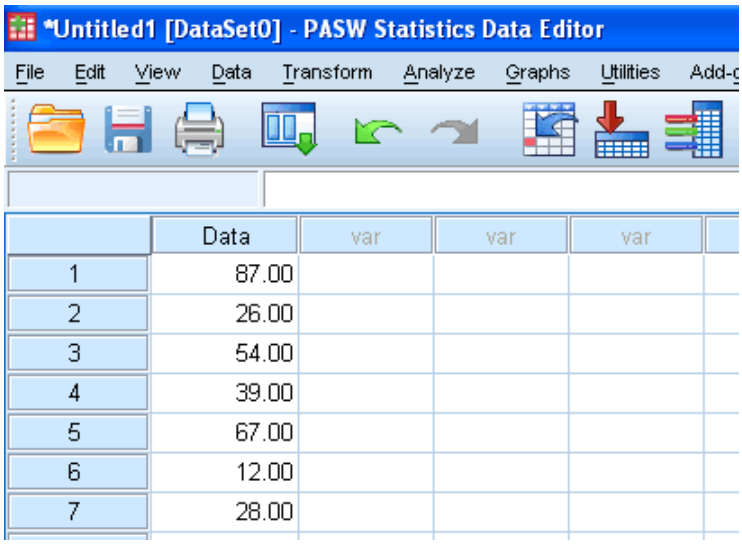
5) Click Old and New Values and specify how to recode values. Optionally, you can define a subset of cases to recode.



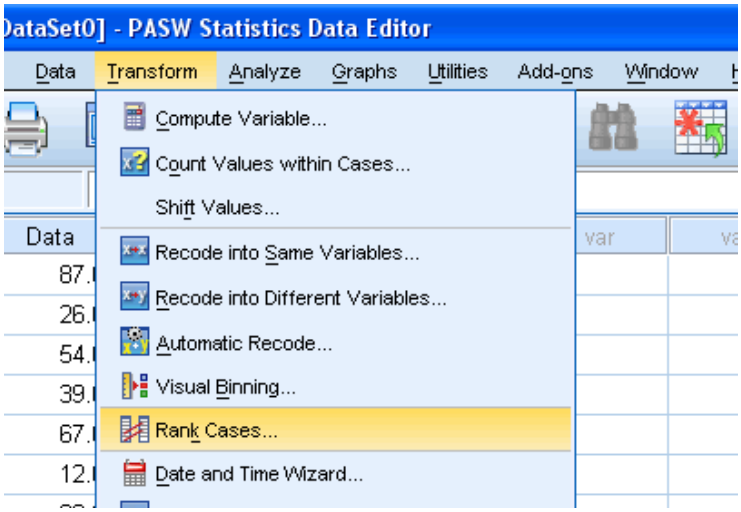
3.6 Ranking Data

Ranking is used to recode the data into their rank ordering from smallest to largest or largest to smallest. We will demonstrate this by entering in some data and ranking it in SPSS.

1) Your data should end up looking like the following (we have named the variable "Data"):



2) Click on Transform > Rank Cases... in the top menu.



3) Click on the Data variable in the left hand box and click the  button to move it to the "*Variable(s):*" box.

Reference

An icon next to each variable provides information about data type and level of measurement.

Measurement Level	Data Type			
	Numeric	String	Date	Time
Scale (Continuous)		n/a		
Ordinal				
Nominal