

Computer
Introduction to
SPSS

Second Stage
2018 -2019

Doctor
Salema Sultan Salman

Introduction to the SPSS software

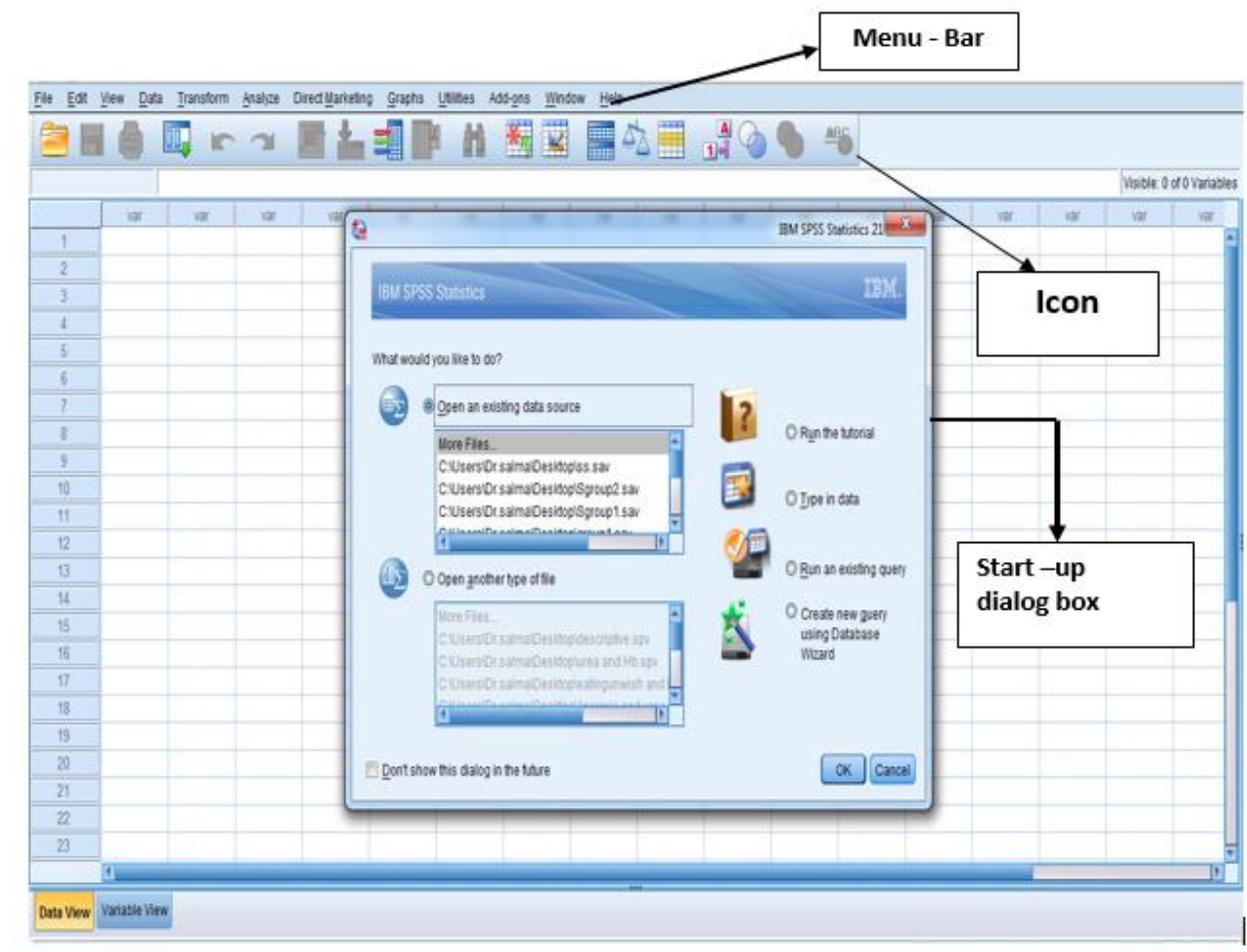
What is SPSS?

SPSS (Statistical Package for the Social Sciences) is a versatile and responsive program designed to undertake a range of statistical procedures.

1-Open SPSS

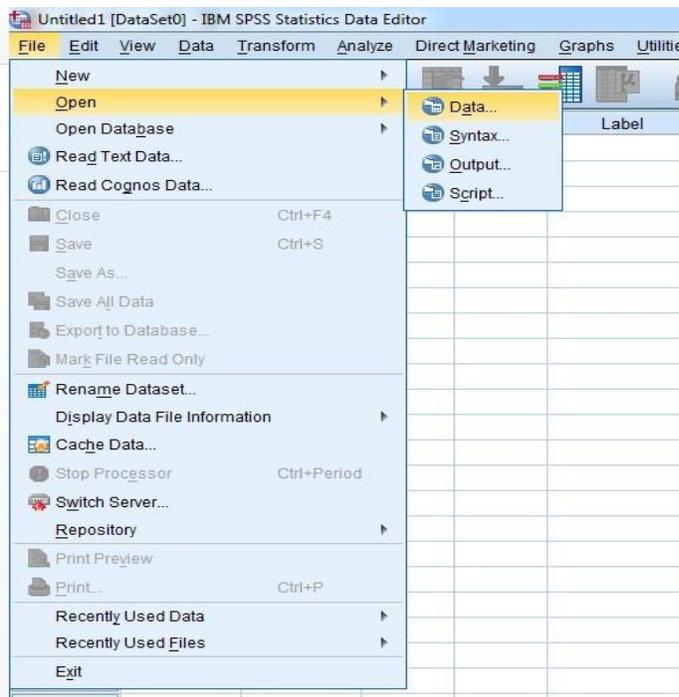


Click on the **Start menu** () > **All Programs** > **IBM SPSS Statistics** > **IBM SPSS Statistics 21** (or whatever is the latest version number) to open the SPSS program. When you first open SPSS on your computer, you should see something that looks similar to following screenshot:



SPSS automatically assumes that you want to open an existing file, and immediately opens a dialogue box to ask which file you'd like to open. It'll make it easier to navigate the interface and windows in SPSS if we open a file. It can be handy to know how to open a dataset once you're already in SPSS, using the drop down menus. How to open a file using the menus:

Click on the **File** menu at the top left of your Data Editor window, then **Open > Data ...**:



Here you can navigate to your working folder and open your data file.

Navigating the SPSS interface

Data editor window

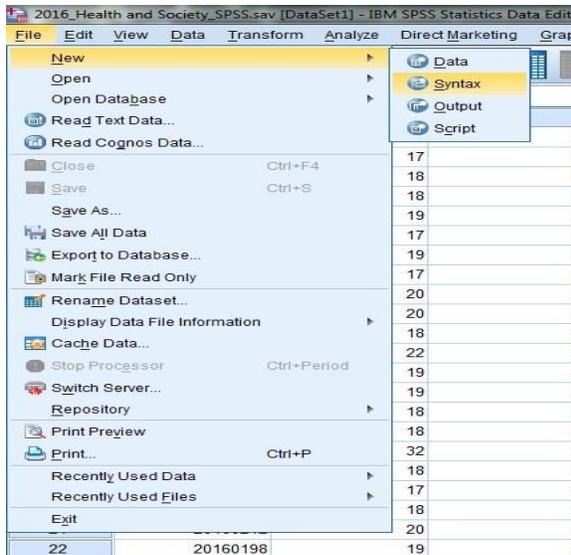
This is the window where you can see your data, and information about the variables in your dataset. It is also possible to change your data in this window. There are two 'views' for the Data Editor window:

- 1) **Data view** – you can see the actual data in your dataset for each record and each variable; and
- 2) **Variable view** – this gives a summary of each variable in your dataset, including the variable name, type, various properties of the way in which the data are stored, any label(s) for the variable itself and variable values (such as value labels for categories of sex, which in the dataset may be represented as 1 and 2, relating to male and female).

<p>Name: - Name of the variable. It is your own choice, but make it understandable and do not use numbers or symbols as the first letter since SPSS will not accept it. Moreover, you cannot use spaces in the name. For example: "edu_level"</p>		
<p>Type: - Indicates the variable type. The most common is Numeric (only accepts numerical data, for example age or number of children) and String (also accepts letters, e.g. for qualitative questions). Typically, all responses in a questionnaire are transformed into numbers. For example: "Man"=0 and "Woman"=1, or "Non-smoker"=1, "Ex-smoker"=2 and "Current smoker"=3.</p>		
 Scale	 Nominal	 Ordinal
<p>Width Corresponds to the number of characters that is allowed to be typed in the data cell. Default for numerical and string variables is 8, which only needs to be altered if you want to type in long strings of numbers or whole sentences.</p>		
<p>Decimals Default is 2 for numerical variables and will automatically be displayed as .00 in the data view, if not otherwise specified.</p>		
<p>Label The description of the variable. Use the question that the variable is based upon or something else accurately describing the variable. For example: "What is your highest level of education?"</p>		
<p>Values Here you can add labels to each response alternative. For example: For the variable gender, "Men" are coded as 0 and "Women" are coded as 1. Through the option Values you tell SPSS to label each number according to the correct response. Next to Value (below Value Labels), type in "0" and next to Label, type in "Men". Then click Add. Next to Value (below Value Labels), type in "1" and next to Label, type in "Women". Then click Add.</p>		
<p>Missing By default, missing values will be coded as "." (dot) for numerical variables in the data set. For missing values in String variables, cells will be left blank.</p>		

Syntax editor window

A second window in the SPSS interface is the Syntax Editor window. SPSS doesn't automatically open a syntax file for us, so let's open one now. Click on the **File** menu in the top left corner of your screen, and select **New > Syntax:**



The syntax window is where we can run commands, such as opening files, editing and managing data, undertaking statistical procedures and tests, and saving files. The Syntax Editor window is essentially a file that we can use to record and save everything that we do for a particular piece of analysis, or when managing/editing a dataset.

File types

SPSS uses three types of files with different functions and extensions:

Type	Extension	Content
Data set	.sav	Data and variables
Syntax	.sps	Commands and comments
Output	.spv	Results

Output viewer window

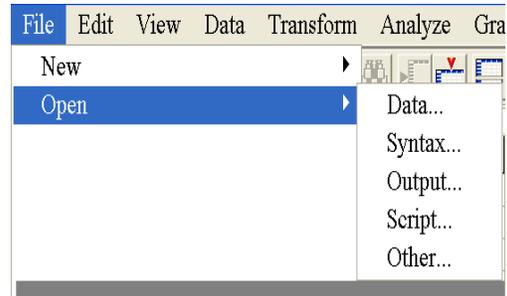
Any time you do anything in SPSS – even just opening a file – SPSS will document it in the Output Viewer window. The output viewer window will keep a record of any and all commands you give SPSS (e.g., open a file, save a file, etc.), and is also the window in which you can view the results of any data/statistical procedures undertaken. You can see in the screenshot below, that SPSS has made a record in the Output file that you opened a particular SPSS file:

SPSS Menus and Icons

Now, let's review the menus and icons.

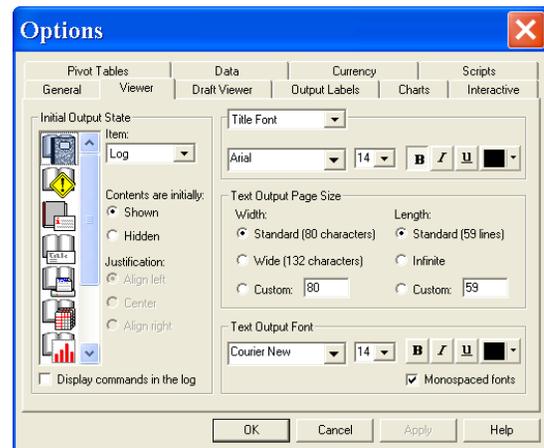
✓ Review the options listed under each menu on the Menu Bar by clicking them one at a time. Follow along with the below descriptions.

File includes all of the options you typically use in other programs, such as open, save, exit. Notice, that you can open or create new files of multiple types as illustrated to the right.



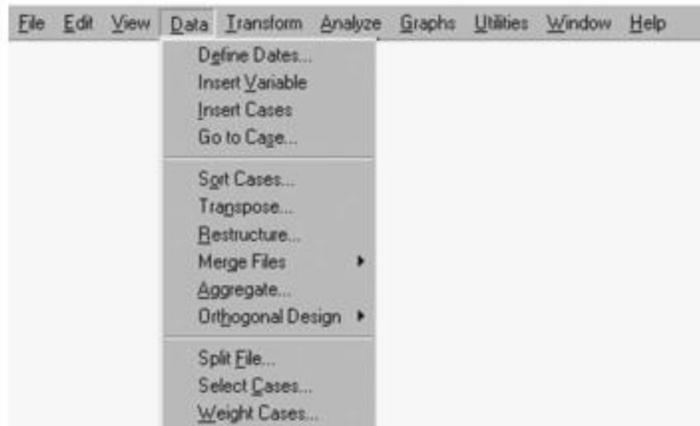
Edit includes the typical cut, copy, and paste commands, and allows you to specify various options for displaying data and output.

✓ Click on **Options**, and you will see the dialog box to the left. You can use this to format the data, output, charts, etc. These choices are rather overwhelming, and you can simply take the default options for now. The author of your text (me) was too dumb to even know these options could easily be set.



View allows you to select which toolbars you want to show, select font size, add or remove the gridlines that separate each piece of data, and to select whether or not to display your raw data or the data labels.

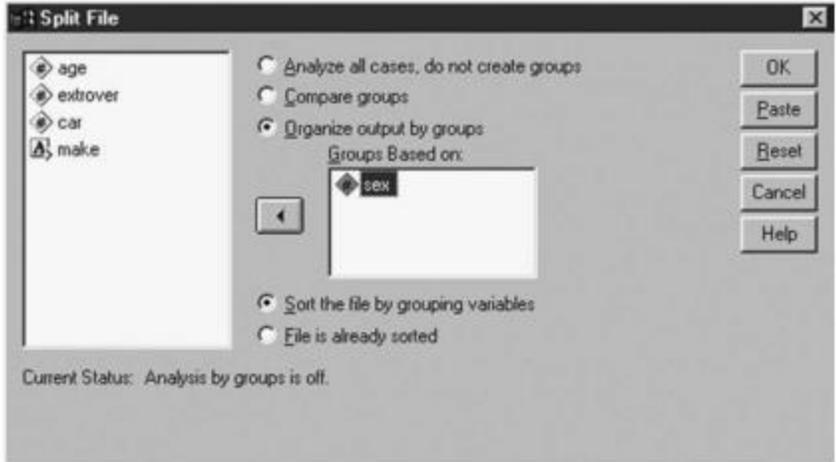
Data allows you to select several options ranging from displaying data that is sorted by a specific variable to selecting certain cases for subsequent analyses.



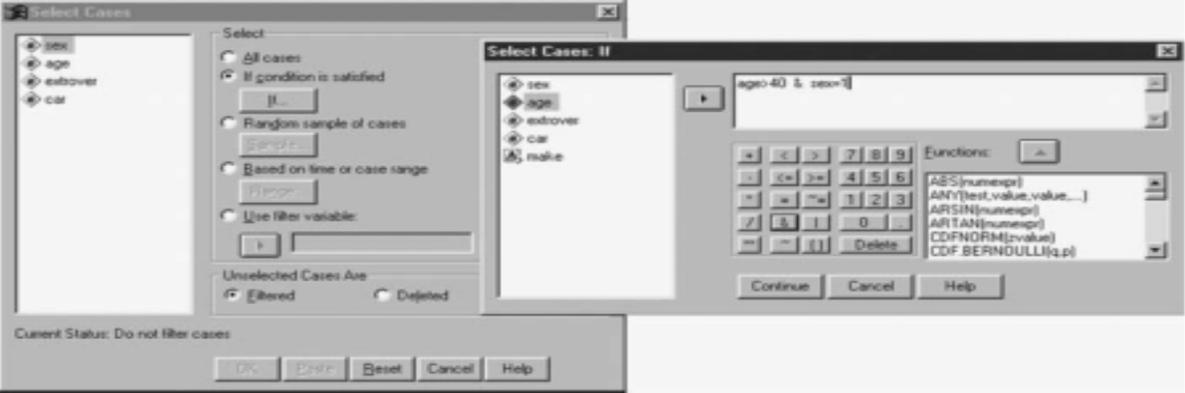
The following set of procedures allows the format of a data file to be changed:

- Sort Cases... opens a dialogue box that allows sorting of cases (rows) in the spreadsheet according to the values of one or more variables. Cases can be arranged in ascending or descending order. When several sorting variables are employed, cases will be sorted by each variable within categories of the prior variable on the Sort by list. Sorting can be useful for generating graphics.
- Transpose... opens a dialogue for swapping the rows and columns of the data spreadsheet. The Variable(s) list contains the columns to be transposed into rows and an ID variable can be supplied as the Name Variable to name the columns of the new transposed spreadsheet. The command can be useful when procedures normally used on the columns of the spreadsheet are to be applied to the rows, for example, to generate summary measures of case-wise repeated measures.
- Merge files allows either Add Cases... or Add Variables... to an existing data file. A dialogue box appears that allows opening a second data file. This file can either contain the same variables but different cases (to add cases) or different variables but the same cases (to add variables). The commands are useful at the database construction stage of a project and offer wide-ranging options for combining data files.
- Aggregate... combines groups of rows (cases) into single summary rows and creates a new aggregated data file. The grouping variables are supplied under the Break Variable(s) list of the Aggregate Data dialogue box and the variables to be aggregated under the Aggregate Variable(s) list. The Function... sub-dialogue box allows for the aggregation function of each variable to be chosen from a number of possibilities (mean, median, value of first case, number of cases, etc.). This command is useful when the data are of a hierarchical structure, for example, patients within wards within hospitals. The data file might be aggregated when the analysis of characteristics of higher level units (e.g., wards, hospitals) is of interest.
- The Split File..., Select Cases..., and Weight Cases... procedures allow using the data file in a particular format without changing its appearance in the Data View spreadsheet. These commands are frequently used in practical data analysis. Here we will describe the Select Cases... and Split

File... commands. The Weight Cases... command  (or from the toolbar) is typically used in connection with categorical data — it internally replicates rows according to the values of a Frequency Variable. It is useful when data is provided in the form of a cross-tabulation;



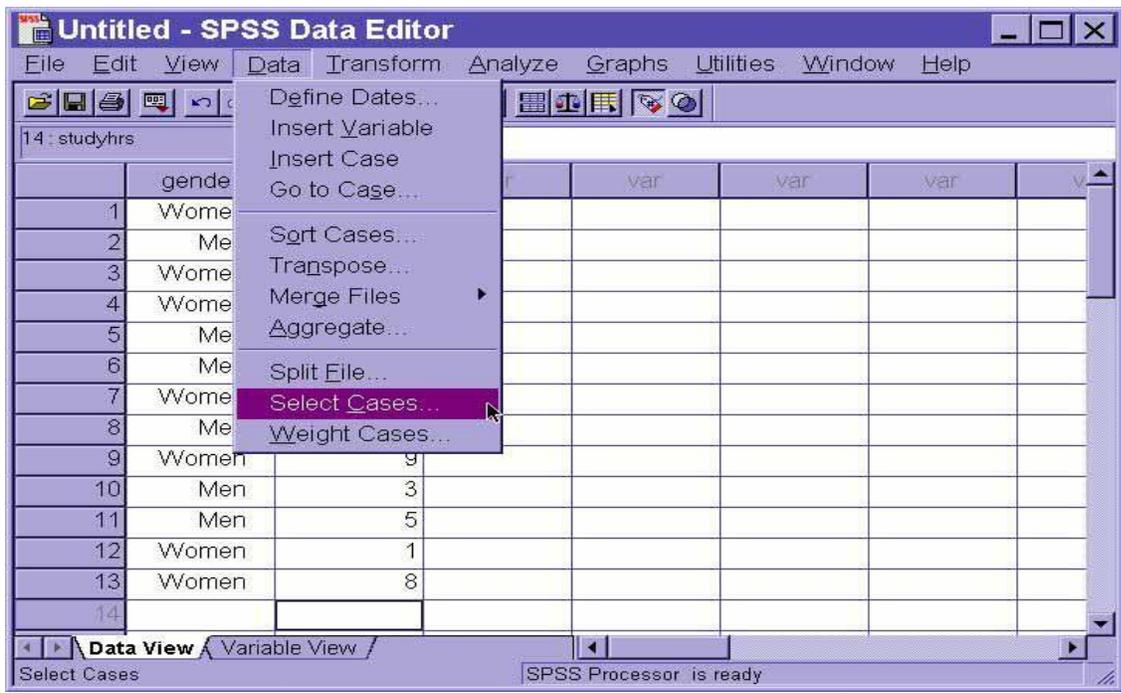
The Split File... command splits rows into several groups with the effect using this command from the drop-down menu (or selecting  that subsequent analyses will be carried out for each group separately. From the toolbar) results in the dialogue box shown in Figure above by default Analyze all cases, do not create groups is checked. A grouping of rows can be introduced by checking either Compare groups or Organize output by groups Figure above. The variable (or variables) that defines the groups is specified under the Groups Based on list .The Select Cases... command  (or selecting from the toolbar) opens the dialogue shown in



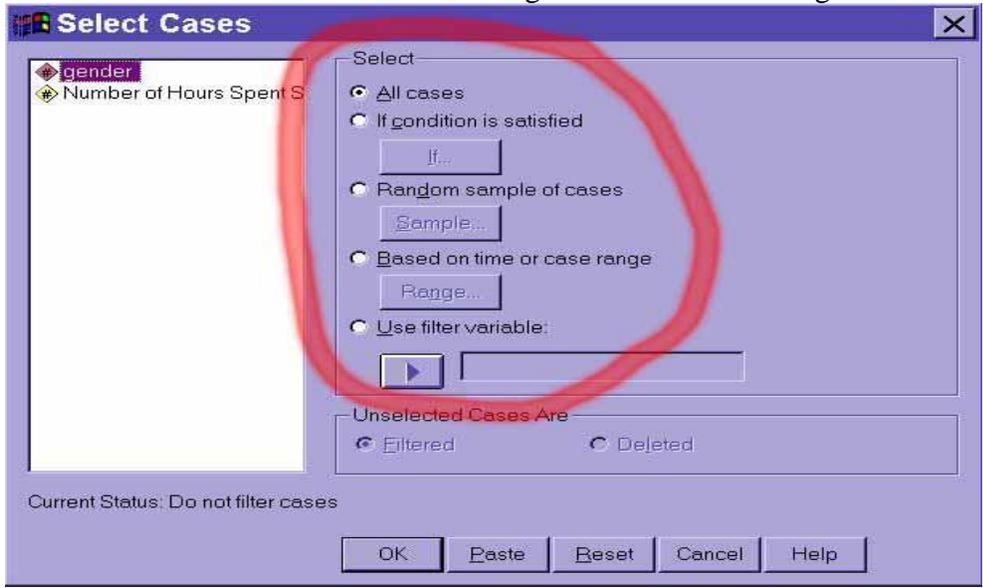
By default All Cases are selected, which means that all cases in the data file are included in subsequent analyses. Checking .If condition is satisfied allows for a subset of the cases (rows) to be selected. The condition for selection is specified using the If button. This opens the Select Cases: If sub-dialogue box where a logical expression for evaluation can be supplied. For example, we chose to select subjects older than 40 from the gender group coded “1” (males).

Select Cases

SPSS allows us to select part of the data set for further analysis, while excluding the remaining cases from these analyses. The procedure is found by choosing **Select** from the **Data Menu**.



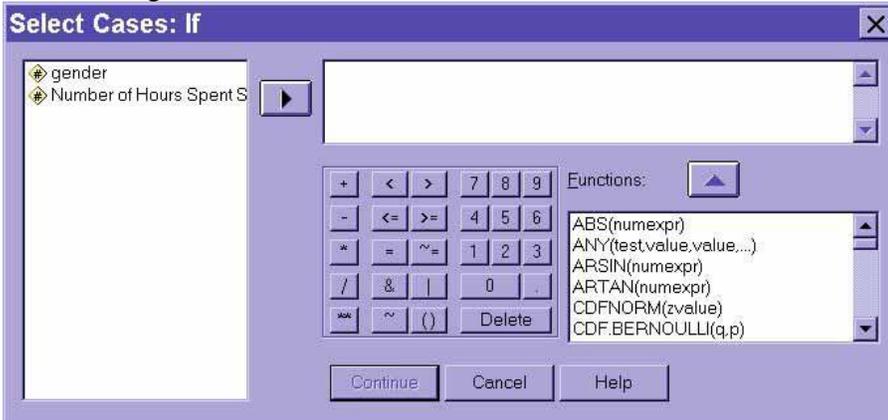
We then have several "Select" options within the dialogue box that comes up so we can tell SPSS which data to select and which to ignore. The select dialogue box looks like this:



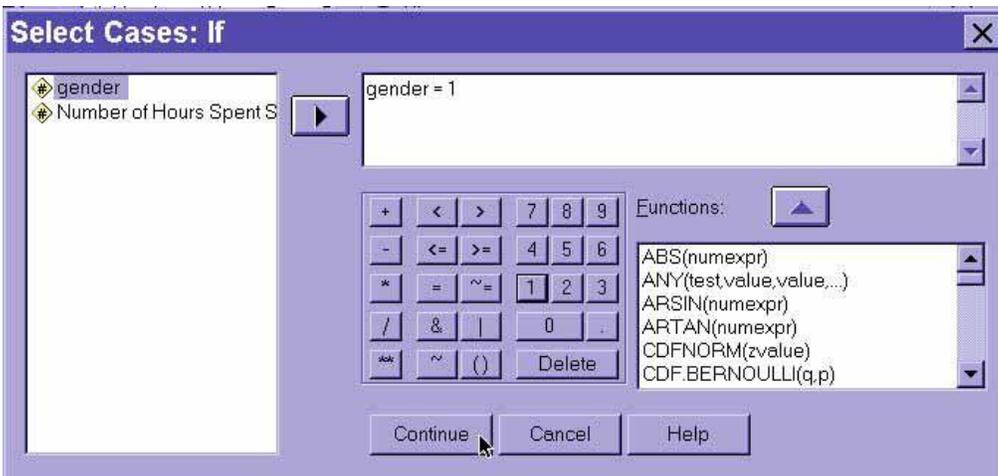
First, we have to specify how to select data and which data to retain for the analyses:

- **All cases**
This option actually turns off any previous selection and uses all data in the file. Click on this radio button and then click on the OK button.
- **If condition is satisfied**
This option allows us to specify a rule based on values of variables; all cases that meet the criteria are retained. After clicking on the radio button for this option, we click on the "If..."

button to bring up an additional dialogue box we can define the rule or rules for including or excluding data:

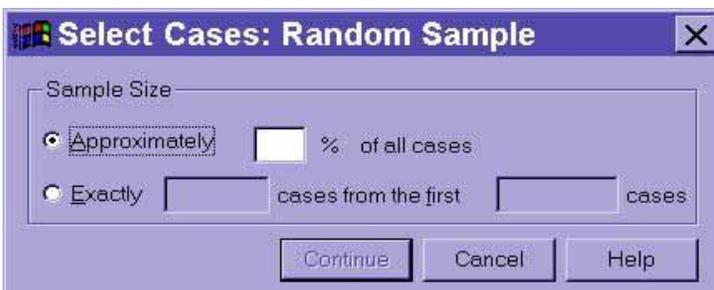


We want to specify a condition in the right box based on one or more variables listed in the box to the left. Click on the HELP button for definitions of each of the "calculator" buttons in this dialogue box. For instance, if we have a variable for gender coded as 1 for women and 2 for men, we can select only the women for analysis by selecting "gender=1" as shown below:



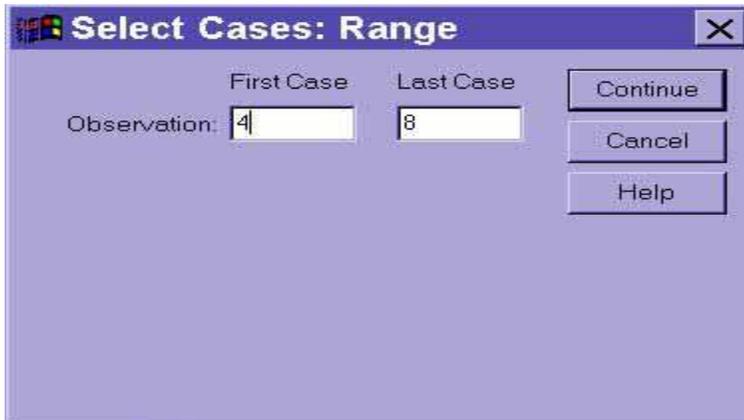
- **Random sample of cases**

This option allows us to select a specified proportion of the cases or a fixed number of cases at random. After selecting the corresponding radio button, we click on the "Sample" button to bring up a dialogue box that allows us to specify either an approximate percentage of cases or to sample an exact number of cases at random.



- **Based on time or case range**

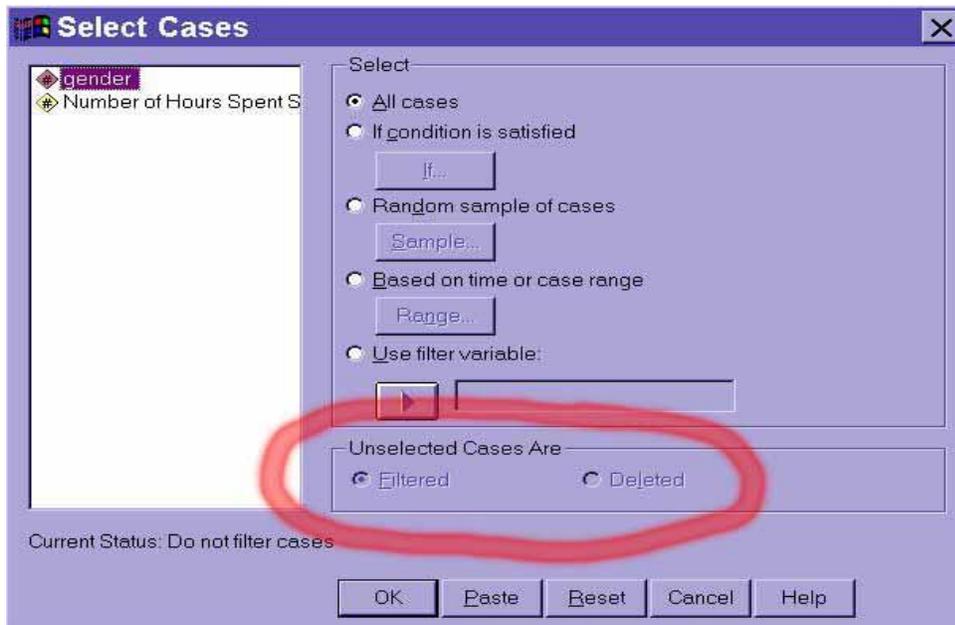
This option allows us to specify a range of cases based on the internal case number. Thus, by typing in "4" and "8", we can select only cases 4, 5, 6, 7, and 8.



- **Use filter variable**

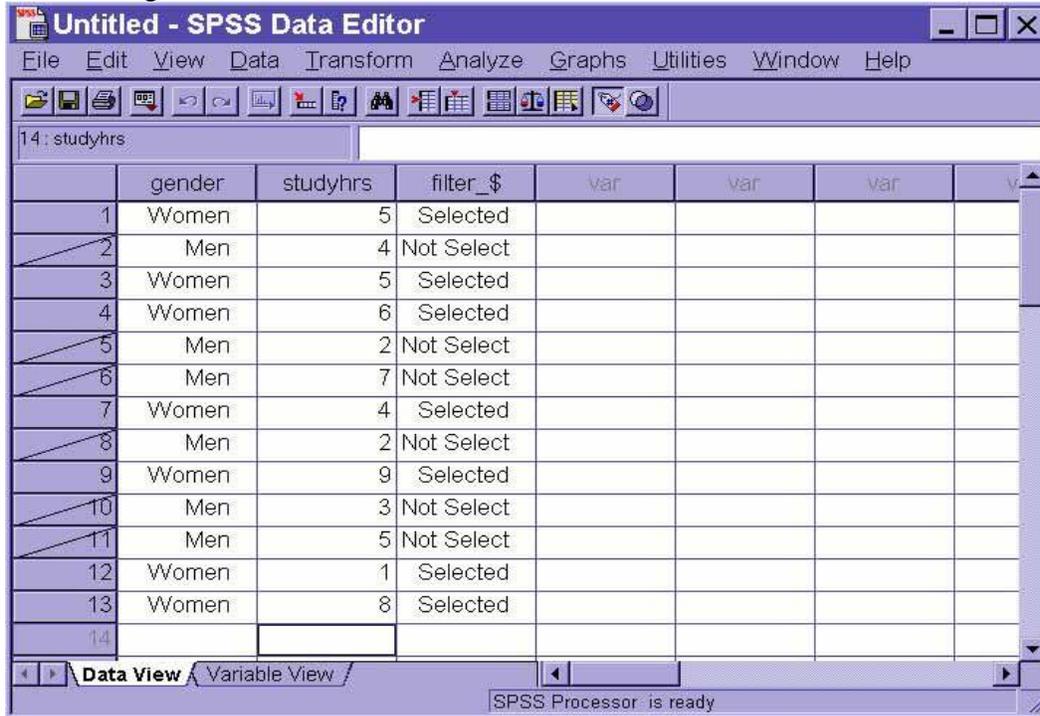
This option uses a special variable type (a filter variable that has values of 1 or 0 only: the cases with a 1 are included, and the cases with 0 on the filter variable are excluded).

We also have to tell SPSS what to do with the unselected data. SPSS can either filter it or delete it. If we choose to delete the unselected data, those cases not meeting the criteria specified above will be **deleted and cannot be recovered!** If we choose to filter the unselected data, then the data will **not** be deleted, but SPSS will ignore the data in any and all analyses until the filter is "turned off" by selecting the **All cases** option described above. This filtering option, therefore, is far safer than the deleting option.



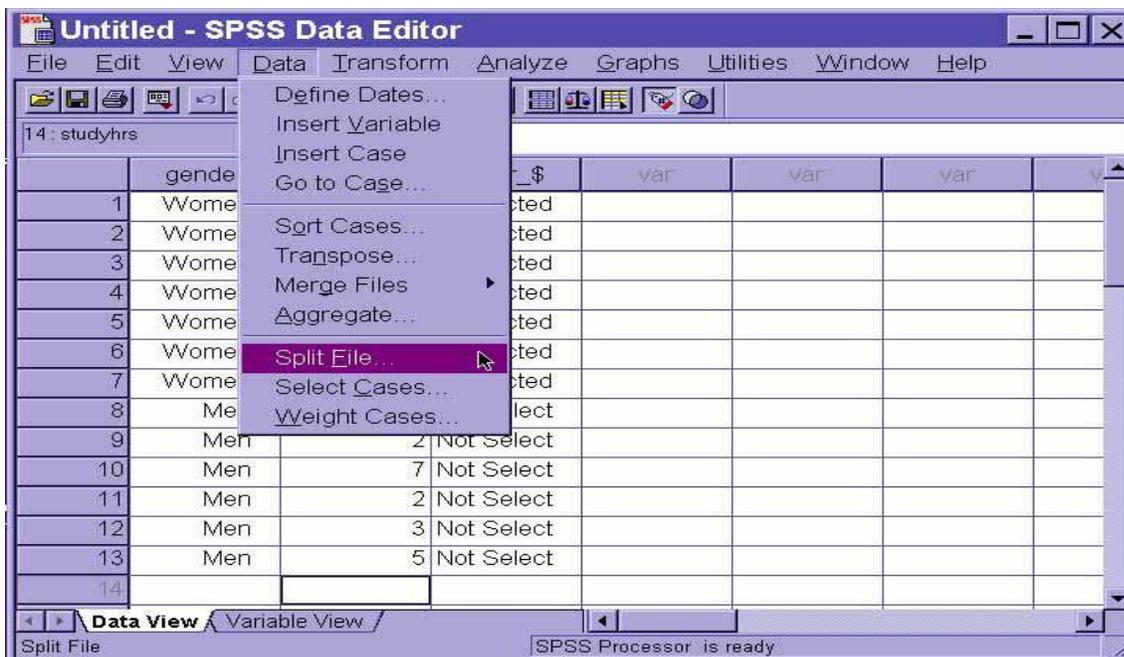
After we have selected one of the radio buttons for the selection method and after we have selected one of the radio buttons for handling unselected data, clicking on the **OK** button will perform the

selection. If we have chosen to filter unselected data, cases that are not being used will have a slash through the case number.

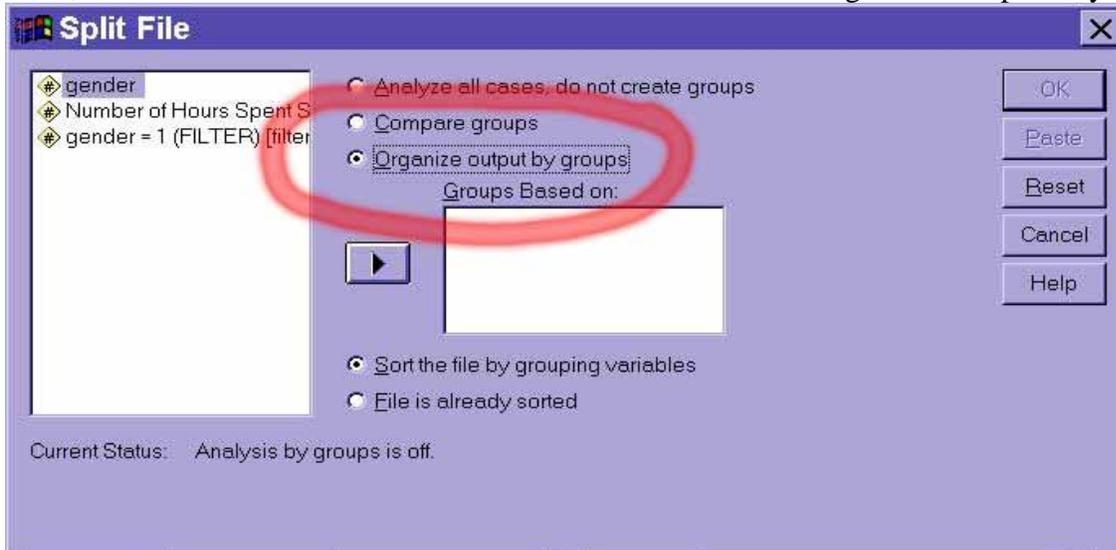


Split File

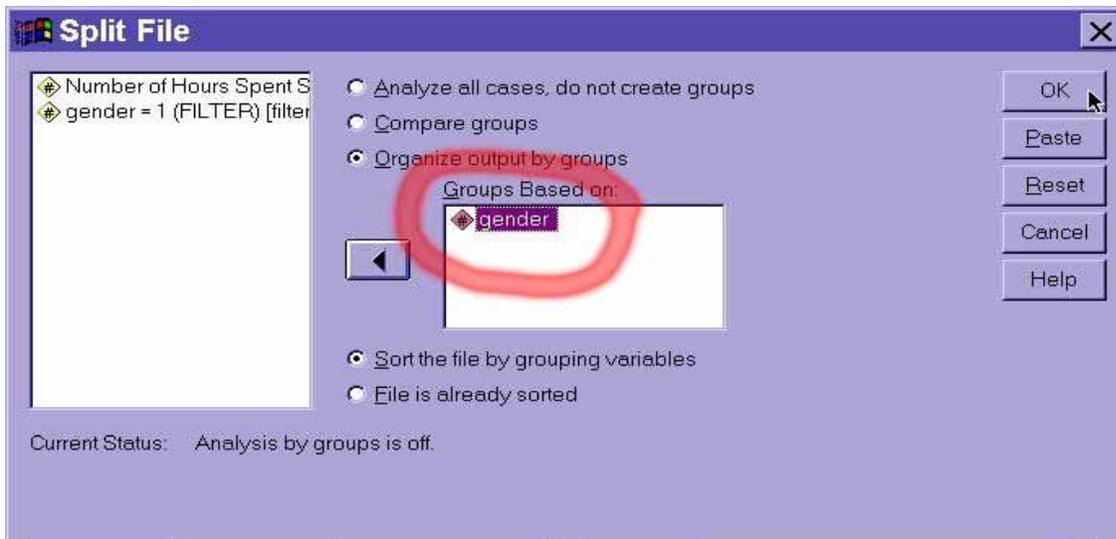
The split file option from the **Data** menu works similarly to the select option. The difference, though, is that we use the split function to repeat the same analyses, separately, on multiple groups. For instance, if we want to compute descriptive statistics of men and women separately, we would select the split option from the **Data** menu.



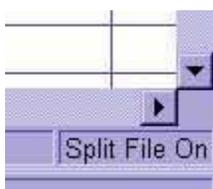
Then, we would click on the radio button for "Organize output by groups":



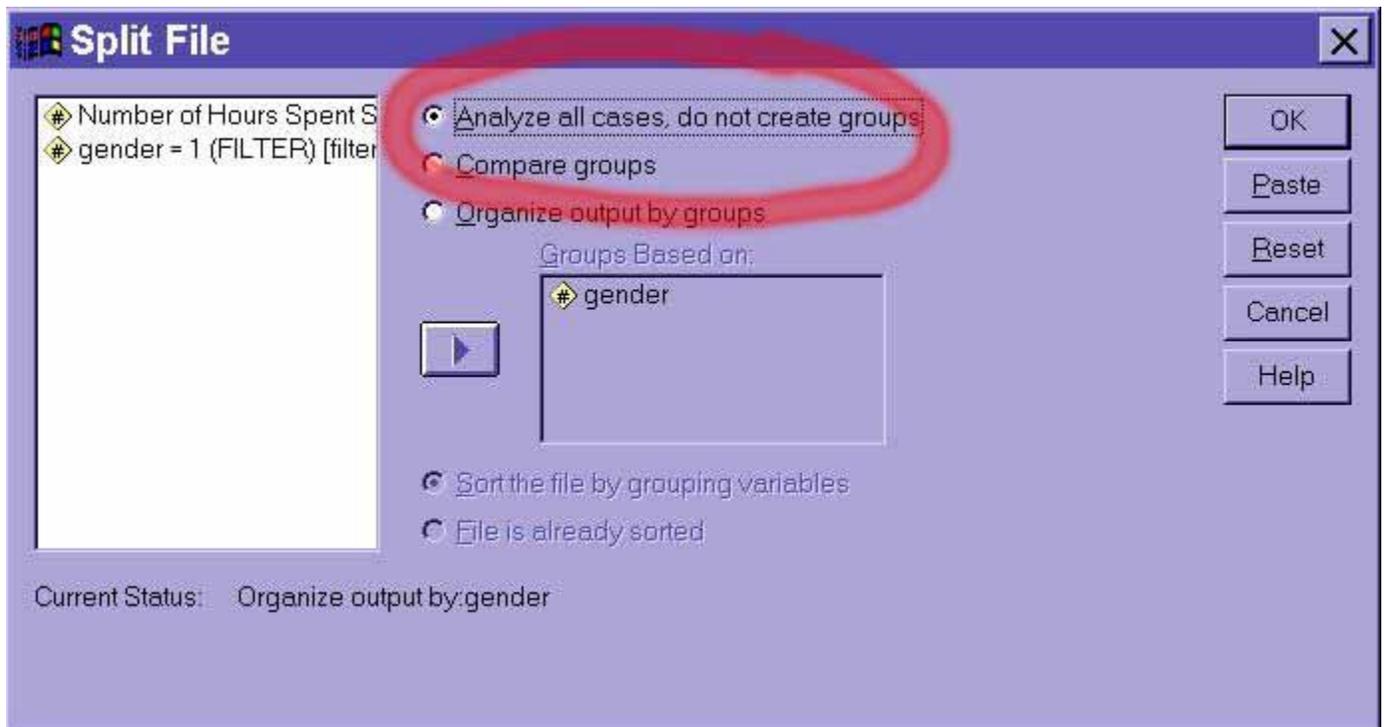
Then, we would select the gender variable in the left field and click the right arrow to move it to the right field:



After clicking on the **OK** button, we would see the "Split file on" message in the lower right-hand corner of the SPSS data window



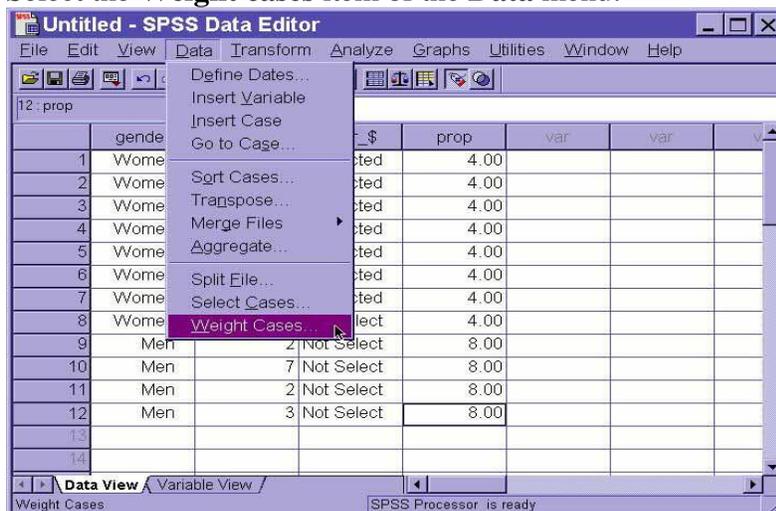
Finally, any further analyses we run will be run separately for men and for women until we turn off the split file option by selecting "Analyze all cases -- do not split groups":



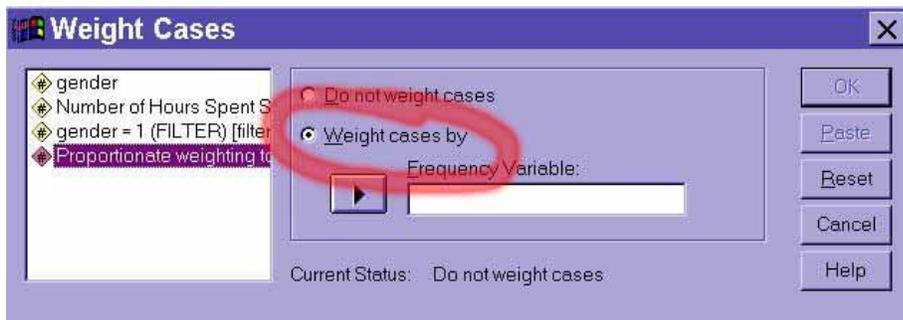
Weight Cases

Sometimes we need to let some cases have more "weight" in the analyses because we under or over sampled from a group. We can determine a set of weights, one weight for each case so that the groups would more closely resemble the proportions we had hoped to sample. Alternatively, sometimes there is a single variable in a data set that represents the number of occurrences of a behavior or a frequency. We can instruct SPSS to treat such a variable as a case weight so that we can create shorter data sets for frequency data without having to enter a separate case for each observation. You may want to refer to this section after chi-square procedures are discussed in class.

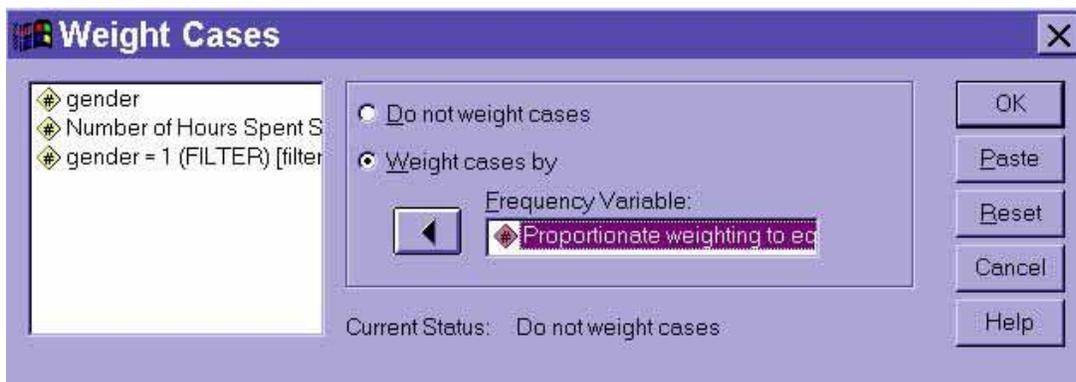
Select the **Weight cases** item of the **Data** menu:



In the dialogue box that pops up, select the option for "Weight cases by":



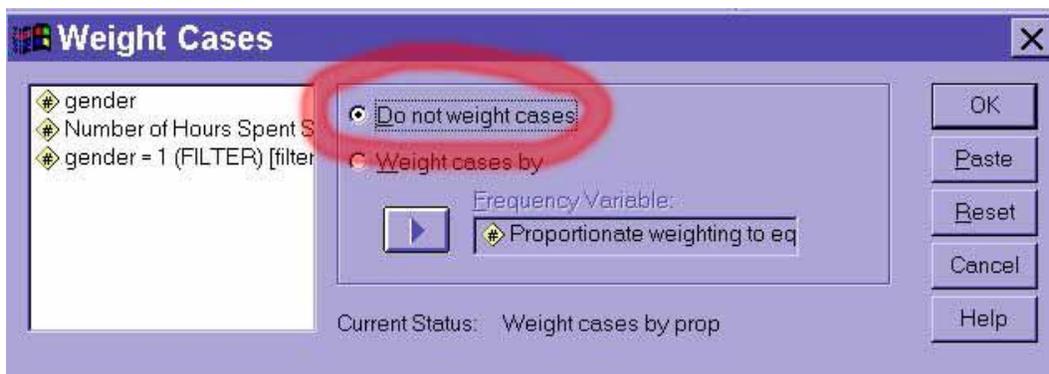
Click on the weighting variable in the left field that contains frequencies or optimal case-weight information, click on the right arrow to move that variable into the right field:



After we click on the **OK** button, the "Weight on" message appears in the lower right hand corner of the SPSS data window.



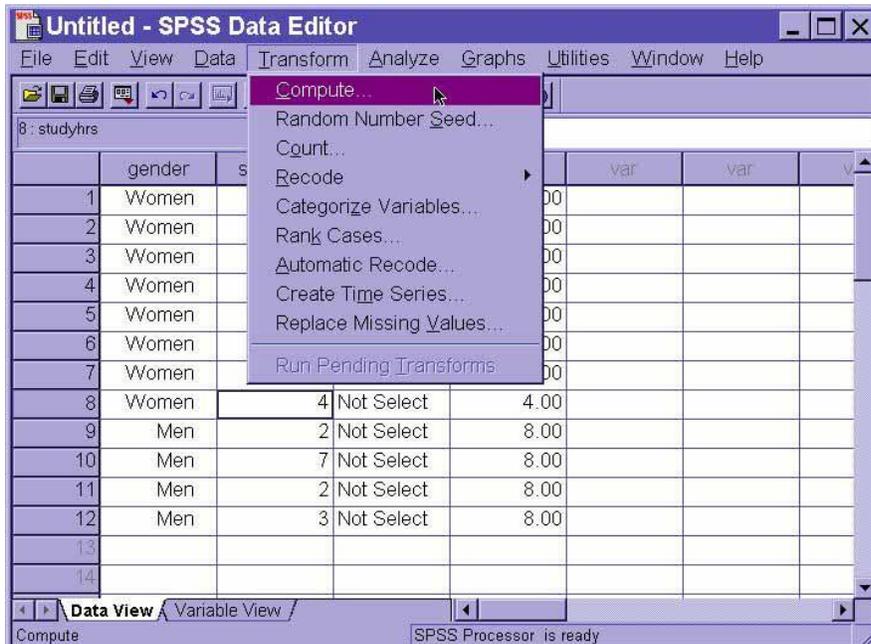
When we select the **Weight** option from the **Data** menu and select the "Do not weight cases" option and click "OK", then cases will no longer be weighted and the weight on message disappears.



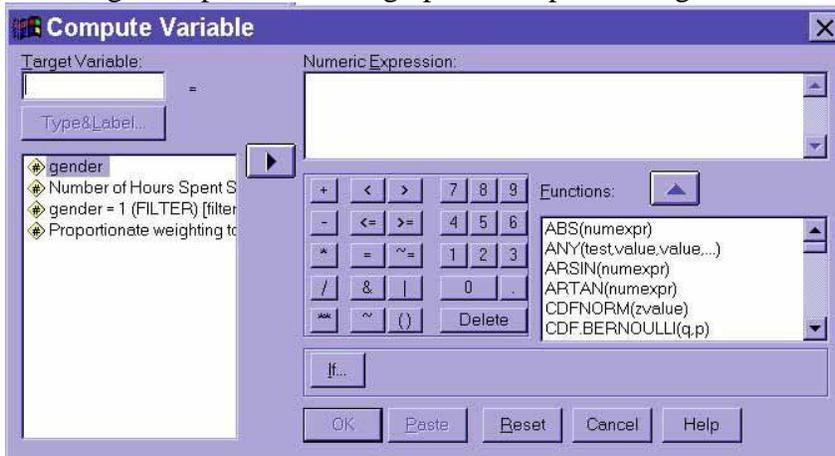
Transform/Compute New Variable

SPSS has very powerful capabilities for creating new variables as a function of existing variables. For instance, we can use these functions to create averages of existing variables, to rescale existing

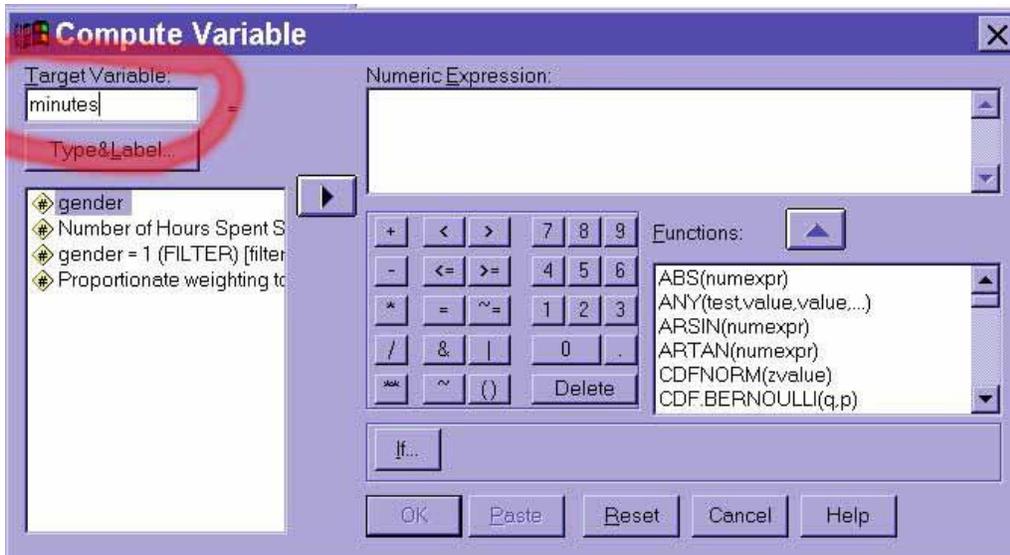
variables, or to compute difference scores by subtracting one variable from another. To do so, we select the **Compute** option from the **Transform** menu:



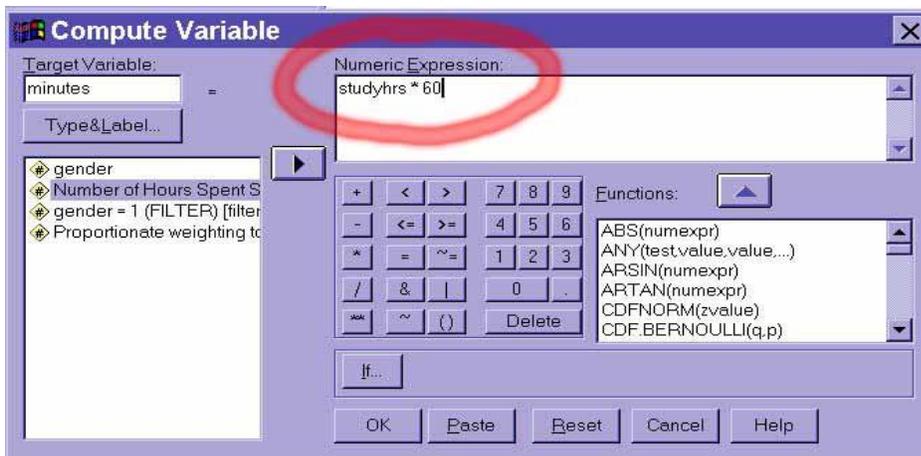
Selecting this option will bring up the compute dialogue box:



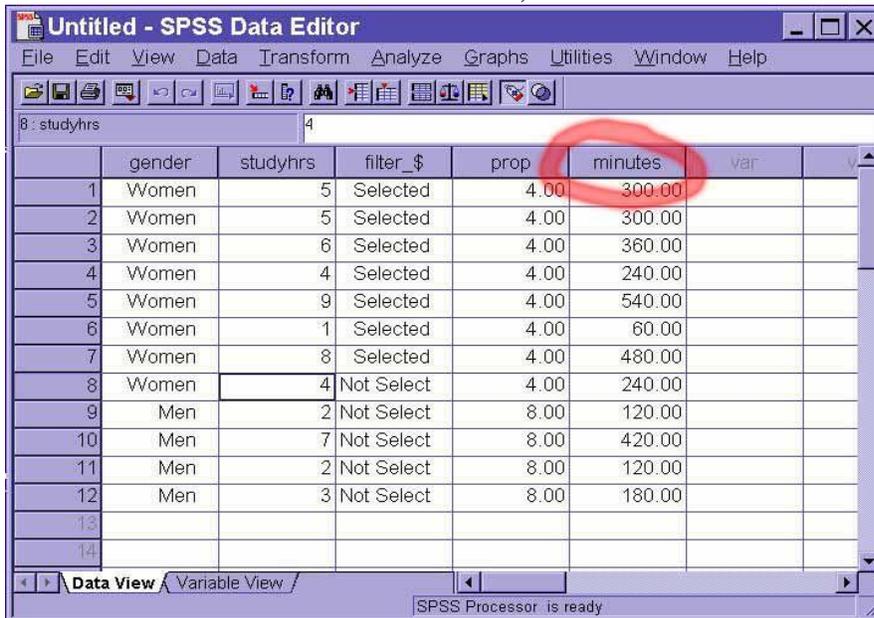
First, we need to supply a name for the target variable (i.e., the new variable SPSS will create to contain the new values. For example, we may want to create a new variable to report the number of minutes studied rather than the number of hours spent studying. Thus, we would name the new variable "minutes":



The next step is to define for SPSS how the new values should be computed, essentially giving SPSS a formula. To convert hours to minutes, we should multiply the studyhrs variable times 60. Thus, we type "studyhrs*60" in the numeric expression field:



After we have clicked on the **OK** button, the new variable "minutes" is created:



Lab Exercises

1. Create a data set for the following data:

Group	Hw1	Hw2	Hw3
expt	92	84	93
expt	77	84	85
expt	87	86	81
expt	89	90	93
expt	64	73	78
control	81	84	93
control	83	90	91
control	84	88	86
control	82	80	78
control	96	91	88

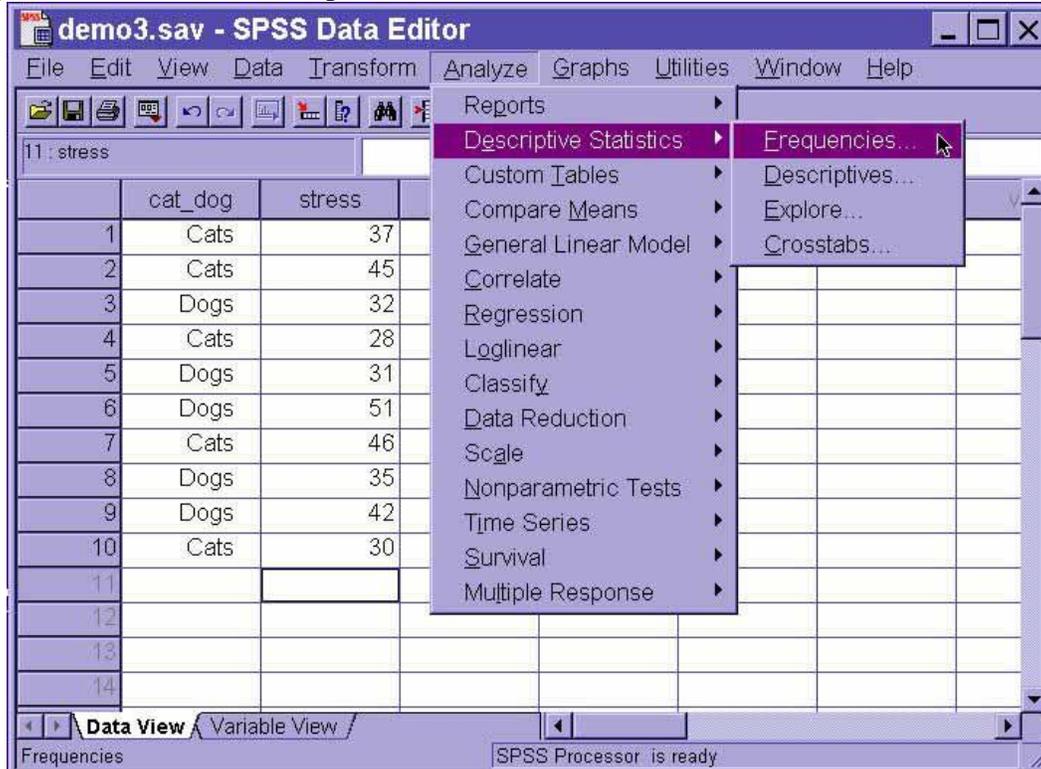
- 2-Select only the experimental group.
- 3-Select approximately 50% of the sample.
- 4-Select only cases 2 through 5.
- 5-Split the file to analyze the experimental group and the control group separately.
- 6-Compute a new variable to represent the SUM of all three homework scores.
- 7-Compute a new variable to represent the average of all three homework scores.

Transform includes several options to change current variables. For example, you can change continuous variables to categorical variables, change scores into rank scores, add a constant to variables, etc.

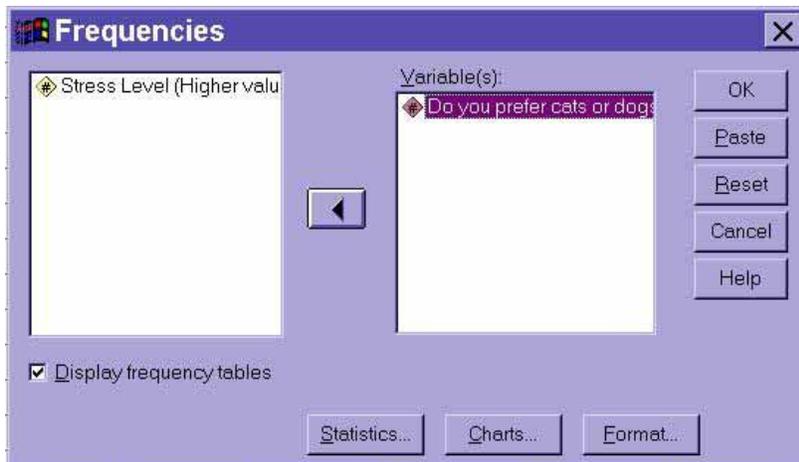
Analyze includes all of the commands to carry out statistical analyses and to calculate descriptive statistics. Much of this book will focus on using commands located in this menu.

1-Frequencies

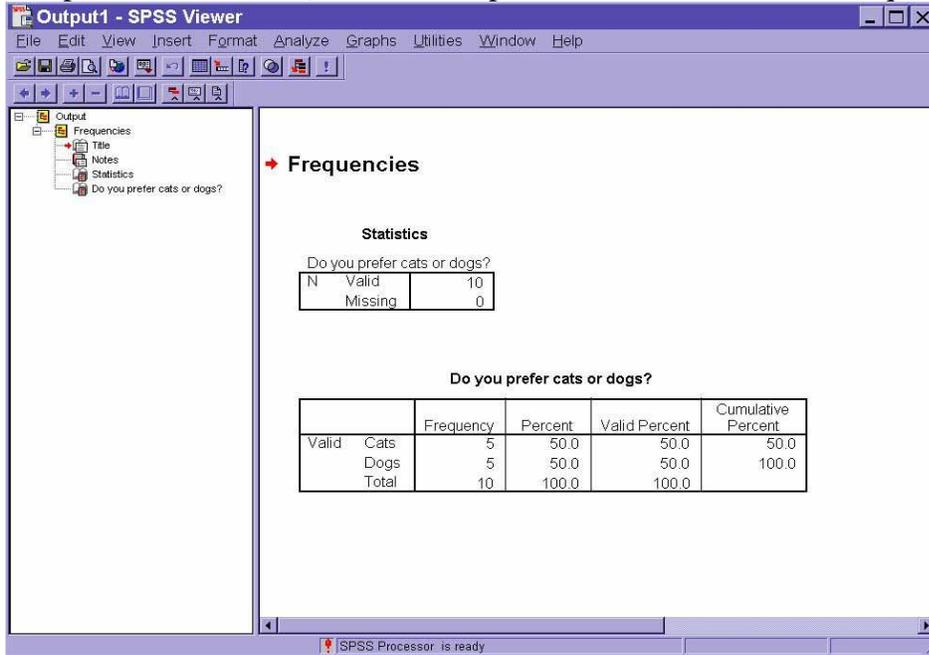
The frequencies procedure is primarily used for discrete data (e.g., nominal and ordinal data), although there are a number of options that are useful for scale level data.



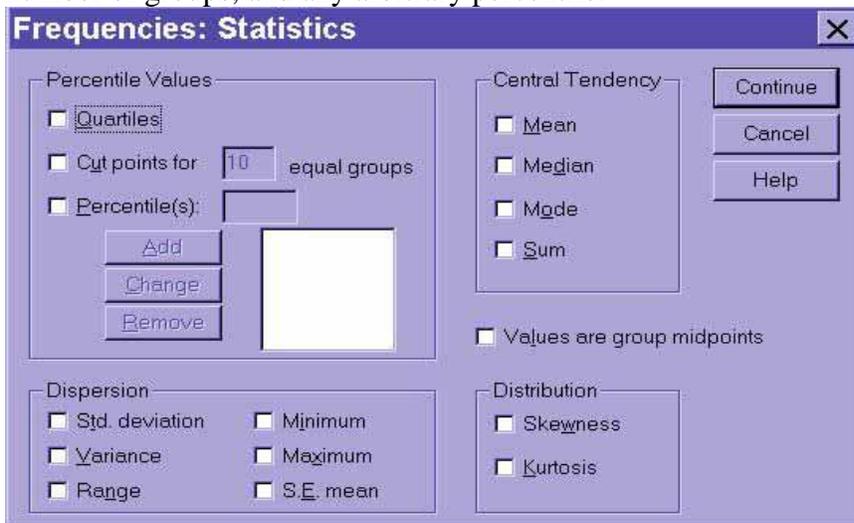
This option brings up a dialogue box, and we need to move the variables of interest from the field on the left to the field on the right.



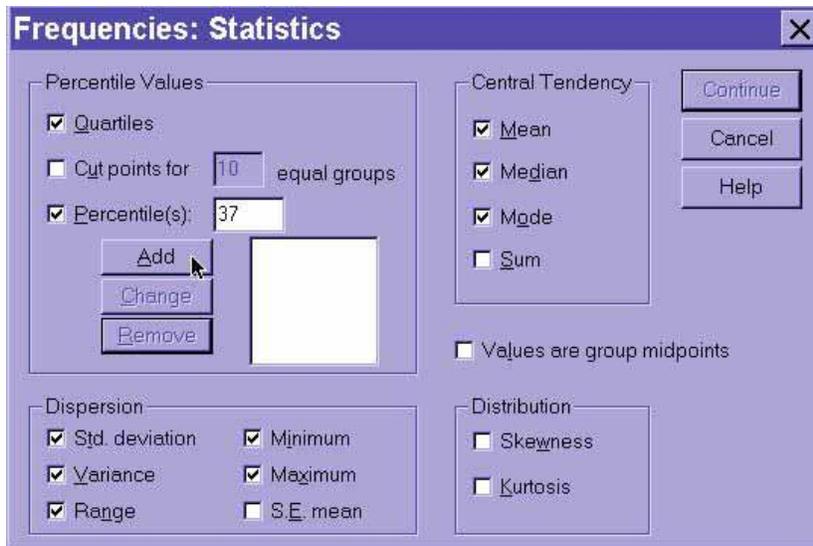
For nominal variables, for which further descriptive statistics are not appropriate (with the exception of the mode), we can skip the **Statistics** to obtain frequencies for each category.



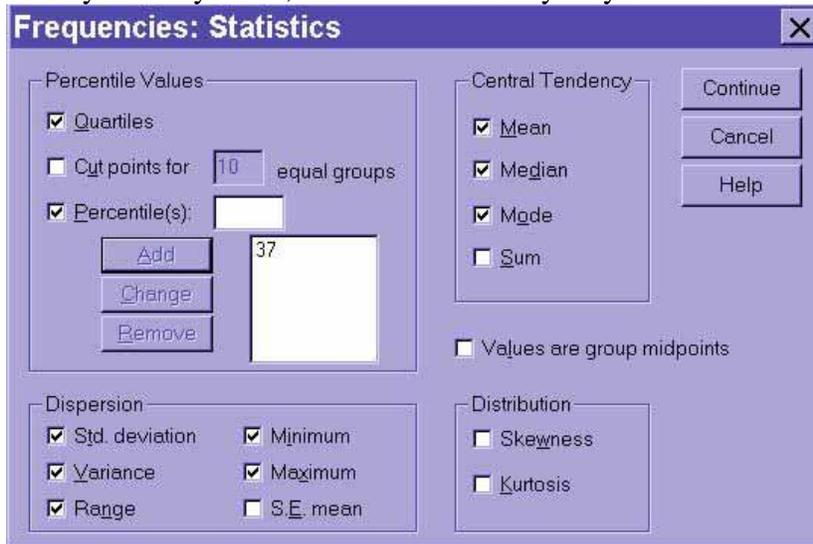
For ordinal and scale variables, though, we will want to specify additional descriptive statistics to be calculated. These can be broken down into measures of central tendency (mean, median, mode, sum), variability (variance, standard deviation, range, minimum, maximum), and percentiles. This last category includes quartiles (25th, 50th, and 75th percentiles), cut-points for an arbitrary number of groups, and any arbitrary percentile.



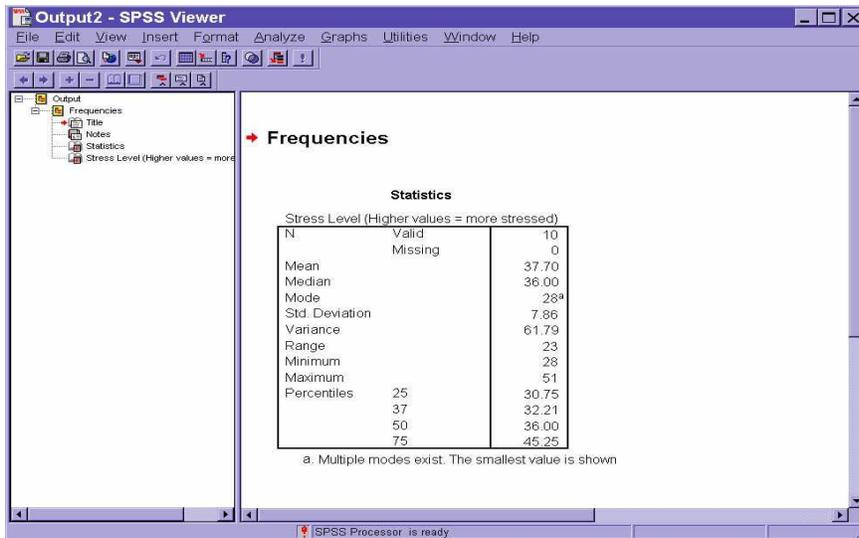
Most options are selected simply by clicking on the box next to each item. For specific, arbitrary percentiles, select the option, type the desired percentile in the field to the right, and then click on the add button below:



Which results in the desired percentile being added to the list. Note that one can always delete or modify an entry. Also, more than one entry may be made.



These options generate the following output:



Lab Exercises

Group	Gender	Hw1	Hw2	Hw3
expt	Male	92	84	93
expt	Female	77	84	85
expt	Male	87	86	81
expt	Female	89	90	93
expt	Male	64	73	78
control	Female	81	84	93
control	Male	83	90	91
control	Female	84	88	86
control	Male	82	80	78
control	Female	96	91	88

2-Using the Frequencies option, find the mean, median, mode, quartiles, 95th percentile, variance, standard deviation, minimum, and maximum of Hw1, Hw2, and Hw3.