**Introduction:**

SPSS (Statistical Package for the Social Sciences) is a versatile and responsive program designed to undertake a range of statistical procedures.  SPSS software is widely used in a range of disciplines and is available from all computer pools within the University of South Australia.

It’s important to note that SPSS is not the only statistical software – there are many others that you may come across if you pursue a career that requires you to work with data.  Some of the other more common statistical packages include Stata and SAS (and there are many others).  The focus for this session, however, is on SPSS.

**Basic Steps Of Data Analysis**

**Steps**

* Load your excel file with all the **data**. Once you have collected all the **data**, keep the excel file ready with all **data** inserted using the right tabular forms.
* Import the **data** into **SPSS**. ...
* Give specific **SPSS** commands. ...
* Retrieve the results. ...
* Analyse the graphs and charts. ...
* Postulate conclusions based on your **analysis**.

**Starting Spss**

To run **SPSS**, log in and click **Start** - Programs - IBM**SPSS** Statistics - IBM **SPSS** Statistics 23. When**SPSS** is first started you are presented with a dialog box asking you to open a file. Pick a recently opened file or pick "Open another file" from the list on the left.

**Basics**

* The SPSS Windows and Files
  + Data Editor (.sav files)
  + Output Viewer (.spv files)
  + Syntax Editor (.sps files)
* Issuing Commands
  + Dialog Boxes
* Working with the Data Editor
  + Data View
  + Variable View
* Working with the Output Viewer
* Working with the Syntax Editor

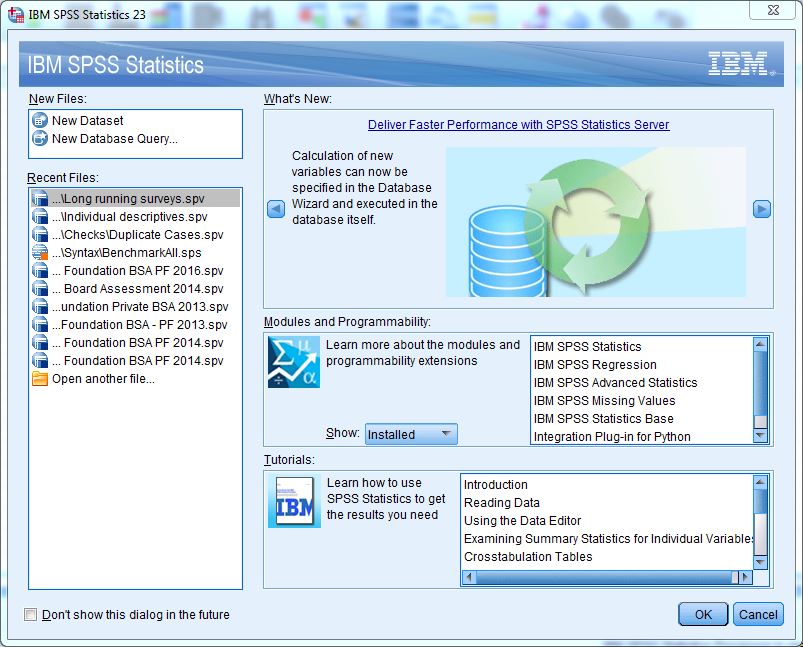
**Opening A Data File**

## Starting SPSS Statistics

The SSCC has SPSS installed in our computer labs (4218 and 3218 Sewell Social Sciences Building) and on some of the Winstats. If you work on a University-owned computer you can also go to DoIT's Campus Software Library, and download and install SPSS on that computer (this requires a NetID, and administrator priviledges).

To run SPSS, log in and click **Start - Programs - IBM SPSS Statistics - IBM SPSS Statistics 23**.

When SPSS is first started you are presented with a dialog box asking you to open a file. Pick a recently opened file or pick "Open another file" from the list on the left.



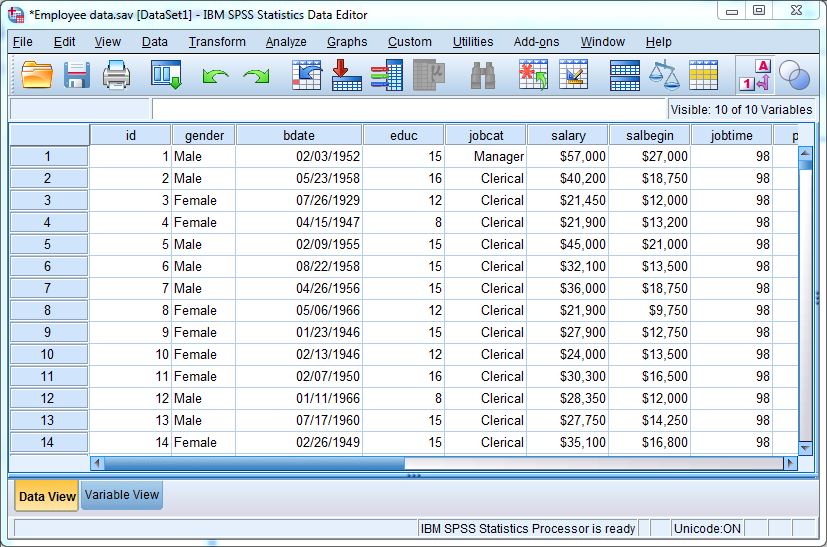
Typically you start your SPSS session by opening the data file that you need to work with.

## The SPSS Windows and Files

SPSS Statistics has three main windows, plus a menu bar at the top. These allow you to (1) see your data, (2) see your statistical output, and (3) see any programming commands you have written. Each window corresponds to a separate type of SPSS file.

### Data Editor (.sav files)

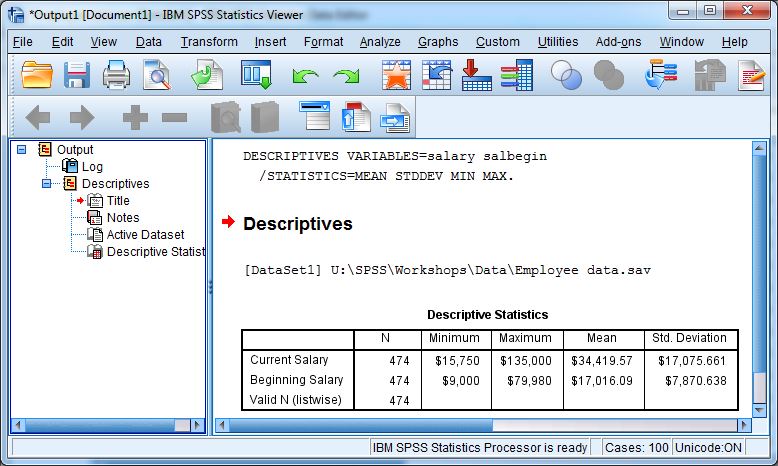
The Data Editor lets you see and manipulate your data. You will always have at least one Data Editor open (even if you have not yet opened a data set). When you open an SPSS data file, what you see is a working copy of your data. Changes you make to your data are not permanent until you save them (click **File - Save** or **Save As**). Data files are saved with a file type of **.sav**, a file type that most other software cannot work with. When you close your last Data Editor you are shutting down SPSS and you will be prompted to save all unsaved files.



To open a different data set, click **File - Open - Data**. (It is also possible to open some non-SPSS data files by this method, such as Excel, Stata, or SAS files.) SPSS lets you have many data sets open simultaneously, and the data set that you are currently working with, the active data set, is always marked with a tiny red "plus" sign on the title bar. In order to avoid confusion it is usually a good strategy to close out any Data Editors you're done using.

### Output Viewer (.spv files)

As you ask SPSS to carry out various computations and other tasks, the results can show up in a variety of places. New data values will show up in the Data Editor. Statistical results will show up in the Output Viewer.

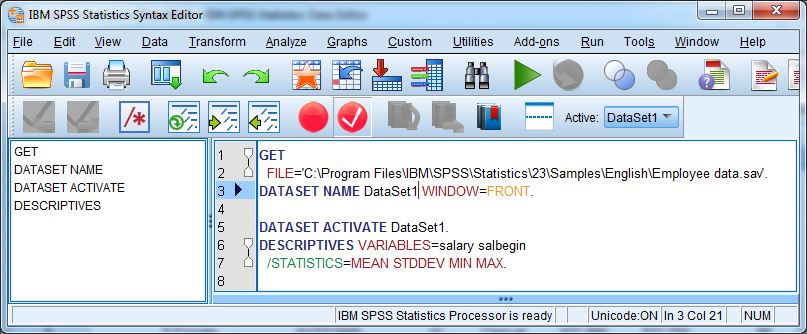


The Output Viewer shows you tables of statistical output and any graphs you create. By default it also show you the programming language for the commands that you issued (called syntax in SPSS jargon), and most error messages will also appear here. The Output Viewer also allows you to edit and print your results. The tables of the Output Viewer are saved (click **File - Save** or **Save As**) with a file type of **.spv**, which can only be opened with SPSS software.

As with Data Editors, it is possible to open more than one Output Viewer to look at more than one output file. The active Viewer, marked with a tiny blue plus sign, will receive the results of any commands that you issue. If you close all the Output Viewers and then issue a new command, a fresh Output Viewer is started.

### Syntax Editor (.sps files)

If you are working with the SPSS programming language directly, you will also open a Syntax Editor.



The Syntax Editor allows you to write, edit, and run commands in the SPSS programming language. If you are also using the menus and dialog boxes, the **Paste** button automatically writes the syntax for the command you have specified into the active Syntax Editor. These files are saved as plain text and almost any text editor can open them, but with a file extension of **.sps**.

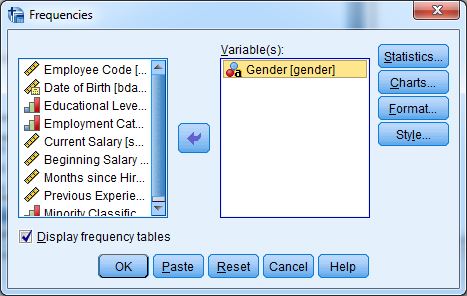
As with the other types of windows, you can have more than one Syntax Editor open and the active window is marked with a tiny orange plus sign. When you paste syntax from dialog boxes, it goes to the active Syntax Editor. If you close out all your Syntax Editors and then paste a command, a fresh Syntax Editor is opened.

## Issuing Commands

Unless you command SPSS to do something, it just sits there looking at you. In general commands may be issued either through menus and dialog boxes that invoke the programming language behind the scenes, or by typing the programming language in a Syntax Editor and running the commands.

### Dialog Boxes

Although each dialog box is unique, they have many common features. A fairly typical example is the dialog box for producing frequency tables (tables with counts and percents). To bring up this dialog box from the menus, click on **Analyze - Descriptive Statistics - Frequencies**.



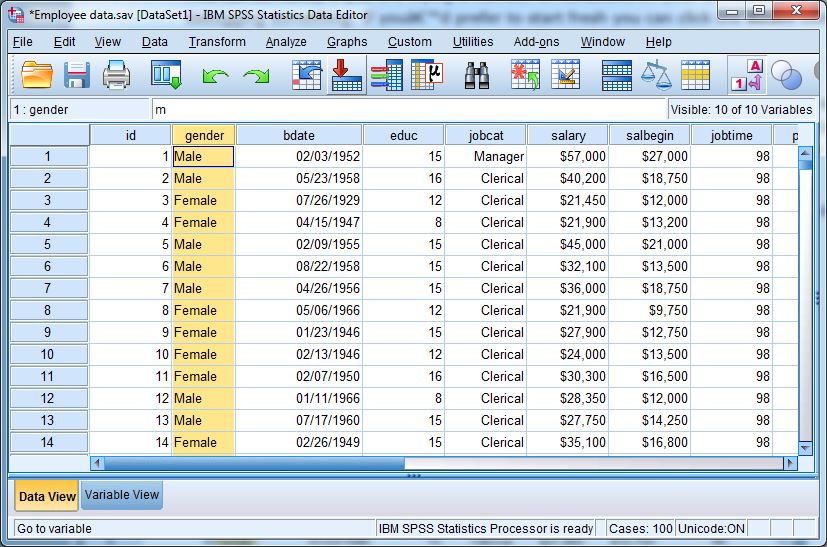
On the left is a variable selection list with all of the variables in your data set. If your variables have variable labels, what you see is the beginning of the variable label. To see the full label as well as the variable name [in square brackets], hold your cursor over the label beginning. Select the variables you want to analyze by clicking on them (you may have to scroll through the list). Then click the arrow button to the right of the selection list, and the variables are moved to the analysis list on the right. If you change your mind about a variable, you can select it in the list on the right and then click the arrow button to move it back out of the analysis list. On the far right of the dialog are several buttons that lead to further dialog boxes with options for the frequencies command. At the bottom of the dialog box, click **OK** to issue your command to SPSS, or **Paste** to have the command written to a Syntax Editor.

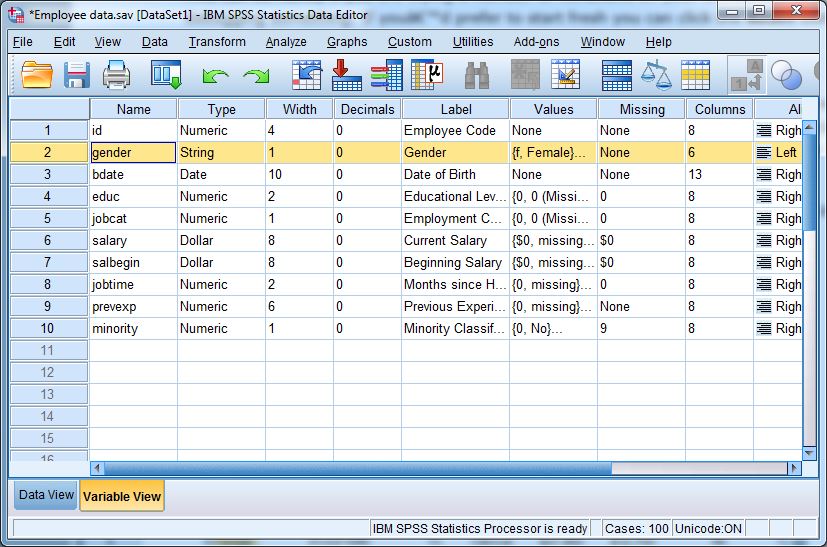
If you return to a dialog box you will find it opens with all the specifications you last used. This can be handy if you are trying a number of variations on your analysis, or if you are debugging something. If you'd prefer to start fresh you can click the **Reset** button.

## Working with the Data Editor

The main use of the Data Editor is to show you (a portion of) the data values you are working with. It can also be used to redefine the characteristics of variables (change the type, add labels, define missing values, etc.), create new variables, and enter data by hand.

The Data Editor gives you two views of your data set: a Data View and a Variable View, selected by clicking on the appropriate tab in the lower left corner of the window.





### Data View

In the Data View, the data are laid out in the standard rectangular format for statistical software. Each row represents a unit of observation, sometimes also referred to as a record or in SPSS as a case. The case (observation) number in the leftmost column is assigned automatically and is not saved as data. Each column represents a variable. All of the data in a column must be of the same type, either numeric or string (also called character).

Each data cell holds a data value. If data are missing, they are displayed as a period (".") or as a blank (" "). Data values may be displayed as either the actual value or as a formatted value. For example, a data value about a person's income might be 15000, while its formatted value might be shown as $15,000. Formats can also take the form of value labels, for instance, data recorded as 1's and 2's might be labeled as "Male" and "Female." While formatting makes it easier to interpret results, it is important to remember that the data values are what SPSS actually processes. In particular, when you set up a command that requires you to specify one or more data values, you use values and not formatted values.

You can switch the Data View between formatted and unformatted data by clicking on the **Value Labels** button on the Toolbar, the fourth button from the right when in the Data View. With value labels on you can also see the actual values for a given variable by clicking on a cell and then looking at the bar just above the data. The box to the left indicates the observation number and variable selected, e.g. 1:gender, while the center box shows you the actual value, e.g. m.

Data values can be edited or added by typing them directly into the Data View. To enter data, type in the actual data value. However, aside from very small data sets for class exercises, you should almost never need to do this.

### Variable View

In the Variable View you can see and edit the information that defines each variable (sometimes called meta-data) in your data set: each column of the Data View is described by a row of the Variable View.

The first attribute of each variable is its **Name**. The variable name is how the data column is identified in the programming language, and in order for the programming language to work gracefully variable names have to abide by certain restrictions: names must begin with a letter, and may be made up of characters, numerals, non-punctuation characters, and the period. Capitalization is ignored. Variable names may be up to 64 characters long. Other restrictions may apply - no coupons please. Variable names may be added or changed simply by typing them in.

The basic variable types are either numeric or string. However, just to make things confusing, SPSS allows you to select among several different standard formats for displaying numeric data (e.g. scientific notation, comma formatting, currencies) and calls this **Type**. You set the variable type by clicking in the column, then clicking on the gray button that appears and working in a dialog box.

The **Label** attribute allows you to give each variable a longer description that is displayed in place of the variable name, analogous to value labels for data values. Both variable labels and value labels are useful for giving you more intelligible output.

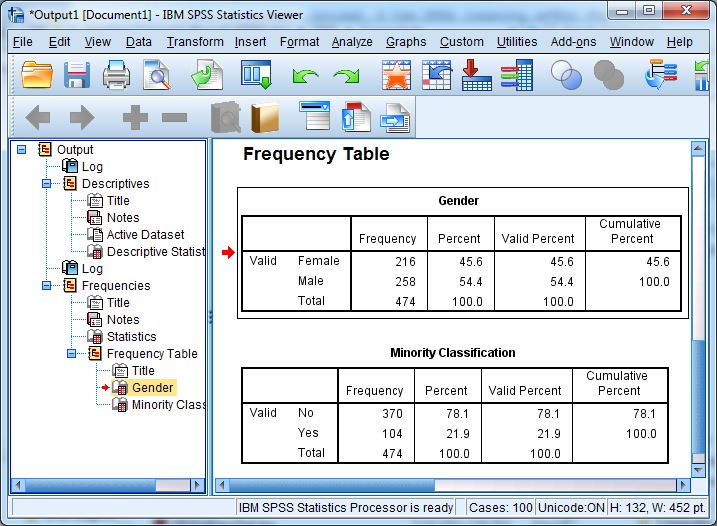
The **Values** attribute allows you to create a list of value labels. Often several variables will share a common set of value labels, and in this window you can copy and paste value label sets. Variable labels are set by simply typing them in, value labels work through a dialog box.

The **Missing** attribute is a place for you to designate certain data values that you want SPSS to ignore when it calculates statistics. For instance, in survey data it is common practice to record a data value of 8 when a respondent says "I don't know" in response to a question, and you can have SPSS treat the 8's in a variable as if they were missing data.

The other attributes, **Width**, **Decimals**, **Columns**, **Align**, **Measure**, and **Role**, are minor settings related to data display. Although Measure (level of measurement) is statistically a very important concept, it has little meaning within the SPSS software.

## Working with the Output Viewer

The Output Viewer collects your statistical tables and graphs, and gives you the opportunity to edit them before you save or print them. The Output Viewer is divided into two main sections, an outline pane on the left, and a tables pane on the right. When you print your output, it is the tables pane that is printed.



When SPSS creates output (tables, syntax, error messages, etc.) it adds them to the tables pane as objects, and each object is noted in the outline pane. Individual objects may be opened and edited, deleted, hidden, rearranged, or printed. To select an object to work with, you can either click on it in the tables pane, or click on the corresponding entry in the outline pane. A red arrow appears next to the object in both panes.

**To edit objects**, double-click on them in the tables pane. Depending on whether you are trying to edit a simple object like a title (which is just a box with some text in it), or something more complicated like a table or a graph, you may be able to simply change the object in the Output Viewer, or another window may open. Except for editing the look of graphs, it will often be easier to edit your output by exporting it to Microsoft Word first, but in principle you can change anything you can see in your output, down to deleting columns and changing numbers. (But if your intent is to fake your results, you should attend our Simulations workshop for better methods of doing this.)

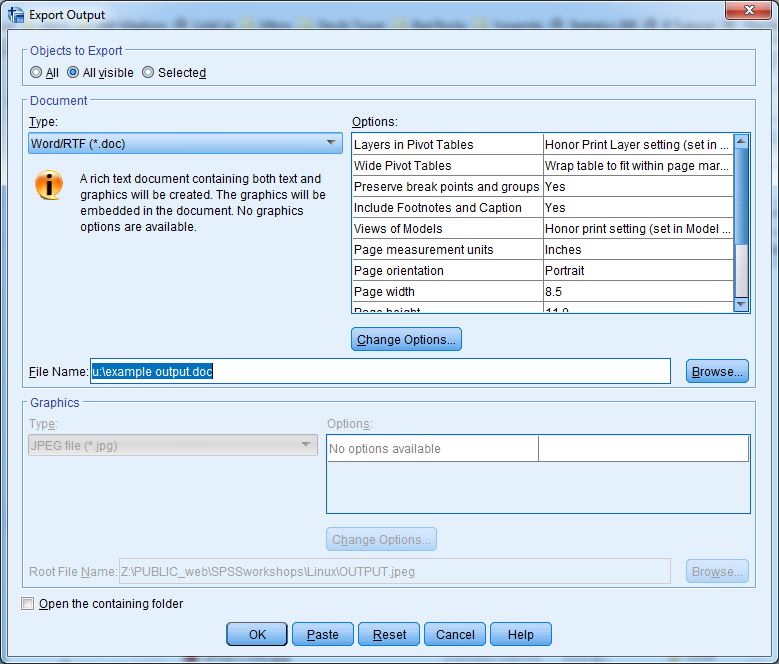
**To delete objects**, select them in either pane and use the **Delete** key.

**To hide objects**, double-click on the icon for each object in the outline pane. To make them visible, just double-click again. You can hide a whole section of the outline by clicking on the minus sign to the left of the group in the outline pane. Hidden objects are not printed, but are saved with the output file.

**To rearrange objects**, select the object (or group of objects) in either pane, and drag them until the red arrow points to the object below which you want them to appear.

**To export your output**, you go through a special procedure. In the Output Viewer click **File - Export** to invoke the Export dialog box. There are three main settings to look at. First, pick the type of file to which you want to export: useful file types include Excel, PDF, PowerPoint, or Word. Next, check that you are exporting as much of your output as you want, the **Objects to Export** at the top of the dialog. If you have a part of your output selected, this option will default to exporting just your selection, otherwise you typically will export all your visible output. Finally, change the default file name to something meaningful, and save your file to a location where you will be able to keep it, like your U:\ drive.

Once your options are set, click **OK**.

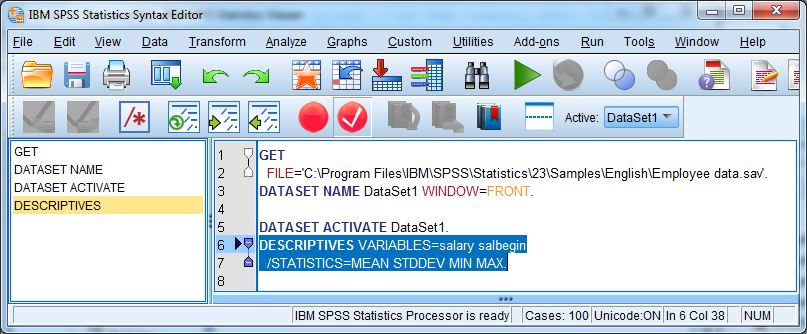


## Working with the Syntax Editor

Learning SPSS programming syntax is a separate topic; the fundamentals are addressed in our SSCC training workshops. But you don't have to memorize a whole new language in order to paste and run SPSS syntax.

The fundamental unit of work in the SPSS language is the command: think of commands as analogous to well-formed sentences. In this language, commands begin with a keyword and end with a period. Commands should begin in the leftmost column in the editor. If they are wrapped onto more than one line, the continuing lines should begin with a blank space. Capitalization does not matter. The Syntax Editor displays syntax that SPSS cannot interpret in red type.

Like the Output Editor, the Syntax Editor has two panes. The tables pane on the right is what is actually saved in the **.sps** file.



**Running syntax**. To have SPSS actually carry out your command(s), you must run them. Click **Run**, and then one of the menu options. There is also an icon on the Toolbar to run your program, a right-facing triangle ("play"). You can run all the commands in the editor, or select a group of commands and run just that (be careful that you highlight full commands, from the first keyword through the final period). You can also run the currentcommand, which is whatever command the cursor is located within.

**Pasting and running**. From most dialog boxes you have the option of pasting commands instead of simply running them. SPSS then writes the command into a Syntax Editor. The syntax tends to be verbose, specifying many options that are the defaults - syntax you write yourself tends to be much shorter and simpler. After you have pasted a command, you still need to run it to get any output.