

## Appendix

### Importing Data from Other Applications

Data entered, or captured by another application can be analysed with SPSS. There are several approaches:

#### ***Do not Use Copy / Paste from Excel***

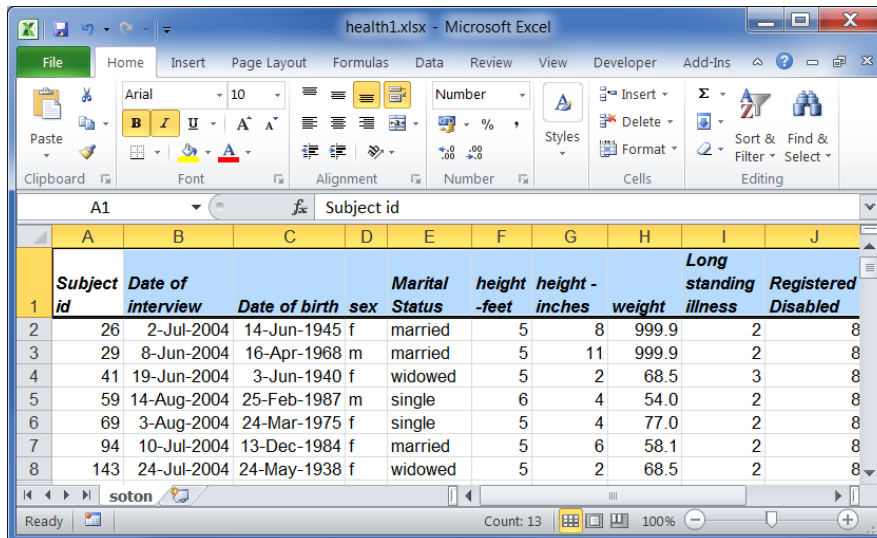
This is not the best approach, for example Dates are pasted in as String variables and most numbers will have two decimal places. Use the method below.

#### ***Using File | Open (Excel files)***

SPSS can read spreadsheet files using **File | Open | Data** menu option. This is more efficient than Copy and Paste described above.

In the **Open File** panel select **Files of type: Excel (\*.xls, \*.xlsx, \*.xlsm)**.

The data should be arranged as shown for an Excel file below in a rectangular grid format; the first row should contain a name for each column and data in the rows 2 onwards.



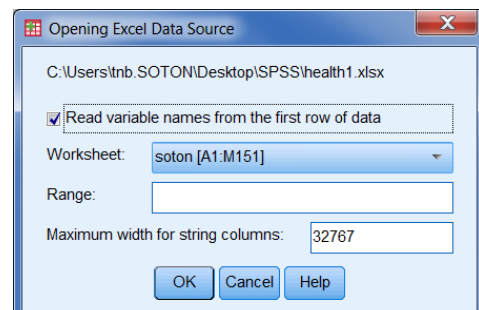
	A	B	C	D	E	F	G	H	I	J
	Subject id	Date of interview	Date of birth	sex	Marital Status	height -feet	height - inches	weight	Long standing illness	Registered Disabled
1										
2	26	2-Jul-2004	14-Jun-1945	f	married	5	8	999.9	2	8
3	29	8-Jun-2004	16-Apr-1968	m	married	5	11	999.9	2	8
4	41	19-Jun-2004	3-Jun-1940	f	widowed	5	2	68.5	3	8
5	59	14-Aug-2004	25-Feb-1987	m	single	6	4	54.0	2	8
6	69	3-Aug-2004	24-Mar-1975	f	single	5	4	77.0	2	8
7	94	10-Jul-2004	13-Dec-1984	f	married	5	6	58.1	2	8
8	143	24-Jul-2004	24-May-1938	f	widowed	5	2	68.5	2	8

*Note: Some versions of Excel may cause a 'file already open' type error message at this point. The solution is to close the file in Excel.*

Check the **Read variable names from the first row of data** box if names are used.

Select the appropriate **Worksheet** in the Excel file.

If the data you want to read does not start in the first row of the spreadsheet, enter the **Range** of cells in the **Opening Excel Data Source** dialogue box.



## Rules for reading spreadsheet files

- Rows are cases and columns are variables.
- The number of cases is fixed by the last row with any nonblank cells.
- The last column with any nonblank cells fixes the number of variables. If the first row supplies variable names, columns are excluded if the cell in the first row is blank.
- Data types are determined by the column width and data type of the first cell. Data that cannot be converted are set to system-missing. If the first cell in a column is blank the default data type of the spreadsheet is used (typically numeric).
- Blank cells for numeric variables are set to system missing and for string variables they are treated as strings
- Numeric and date variables are set to Scale and Text data to Nominal.

You will need to add Labels, Values and define Missing and, for some variables, change the Measure to the appropriate Type.

## Using the Database Wizard

The wizard allows you to capture data from various sources. These can be a database format file, e.g. MS Access, an Excel file, or a text file. You can read data from any database, or other application for which you have an ODBC (Open DataBase Connectivity) driver.

The process requires that you:

- Select the file.
- Enter a login name and password for the database if one is set.
- Select the table(s) and fields you want to read into SPSS.
- Specify any relationships between your tables.

Optionally, you can:

- Specify any selection criteria for your data.
- Add a prompt for user input to create a parameter query.
- Define any variable attributes.
- Save the query you have constructed before running it.

The example described on the next page uses a Microsoft Access database.

Click on **File | Open Database | New Query**.

A screen similar to the one opposite will appear, though you may have more or less items on the ODBC Data Sources list.

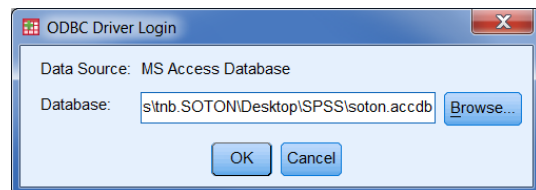
Select *MS Access Database* from the *ODBC Data Sources*.

Click on **Next**



Click the **Browse** button to display an *Open File* dialogue box that enables you to select the Access database, which contains the data.


Click on **OK**



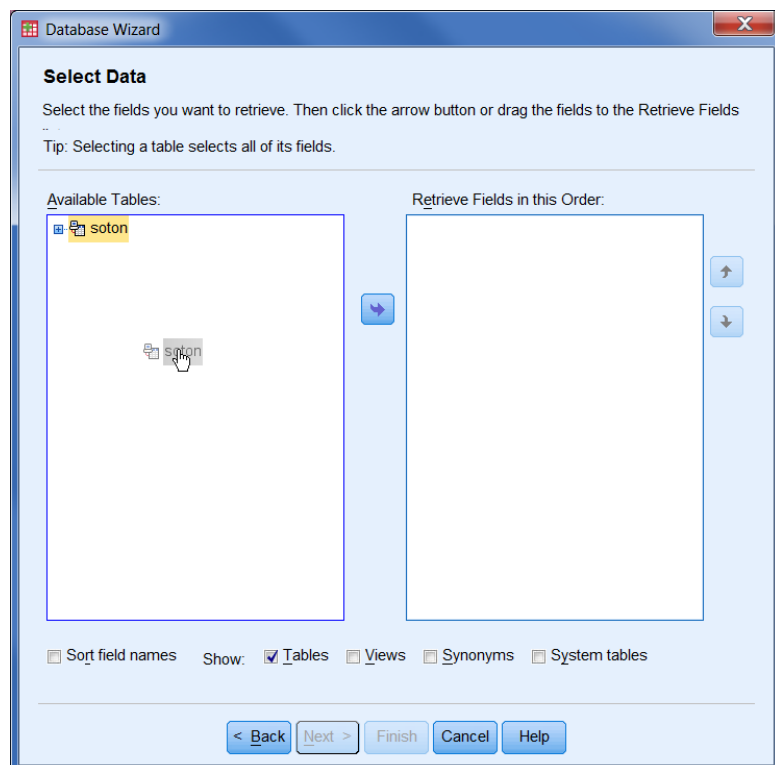
A *Select Data* window is displayed.


Move the cursor over the table you wish to use. In this example, there is only one, *soton*.

Select the table (*soton* in this example) by clicking on it.

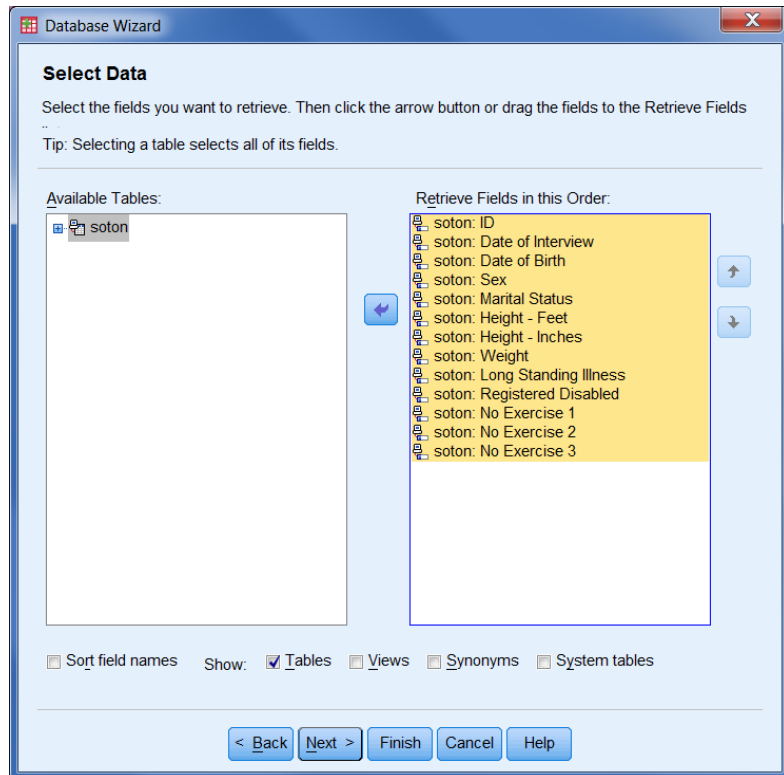
Either **click** the arrow button  or **Drag** *soton* from the *Available Tables* list to the *Retrieve Fields in this Order* list. The cursor changes shape

to 



If only some of the fields in the table are required, clicking on the plus symbol **+** next to the *soton* table name  displays all the field names in the table. One or more fields can be selected and dragged into the **Retrieve Fields in this Order** panel.

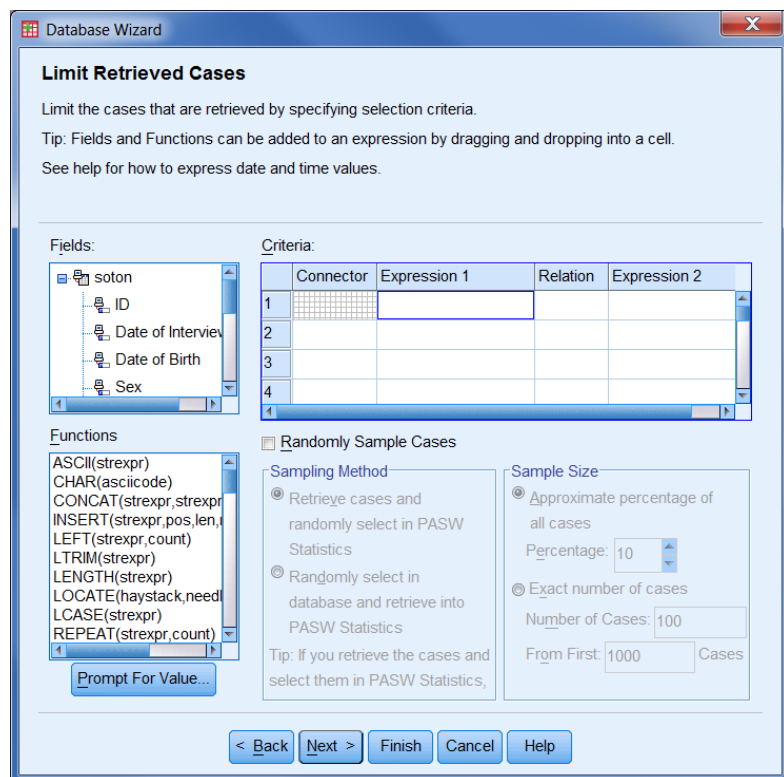
Clicking on **Next** displays the Limit Retrieved Cases dialogue.



The **Limit Retrieved Cases** dialogue allows selected data to be extracted from the database using various criteria. In this example all the data is required, so no criteria are specified.

*Note: There are features in SPSS Statistics to analyse selected cases from the whole data set, these are described elsewhere in these notes.*

**Click on Next**

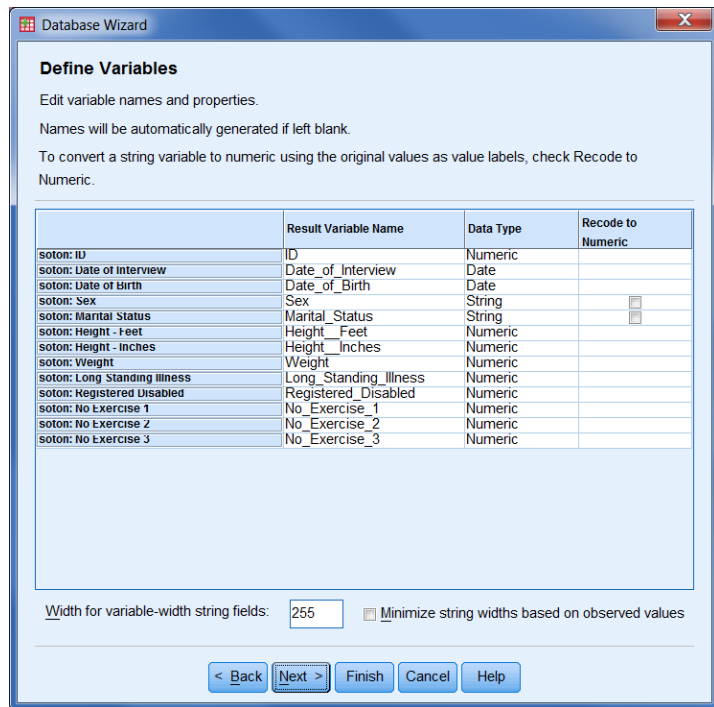


*Note: The next optional step of the Wizard does not appear in this example because only one table was selected. If two or more tables are selected, then a Specify Relationships dialogue is displayed.*

The *Define Variables* dialogue displays the database field names (in the grey column) and the names that SPSS will use (*Result Variable Name*).

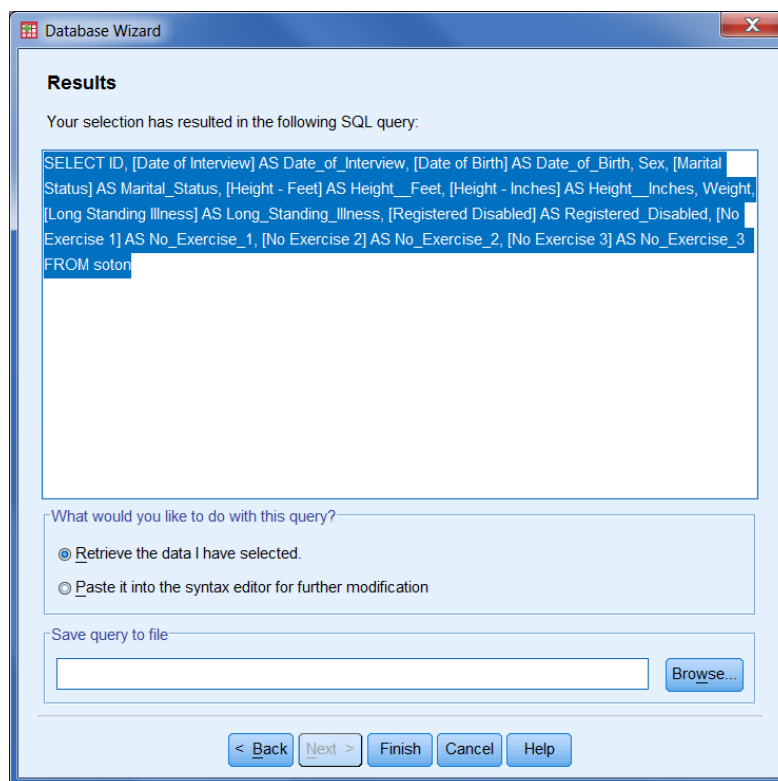
SPSS does not allow the space character in a variable name, so the underscore `_` is substituted wherever a space occurs. The *Result Variable Name* is modified to conform to SPSS naming rules. In this example the minus character has been removed from Height – Feet and Height – Inches and two adjacent underscores replace spaces.

Text (string) database fields can be *Recoded to Numeric* variables, using the original text values as labels, by checking the box  .



### Click on Next

A *Results* dialogue with an SQL query is shown. At this stage the data can be retrieved directly from the Database table(s), the SQL query can be saved to a file for later recall, or the query can be pasted into a Syntax window as a GET CAPTURE ODBC command.



Finally, **Click** on the **Finish** button and run any commands pasted into the Syntax Window.

The data is placed in the Data Editor.

\*Untitled3 [DataSet2] - PASW Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : Date\_of\_Birth 14-Jun-1945 00:00:00 Visible: 13 of 13 Variables

	ID	Date_of_Interview	Date_of_Birth	Sex	Marital_Status	Height_Feet	Height_Inches	Weight	Long_Standings_Illness
1	26.00	2-Jul-2004 ...	14-Jun-19...	f	married	5.00	8.00	.	2.00
2	29.00	8-Jun-2004 ...	16-Apr-19...	m	married	5.00	11.00	.	2.00
3	41.00	19-Jun-200...	3-Jun-194...	f	widowed	5.00	2.00	68.50	3.00
4	59.00	14-Aug-200...	25-Feb-19...	m	single	6.00	4.00	54.00	2.00
5	69.00	3-Aug-2004...	24-Mar-19...	f	single	5.00	4.00	77.00	2.00
6	94.00	10-Jul-2004...	13-Dec-19...	f	married	5.00	6.00	58.10	2.00
7	143.00	24-Jul-2004...	24-May-1...	f	widowed	5.00	2.00	68.50	2.00
8	160.00	23-Jul-2004...	19-Aug-19...	f	married	5.00	8.00	54.00	2.00
9	186.00	15-Jul-2004...	11-Jun-19...	m	single	6.00	2.00	76.20	1.00
10	211.00	9-Aug-2004...	17-Jan-19...	m	married	6.00	.0	80.70	1.00
11	224.00	11-Aug-200...	16-Jan-19...	f	widowed	5.00	2.00	89.80	2.00
12	244.00	3-Jun-2004 ...	3-Jan-197...	m	married	5.00	4.00	52.20	2.00
13	250.00	25-Jul-2004 ...	28-Aug-19...	f	single	5.00	7.00	60.00	2.00

Data View Variable View

PASW Statistics Processor is ready

Some changes to the variable definitions may be required, for example date variables have a width of 20 which includes a time element. To remove the time part, change to the **Variable View** tab in the Data Editor select the Type cell and change the format of the variable type to dd-mmm-yyyy. Modify the attributes for each variable, changing Name, Width, Decimals, Labels, Values, Missing, and Measures as necessary.

*Note: Database Capture offers more flexibility than the earlier two approaches (Copy & Paste and File Open) and can be used with Excel files.*