

SPSS :

(Predictive Analytics SoftWare)

Getting started



Peter Browne
IT Services
December 2009

Table of Contents

| | |
|--|----|
| Overview of the course | 1 |
| 1 Brief description and history of SPSS (PASW) | 1 |
| 2 Running SPSS and the Initial window(s) | 1 |
| 2.1 Running SPSS | 1 |
| 2.2 The Initial SPSS window(s) | 1 |
| Overview | 1 |
| The Title Bar | 2 |
| The Menu Bar | 2 |
| The (Power) Tool Bar | 2 |
| The Status Bar | 2 |
| 3 Sample SPSS session | 3 |
| Overview of this guided exercise | 3 |
| 3.1 Open File | 3 |
| 3.2 Summarize Cases (formerly List Cases) | 4 |
| 3.3 Frequencies | 6 |
| 3.4 Explore | 6 |
| 3.5 Graphics | 9 |
| 3.6 Non Parametric Wilcoxon Test | 12 |
| 3.7 Correlation | 13 |
| 3.8 Regression (Optional) | 14 |
| 4 Creation of a small data file and computation of new variables | 15 |
| Overview | 15 |
| 4.1 Preliminary considerations about data structures | 15 |
| 4.2 Creation of a data dictionary | 16 |
| 4.3 Entering data | 18 |
| 4.4 Moving around the data | 19 |
| 4.5 Editing data | 19 |
| 4.6 Computation of new (or existing) variables | 20 |
| 5 Help | 21 |
| SPSS, Facebook, Twitter and other web links | 21 |
| Acknowledgements | 21 |
| Appendix A | 22 |
| Menu Bar | 22 |
| Appendix B | 23 |
| The (Power) Tool Bar | 23 |
| Appendix C | 25 |
| Running (and installing) a networked version of SPSS | 25 |
| The School of Medicine and Dentistry Service | 25 |
| The Teaching Service (SPSS12.0.1) | 25 |
| Appendix D | 26 |
| A quick method for Contingency Tables | 26 |

Major (and minor) Changes.

V1. Major interim revision for version 16

Extra text explaining the purpose of commands added.

Dialog boxes replaced and important features highlighted.

Extra exercises for quick workers.

V2. New name (PASW), cover icon, owner (IBM) and refs to newer versions of SPSS & Windows. Revised Appendices.

V3. Minor corrections, pictures of the Normal Distribution and completed data in 3.4

V4. Locations of World95.sav in 3.1, New Designate icon in App B and tidying up sections 4.2, 4.3 and 4.4.

SPSS: Getting Started

Overview of the course

1. Brief description and history of SPSS
2. The Initial SPSS window(s)
3. Sample SPSS session with an existing file
4. Creation of a small data file and computation of new variables
5. Help, Tutorials and Web links (if time permits)

1 Brief description and history of SPSS (PASW)

SPSS is an acronym for **S**tatistical and **P**resentational **S**ystem **S**oftware. (Previously it stood for Statistical Package for Social Scientists). It is a fairly comprehensive program and you should be able to perform 95%+ of your analyses with it. It encompasses a wide range of statistical techniques from basic descriptive statistics through to regression analysis and graphics.

These notes are based on SPSS version 16 for PCs, although reference is made to other versions. Macintosh versions are broadly similar. Version 16+ has also been implemented under Linux (Red Hat and Debian).

By the way, the computer giant IBM acquired SPSS on 2/10/2009. SPSS 17 and 18 are now known as PASW (**P**redictive **A**nalytics **S**oft**W**are).

2 Running SPSS and the Initial window(s)

2.1 Running SPSS

Start the SPSS program by one of the following ways ...

- 1) From these training lab PCs, click on the following sequence ...

Start ⇒ Programs ⇒ SPSS ⇒ SPSS 16.0

Or

- 2) From the Student Teaching Service computers, the sequence is ...

The Task bar and click on the Public Taskbar option or fly out 

Select the 1-Maths and Stats group and then the SPSS 12 program.

For more information about starting & installing SPSS, see Appendix C.

2.2 The Initial SPSS window(s)

Overview

1. The Title Bar
2. The Menu Bar
3. The (Power) Tool Bar
4. The Data Editor (Data View and Variable View)
5. The Status Bar

2.2 The Initial SPSS window(s) (continued)

The Title Bar

The traditional Windows bar at the top of screen with Minimize, Restore/Full Size and Close buttons.

The Menu Bar

Many of the tasks that you want to perform with SPSS start with menu selections. Different SPSS windows have their own menu bars with menu selections appropriate for that window type. The Analyze (formerly Statistics) and Graphs menus are available on all windows, making it easy to generate new output without having to switch windows. See Appendix A for further details.

The (Power) Tool Bar

Different SPSS windows have their own Tool Bars that provides quick, easy access to common tasks. Some windows have more than one Tool Bar. Tool Tips (SPSS) or Screen Tips (Windows) provide a brief description of each tool when you put the mouse pointer on the tool. More information is provided in the Status bar at the bottom of the screen. See Appendix B for further details.

The Data Editor (Data View and Variable View)

The Data Editor provides a convenient, spreadsheet-like method for creating and editing data files. The Data Editor window opens automatically when you start a session.

The Data Editor provides two views of your data:

1. Data view. Displays the actual data values or defined value labels.
2. Variable view. Displays variable definition information, including defined variable and value labels, data type (for example, string, date, and numeric), measurement scale (nominal, ordinal, or scale), and user-defined missing values.

In both views, you can add, change, and delete information contained in the data file.

The Status Bar

The status bar at the bottom of each SPSS window provides the following information: Expanded screen tips, Commands, Filters, Weights and Split Files. See the SPSS Help system for explanations of these terms.

| |
|---|
| For more detail on this section, visit the hidden local SPSS Primer web page. |
|---|

| |
|---|
| http://www.stu.qmul.ac.uk/primer/spss.html |
|---|

3 Sample SPSS session

Overview of this guided exercise

1. Opening data files (Open)
2. Listing your data (Summarize)
3. Exploratory Data Analysis (Frequencies and Explore)
4. Further Data Analysis (Graphics, Wilcoxon test, Correlation & Regression)

Note. Exercises are denoted with the word 'Exercises' underlined. There are extra exercises for '*Quick Workers*', which are in italics.

3.1 Open File

You may either work with multiple existing saved files and / or create a file yourself. SPSS has variety of bundled saved files for you to experiment with. One such general file is called "World95.sav" and another called "GSS93 subset.sav".

Exercise.

Open the data file called World95.sav using the below navigation

Navigation Notation example for the PC Lab computers.

Within SPSS, go to File menu, click on the Open menu & Data... dialog box. Look for the 'My Documents', click on it and then select the SPSS folder.

To reduce verbiage this becomes ...

File ⇒ Open ⇒ Data... ⇒ My Documents ⇒ SPSS folder

The file World95.sav should now appear in the Data View.

Observe what happens when you click on the Variable View tab.

Examine the Variable spreadsheet.

Now, return to the Data spreadsheet and find climate (at the end).

Recall just beneath the Menu bar is the tool bar (TB).

When you toggle the Value Labels button (the last but two buttons) on the Tool Bar,



Note what happens.

Location of the sample files.

The location of this sample file (and others) is system dependent.

In the Mile End PC Training room, World95.sav is located in several places.

- 1) The SPSS sub folder within My Documents folder for convenience.
- 2) The SPSS folder on the networked H: drive.

Student Teaching Service (including the RBC and Cancer IT Training Labs).

The SPSS folder on the networked L: drive.

SMD Staff Novell Netware servers for users with a networked SPSS 12 version installed.

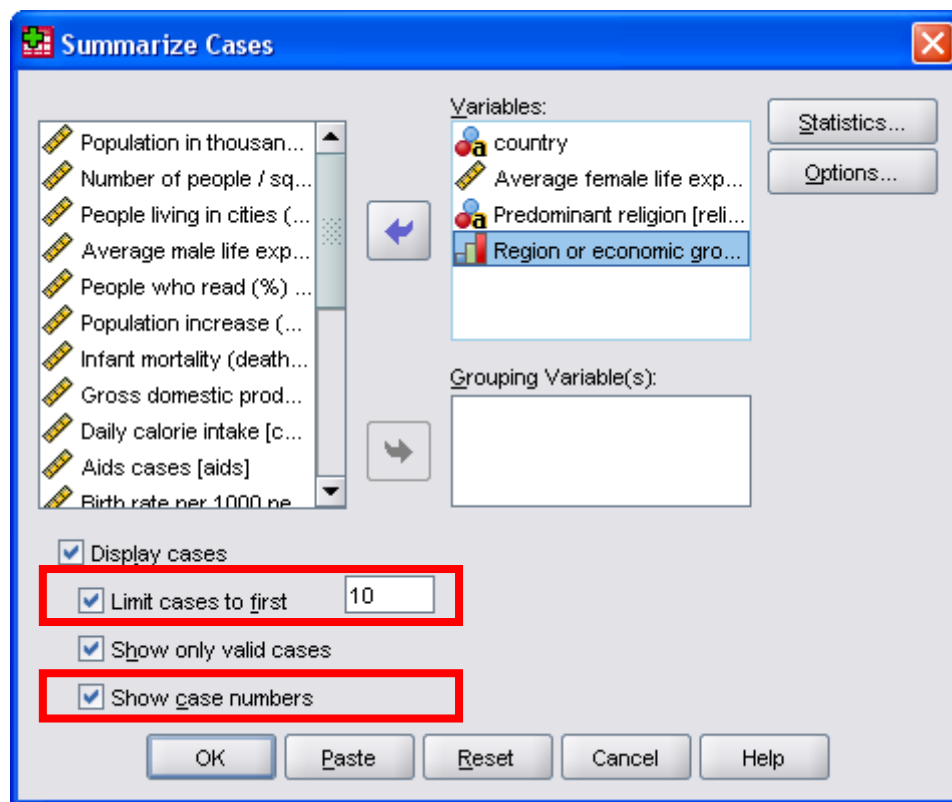
The SPSS 12.0.1 folder on the networked I: drive.

On a personal PC with SPSS installed, the typical location is given below.

C:\Program Files\SPSSInc\SPSS16\Samples

3.2 Summarize Cases (formerly List Cases)

When you have created a data file, you check that your data is correct. One simple way is use the Summarize Cases command (and print out your data).




Exercise

Navigate to the dialog box, via the navigation notation (see the previous page) ...
Analyze ⇒ Reports ⇒ Case summaries...

Complete the Summarize Cases dialog box.

Select the (source) variables on the Left Hand Side, which you are interested in and move them to the (target) Variables: list on the Right Hand Side by way of the curved blue arrow button between the two areas.

Hot Tip. You may use the mouse together with the Shift or Ctrl keys to select adjacent or non-adjacent groups of variables respectively.
Aside. This technique also works in Windows (File) Explorer. ( and E)

Limit the cases to first 10 (as opposed to 100, which is the default).

Hot Tip. Additionally for large data sets, select the Show case numbers checkbox. This is particularly useful for tracking down and correcting errors.

When the dialog box is complete, all the buttons on the bottom left hand side are active, that is to say, they are NOT greyed out.
Hit the OK button.

Note. OK etc. buttons have been 'transposed' with Statistics etc. ones in SPSS16+.

3.2 Summarize Cases (continued)

SPSS now opens an output window, which contains your results. Note the resemblance to the Windows (File) Explorer. The left hand pane displays your results in a tree like headings. The right hand pane has the current highlighted results.

Note. The Menu bar and (Power) tool bar buttons have changed in the Output Window. The same is true for the Chart Editor and Syntax / Paste Windows.

Hot Tip. You can export to other formats, for example SPSS 11.5+ to Word / RTF Documents and SPSS 15+ to Portable Document Files (PDFs).

The screenshot shows the SPSS Viewer window with the following content:

```

GET
FILE='C:\Program Files\SPSSInc\SPSS16\Samples\world95.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
SUMMARIZE
/TABLES=country lifeexpf religion region
/FORMAT=VALIDLIST CASENUM TOTAL LIMIT=10
/TITLE='Case Summaries'
/MISSING=VARIABLE
/CELLS=COUNT.
  
```

→ Summarize

[DataSet1] C:\Program Files\SPSSInc\SPSS16\Samples\world95.sav

Case Processing Summary^a

| | Cases | | | | | |
|--------------------------------|----------|---------|----------|---------|-------|---------|
| | Included | | Excluded | | Total | |
| | N | Percent | N | Percent | N | Percent |
| country | 10 | 100.0% | 0 | .0% | 10 | 100.0% |
| Average female life expectancy | 10 | 100.0% | 0 | .0% | 10 | 100.0% |
| Predominant religion | 10 | 100.0% | 0 | .0% | 10 | 100.0% |
| Region or economic group | 10 | 100.0% | 0 | .0% | 10 | 100.0% |

a. Limited to first 10 cases.

Case Summaries^a

| | Case Number | country | Average female life expectancy | Predominant religion | Region or economic group |
|-------|-------------|-------------|--------------------------------|----------------------|--------------------------|
| 1 | 1 | Afghanistan | 44 | Muslim | Pacific/Asia |
| 2 | 2 | Argentina | 75 | Catholic | Latn America |
| 3 | 3 | Armenia | 75 | Orthodox | Middle East |
| 4 | 4 | Australia | 80 | Protstnt | OECD |
| 5 | 5 | Austria | 79 | Catholic | OECD |
| 6 | 6 | Azerbaijan | 75 | Muslim | Middle East |
| 7 | 7 | Bahrain | 74 | Muslim | Middle East |
| 8 | 8 | Bangladesh | 53 | Muslim | Pacific/Asia |
| 9 | 9 | Barbados | 78 | Protstnt | Latn America |
| 10 | 10 | Belarus | 76 | Orthodox | East Europe |
| Total | N | | 10 | 10 | 10 |

a. Limited to first 10 cases.

3.3 Frequencies

The purpose of Frequencies is to provide simple descriptive frequencies (and optionally graphics) on (both nominal and ordinal) qualitative data. Frequencies may be used for quantitative data, but Explore is a better command.

Exercise.

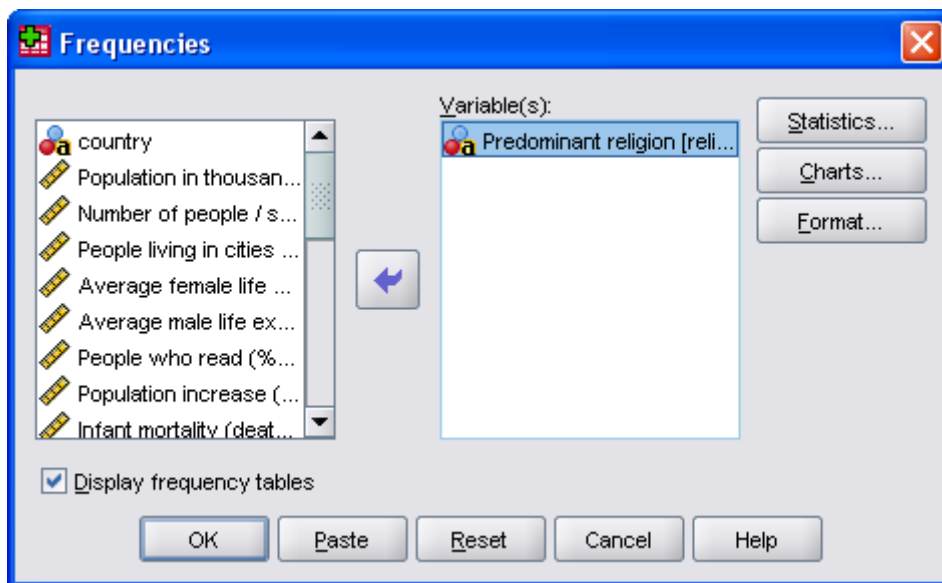
Look at Religion, which is a string or text variable.
(Note the **a** with blue and red circles.)

Quick workers - look at Region too.

Use the concise navigation notation:

Analyze ⇒ Descriptive Statistics ⇒ Frequencies...

Move the Religion variable from the source list to the (target) Variable(s): list using the blue curved arrow. When complete, hit the OK button.



3.4 Explore

Explore command is very useful producing statistics on continuous quantitative data. The variable(s) may be broken down by Factors. Usually, this is one at a time. However a special Syntax box technique is employed to breakdown a variable(s) by two or more factors simultaneously. See the Hot Tip box at the end of this section. Explore allows the user to produce graphics, such as Box and Whisker Box Plots, Histograms and Stem & Leaf plots. More importantly, the user can produce Probability plots with Tests of Normality. The Shapiro-Wilks is the 'Gold standard'. The purpose for testing for Normality, helps you decide, whether to employ Parametric or Non-Parametric Statistical tests for future analyses. Here a non statistically significant test implies an underlying bell shaped Normal distribution.

Lifeexpf is the short name for Average Female Life Expectancy and is the main focus of this sample SPSS session.

Note the diagonal yellow ruler. This denotes, that it is a (continuous) measurement.

Exercise

Produce descriptive statistics for lifeexpf, and test for Normality.
What information is provided ? Is the data Normally distributed ?

Quick workers - use the help system to find about a) Normal Plots and b) Boxplots.

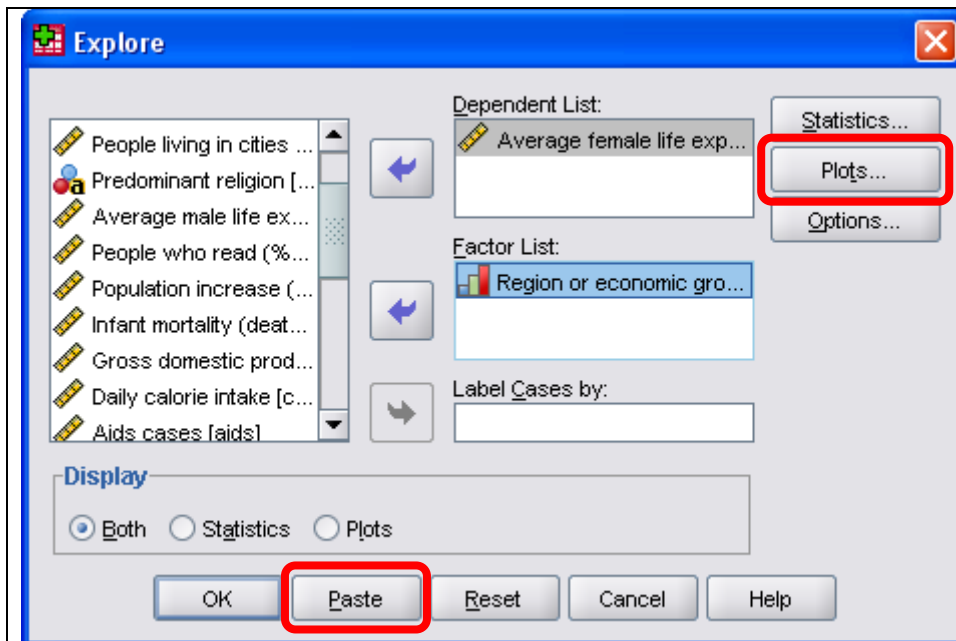
Use the notes below to help you.

Analyze ⇒ Descriptive Statistics ⇒ Explore...

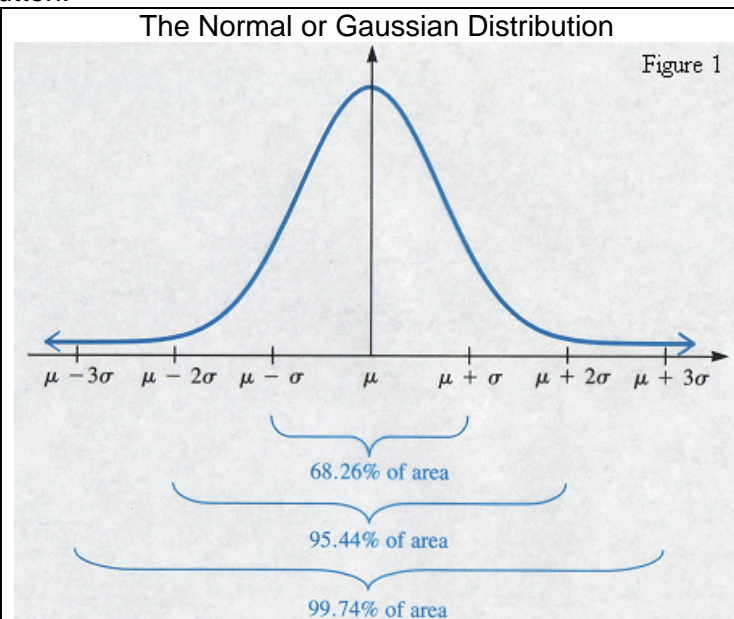
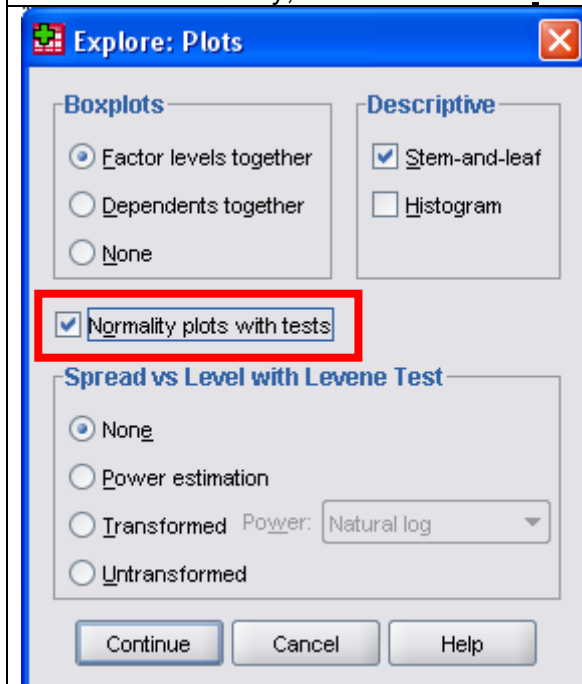
From the source list, move Lifeexpf into the Dependent List: textbox.

If you wish, you may breakdown the data by Region or Religion.

To do this move Region (or Religion) into the Factor List: box



To test for Normality, first click on the Plots button.



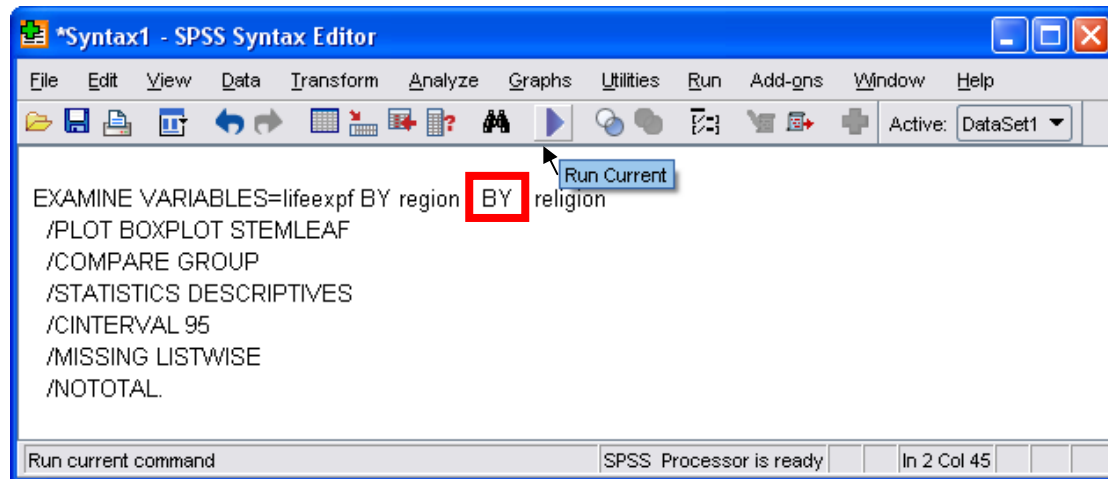
Courtesy of Hawaii University

Tick the Normality plots with tests check box. Click on the Continue and OK buttons.

3.4 Explore (continued.)

Hot Tip.

Suppose you wish to break down your analysis by TWO or more variables at the SAME time. This can be achieved via the Paste button. See the previous page Click on this button and a SPSS Syntax / Paste Window appears. See below.



Go to the EXAMINE VARIABLES=Lifexpf line BY region and append 'BY religion' without the quotes.

Thus two BYs are required for a breakdown with two variables.

Aside. The keyword is BY, which is a reserved word. You cannot have a variable called BY.

Hit the play or Run button.



3.5 Graphics

One of the biggest changes between SPSS 16 and much earlier versions, is the way that Graphics are produced. The newer versions use a more interactive technique called Chart Builder for creating graphs, whereas older versions use the more traditional legacy method of dialog boxes.

Hot Tip. With version 13 and above, graphics may be exported as PowerPoint files.

Graphics Exercise.

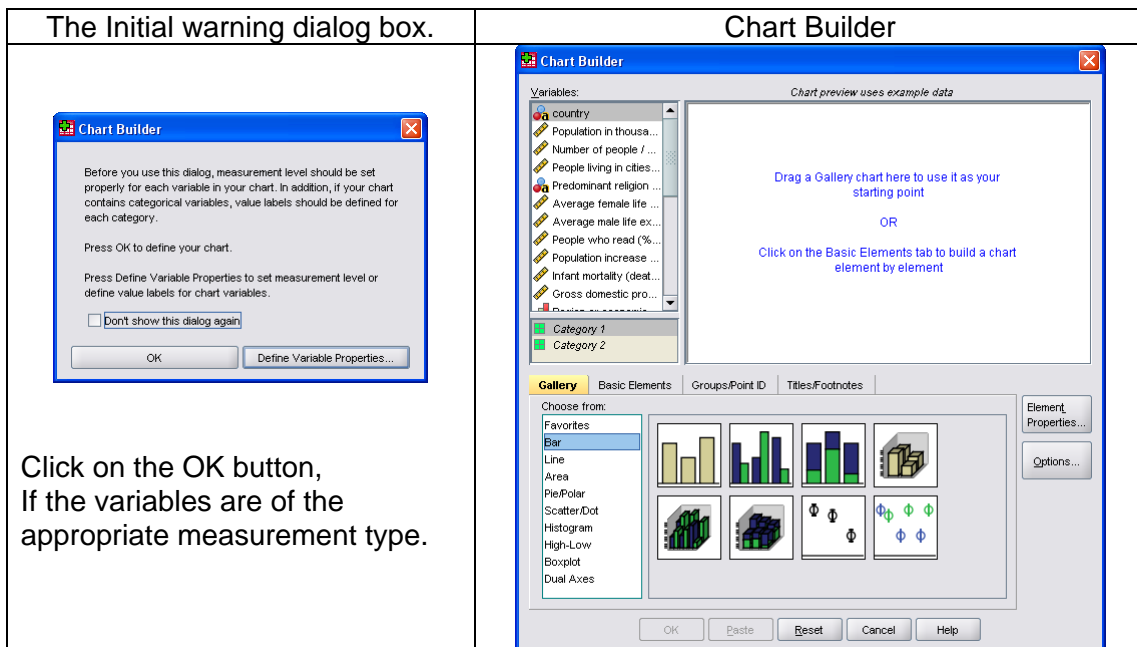
Create a Scattergram between Male and Female Life Expectancy (Lifeexpfm and Lifeexpm respectively) with the Region as the marker.

Use the notes below to guide you.

Quick workers - do BOTH methods

Chart Builder

Graphics ⇒ Chart Builder ...



Spend a few moments examining the tabbed dialog box.

To produce a scattergram, go to the **Gallery** tab and from the **Choose from:** list select the **Scatter/Dot** option. The sub types of scatter plot become available.

Next within the Gallery tab, select the first row & second column and a mock up of the chart appears in the top right corner

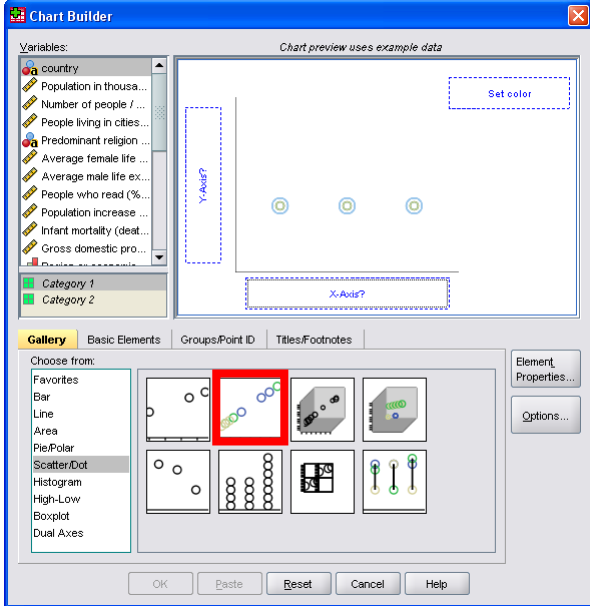
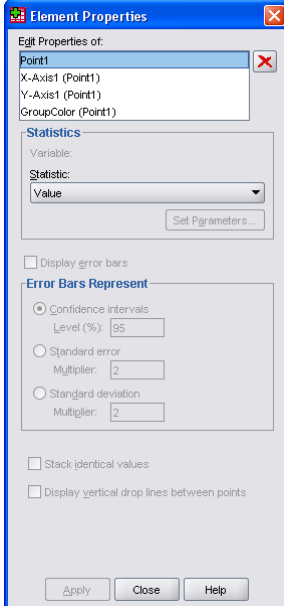
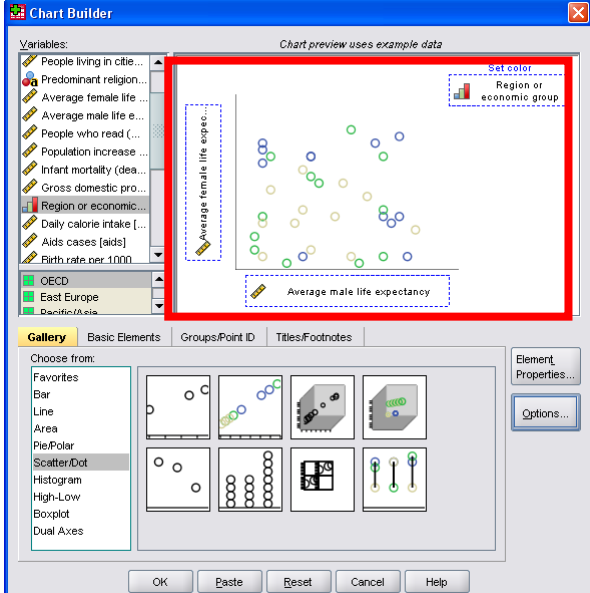
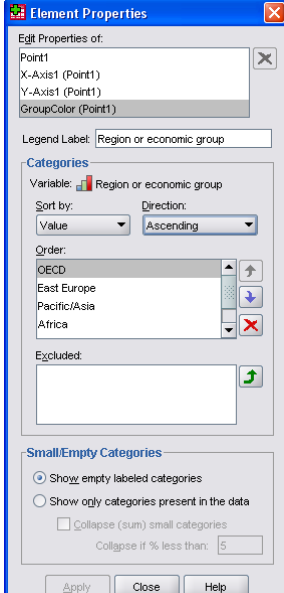
To complete the chart, just **drop and drop** the required variables in the appropriate positions.

Y axis: lifeexpf (Average Female life expectancy)

X axis: lifeexpm (Average Male life expectancy)

Setmarkers by: region (World region including OECD)

3.5 Graphics (continued)

| Chart Builder (Scatter/Dot) | Element Properties |
|---|---|
|  |  |
| Completed Chart | Element Properties |
|  |  |

If you wish to make further modifications, use the Element Properties.
 To close the Element Properties, click on the Element Properties button.
 (This button is a flip-flop switch. Click to close, click again to open and click to close.)
 Finally, click on the OK button.

The chart now appears in the Output window.

Observe the plot. What do you note ?
 Hint. Are there any sub-populations ?

3.5 Graphics (continued)

Traditional legacy dialog method

Version 16

Graphics ⇒ Legacy Dialogs ⇒ Scatter/Dot ⇒ Simple Scatter ⇒ Define

Old versions.

Graphics ⇒ Scatter ⇒ Simple ⇒ Define

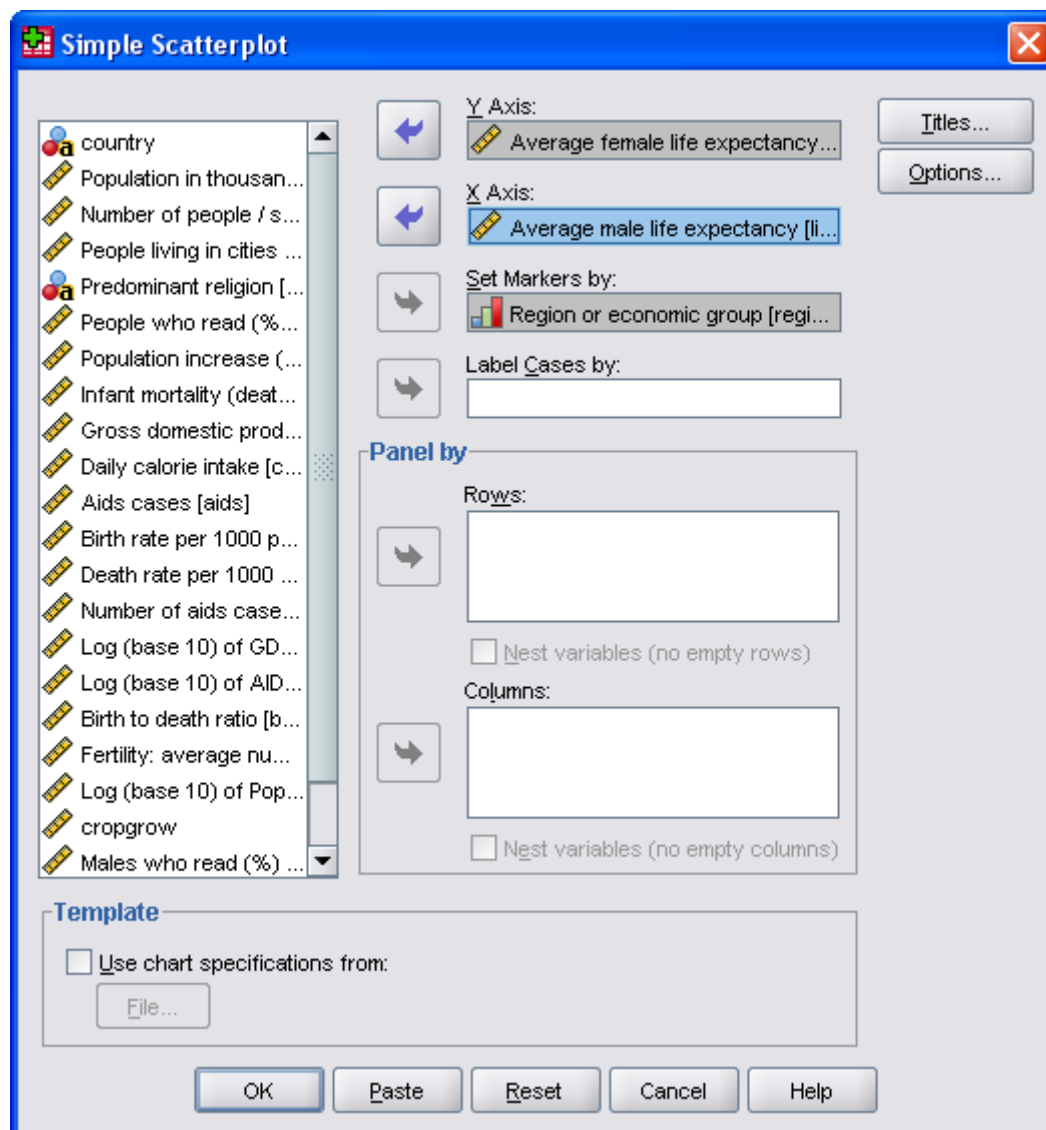
Complete the Simple Scattergram dialog box, as shown on this page.

Y axis: lifeexpf (Average Female life expectancy)

X axis: lifeexpm (Average Male life expectancy)

Set Markers by: region (World region including OECD)

Hit the OK button.



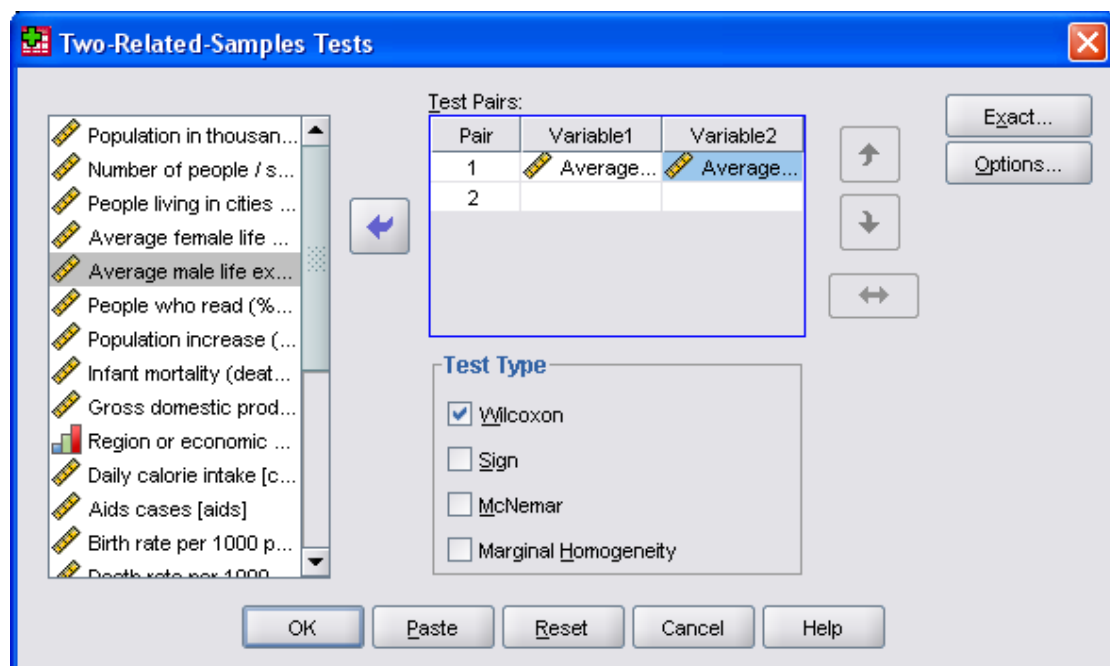
The chart, which may be further edited, now appears in the Output window. Double click on the chart to edit it. Use the Help system to assist you.

Observe the plot. What do you note ? Hint. Are there any sub-populations ?

3.6 Non Parametric Wilcoxon Test

Non Parametric tests are employed, when the data does NOT have an underlying Bell shaped Normal (or Gaussian) distribution. (Aside. Sometimes, they may be transformed to approximate Normality.) This family of tests are based on the ranks of the data, rather the raw data itself. These tests often have analogues of Normal based Parametric tests.

| Analogue tests examples | |
|------------------------------------|---------------------|
| Parametric | Non Parametric |
| Paired Student's t-test | Wilcoxon test |
| Two sample Student's t-test | Mann Whitney test |
| 1 way Analysis of Variance (ANOVA) | Kruskal Wallis test |



Exercise.

Is there a difference between men & women with regard to average life expectancy ? (Recall. The men and women are paired within each country.)

Use the information below and the dialog box above to complete this task. Perform a paired sample signed rank sums Wilcoxon test.

Analyze ⇒ Nonparametric tests ⇒ 2 Related Samples...

Move lifeexpf and lifeexpm to Test Pair(s) List:
Ensure that the Wilcoxon checkbox is selected.

Exercise.

Quick workers.

Do the above exercise using a Paired Student's T test instead.

Hint. Analyze ⇒ Compare means ⇒ Paired Samples T Test...

3.7 Correlation

Correlation is a dimensionless (that is no units) measure of linear association between two continuous variables placed on an equal footing, for example, height and weight. The measure has a range from -1 to +1. If there is a negative correlation, then small values of one variable are associated with large of the other and vice versa. Negative correlation is also known as either an indirect or inverse relationship. Positive correlation is the reverse of negative correlation and is known as a direct relationship.

Pearson is the (Normal) Parametric Correlation Co-efficient (CC) ; Spearman is the Non Parametric (CC) analogue.

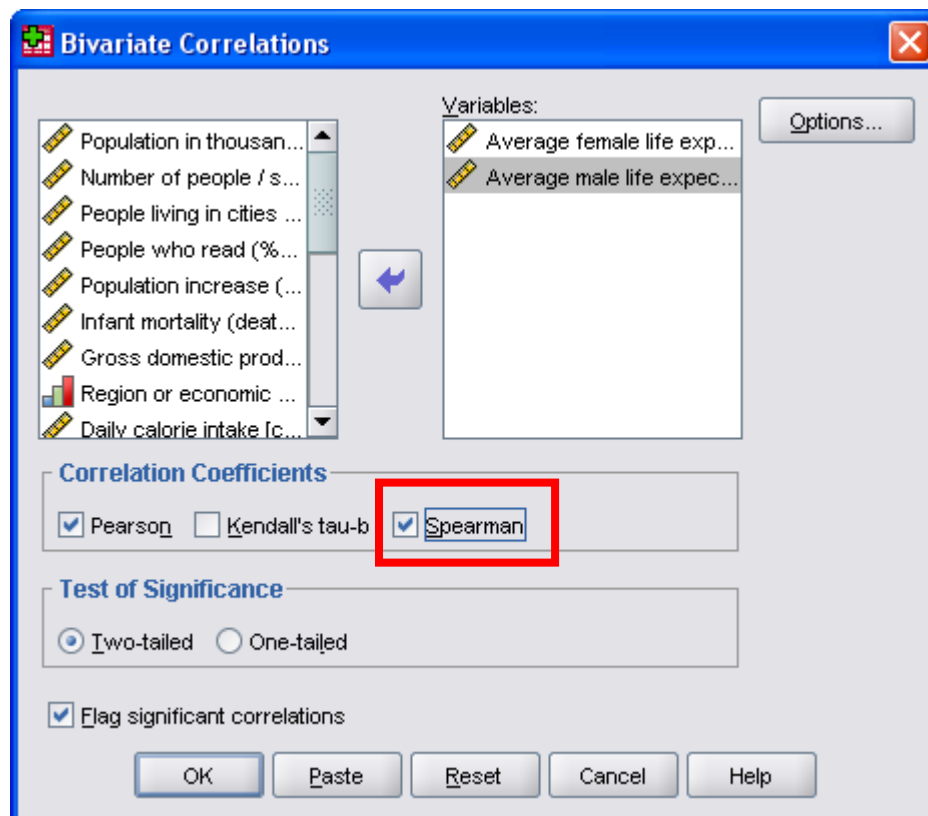
The sample cc is tested against a zero correlation (usually no association).

Exercise.

Are the two life expectancies correlated ?
(Consider Pearson and Spearman Correlation Co-efficients)

Use to the notes and dialog box to guide you.

Analyze ⇒ Correlate ⇒ Bivariate...



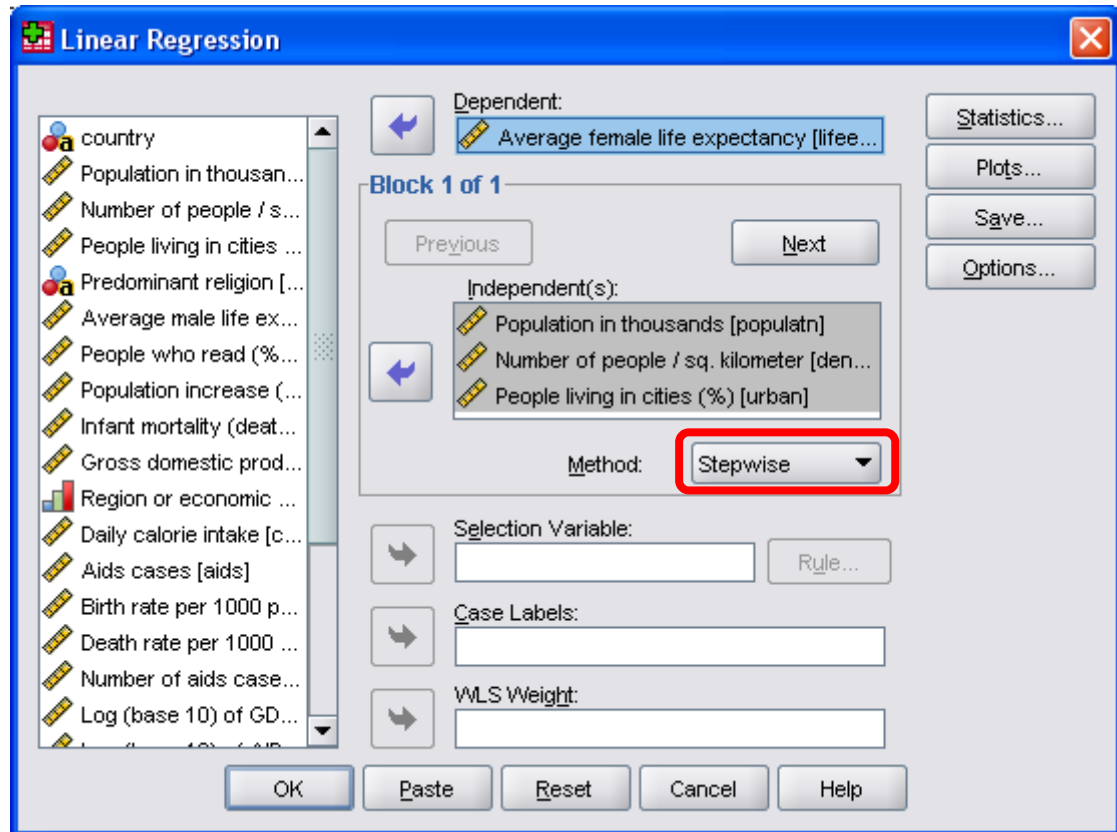
Complete the Bivariate dialog box.
Choose Lifeexpf and Lifeexpm as the (target) Variables:
Select the (Non Parametric) Spearman correlation checkbox.
Click on the OK button.

3.8 Regression (Optional)

Exercise (Fast workers only).

Which variables would help you predict the female life expectancy ?
(Ignore Country, Religion, Region and Climate, as these would need special treatment before being included in the analysis, the construction of so-called indicator or dummy variables. Also ignore MALE life expectancy - lifeexpm)

Analyze ⇒ Regression ⇒ Linear...



Make Lifeexpf the Dependent Variable.
Make ALL the other variables Independent,
except country, religion, region, climate and lifeexpm.
Method is Stepwise.

Additionally, you may specify Statistics by way of the appropriate button.
Select the following checkboxes, Confidence Intervals, R square change,
Durbin-Watson and Casewise diagnostics.
Click on the Continue button.
Also, you may specify Plots
Select the Normal probability plot checkbox.
Click on the Continue button.
Finally, click on the OK button.

Examine your results carefully.

Repeat the regression analysis but this time use Backward (elimination) as your Method.

Hot Tip. Click on the dialog recall button on the (power) tool bar
and modify the Linear Regression dialog box and set Method to Backward.

Examine your results. Are they the same as previously ? If not, what is different ?

4 Creation of a small data file and computation of new variables

Overview

1. Preliminary considerations about your data structure
2. Creation of a data dictionary
3. Entering data
4. Moving about the Data Editor
5. Editing data
6. Computation of new (or existing) variables

4.1 Preliminary considerations about data structures

Before you start to use SPSS to create your data file, you should think about the data structure, as this will have an impact on the type of analysis that you can perform. SPSS is fussy!

For example, sometimes you may be looking at two or more independent groups. In this instance, the data would look like below.

| Group type | Measurement. |
|------------|--------------|
| 1 | 4.5 |
| 1 | 4.7 |
| ... | |
| 1 | 4.6 |
| 2 | 5.2 |
| ... | |
| 2 | 5.1 |
| 3 | 5.2 |
| ... | |
| 3 | 4.9 |

On the other hand, if the data is related, for example paired readings and repeated measures, it will have this type of structure.

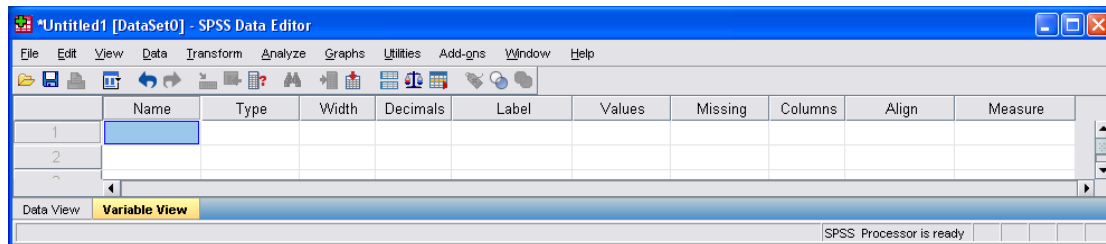
| Time 0 | Time 15 | Time 30 | Time 45 | Time 60 |
|--------|---------|---------|---------|---------|
| 11 | 12 | 13 | 14 | 15 |
| 10 | 13 | 14 | 14.5 | 15 |
| 11 | 11.5 | 12 | 13 | 14 |

4.2 Creation of a data dictionary

Refer section 2.1 for starting to SPSS and Appendix C.

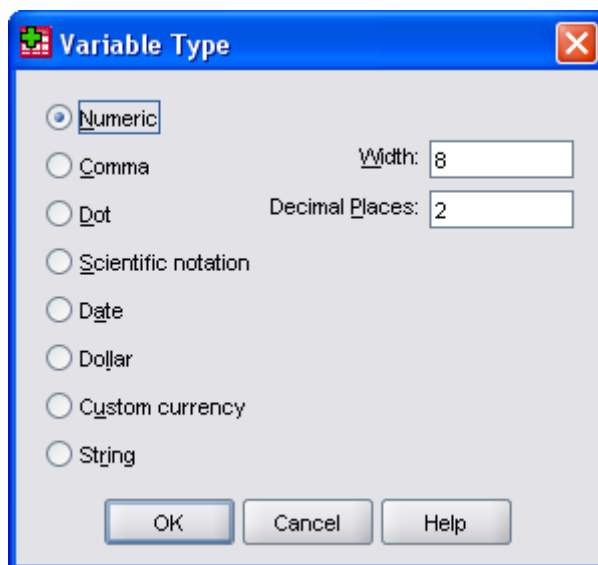
Within SPSS, click on the Variable View tab near the bottom left hand corner of the Window. This is new Spreadsheet style of building a data dictionary. Each row represents a variable name and the columns represent the variable's attributes.

The important attributes are Name, Type, Label and Label Values.



Name A variable starts with a letter of the alphabet and is up to 255 characters long.

Type Type of information to be stored, for example Numeric, String (or textual), Currency and Dates. Numeric is the default type. Make your data as numeric as possible. For dates, choose dd/mm/yyyy format to avoid Y2K problems. For more information about dates, use the Help System.



Width Different amounts of space allocated to different variable types. For example, numeric has a width of 8.

Decimal Number of decimal places (dps) for numeric data. The default is 2.

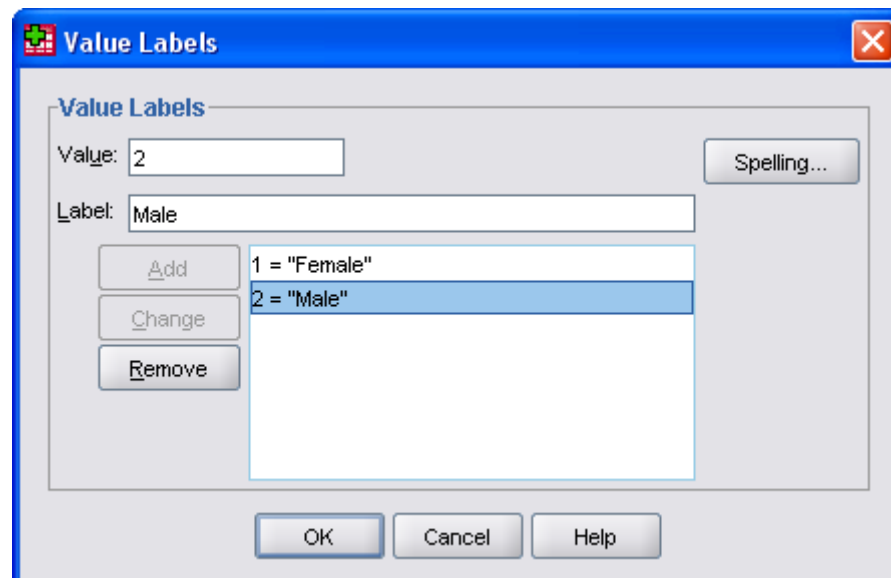
Hot Tip. Minimise the number of dps. This makes data checking easier.

(Aside. Combining the last above two attributes for numeric data, the default result is NNNNN.DD)

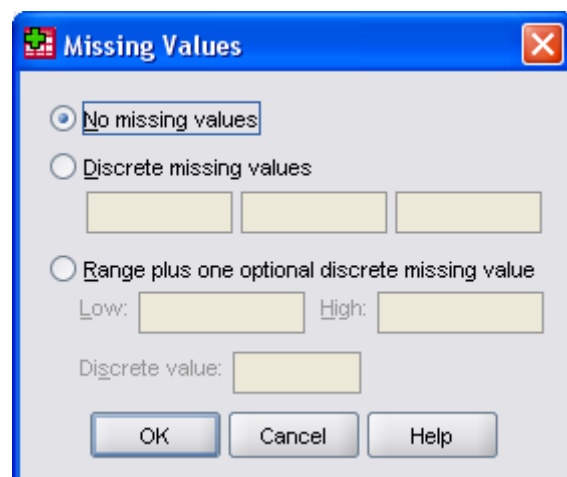
4.2 Creation of a data dictionary (continued)

Label A longer more explanatory label for the variable.
This makes printouts more readable.

Value Value Labels are very useful for qualitative data.
For example, Gender may have codes 1 and 2. Value Labels enable you to assign a piece of text to the variables, that is 1="Female" and 2="Male".



Missing. You may set up to three discrete numbers as missing.
Useful in questionnaires, if a question was not answered.
For example, 7=Interviewee did not know the answer, 8=Interviewer forgot to ask the question and 9=Interviewee refused to answer.
The default for numeric data is known as the (system) missing is ".", which represents a very large negative number.



Columns The column width in the Data View sheet. You can use the Excel trick to the column bigger/smaller. Within the Data view move the cursor the variable list heading. The cursor changes and you hold down the left mouse button and drag the mouse to a wider or narrower column.

Align Justification. Right for numeric.

Measure One of Nominal, Ordinal and (default) Scale. It has implications for graphics.

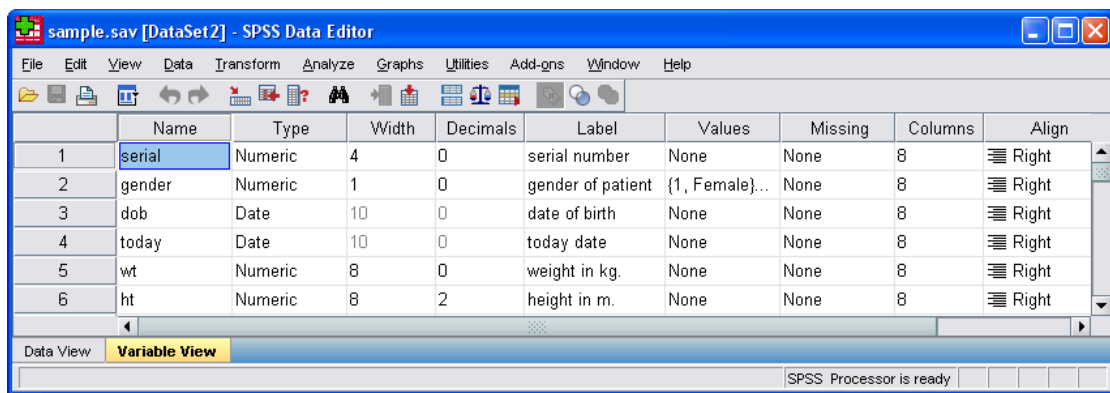
4.2 Creation of a data dictionary (continued)

Exercise

Let us create some variables. Some variables have decimal places (dps).

Serial (Serial Number - Numeric, Width of 4 and no dps).
 Gender (Gender - Numeric, Width of 1 and no dps. Code as earlier.)
 DOB (Date of Birth, Date, dd.mm.yyyy).
 TODAY (Today's date, dd.mm.yyyy)
 WT (Weight in Kgs, Numeric, Width 3 and 0 dps).
 HT (Height in Metres, Numeric, Width 5 and 2 dps).

The completed Variable View Window should look like the below screen.



When completed, please save your file. Save via the below procedure. Go to the File Menu, then Save As ... then navigate to suitable disk and folder and enter a suitable name. Note the file extension is SAV (for SAVE file). Only use SPSS to read these .SAV files, otherwise trouble !


4.3 Entering data

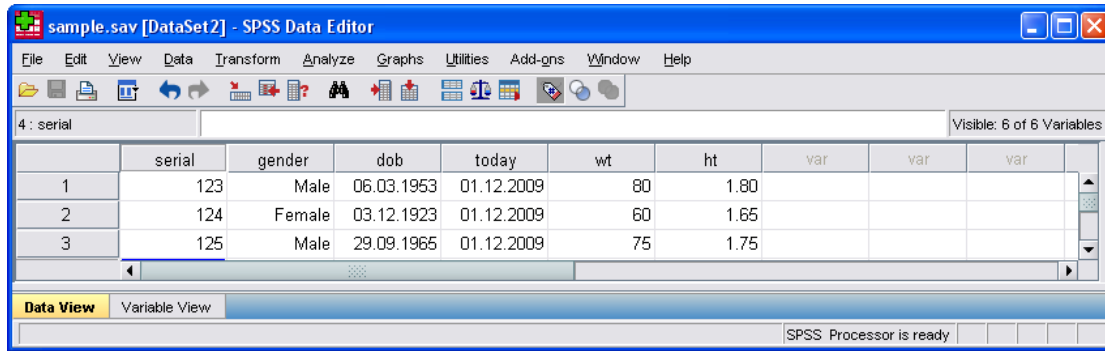
Return to the Data View and it appears like an Excel spreadsheet. However, you cannot store formulae in the data cells. The rows are cases, which may be a questionnaire, an animal in an experiment or a hospital record. The columns are items of information or variables pertaining to the case.

Exercise

Enter some fictitious data for the variables (.with the numeric key pad on).

| Serial | Gender | DOB | TODAY | WT | HT |
|--------|--------|------------|-----------|----|------|
| 123 | 2 | 06.03.1953 | 1.12.2009 | 80 | 1.80 |
| 124 | 1 | 03.12.1923 | 1.12.2009 | 60 | 1.65 |
| 125 | 2 | 29.09.1965 | 1.12.2009 | 75 | 1.75 |

Note the effect of the Value Labels button  on the (Power) Toolbar.



4.4 Moving around the data

Apart from pointing and clicking with the mouse, you may use the keyboard to help navigation. (These keyboard short cuts work in the same way for Excel.)

| | |
|-------------|---|
| Tab | to move a column at a time across a case (or row). |
| Return | to move down the same variable (or column). |
| Home | the first column of the current case (or row). |
| End | the last column of the current case (or row). |
| Ctrl + Home | the first case (or row) and first variable (or column). |
| Ctrl + End | the last case (or row) and last variable (or column). |

Exercise

Move around your data using the above mentioned short cut keys.

4.5 Editing data

Click on the data cell, which you want to change.

Either edit the data within the Data Editor, just below the (power) Tool Bar, or hit the F2 function key and edit within the data cell itself, just as in Excel.

Note, that you may 'Cut and Paste' too.

Hot tips. Keyboard shortcuts.

| | |
|----------|-------|
| Ctrl + X | Cut |
| Ctrl + C | Copy |
| Ctrl + V | Paste |

You may use the Edit Menu instead, if you prefer.

Exercise

Modify some of your data and then save your file.

Hot tip. Use the keyboard short cut of Ctrl + S.

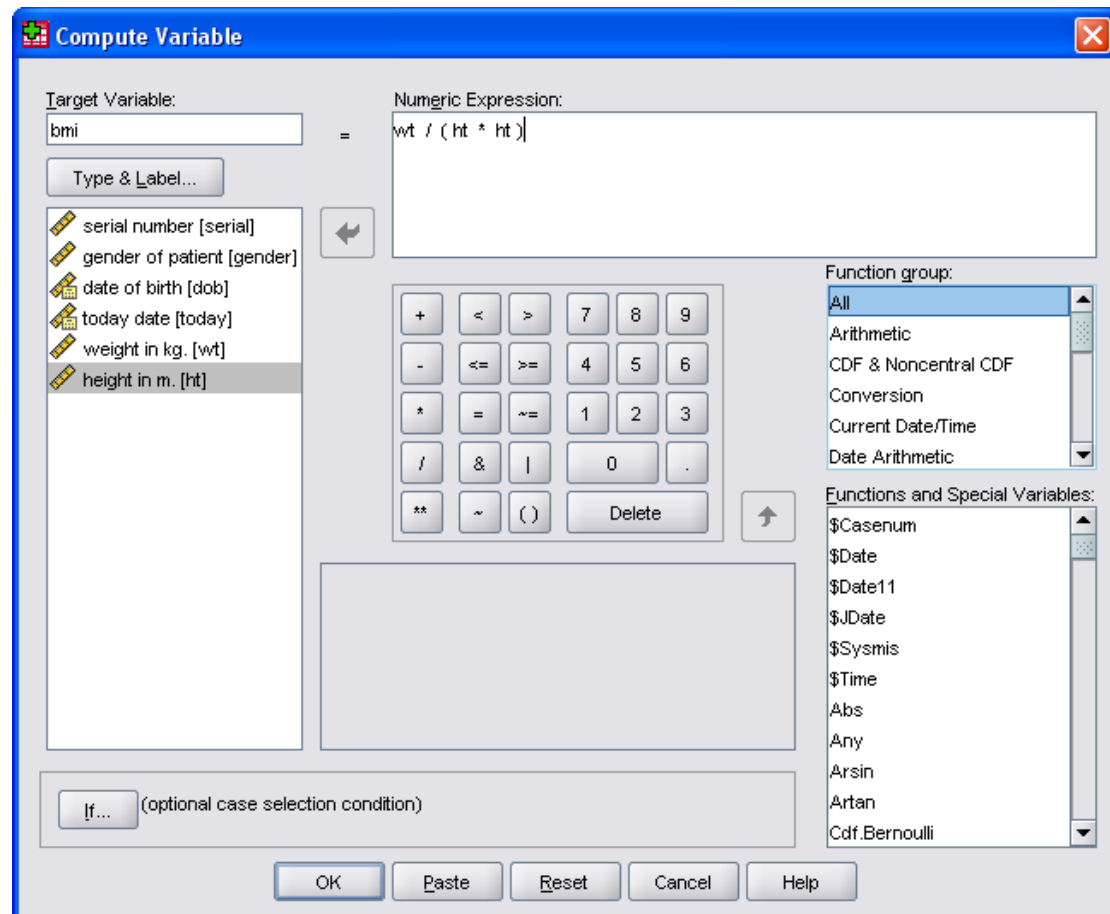
You may use either the File Menu or (Power) Toolbar Floppy button instead.

4.6 Computation of new (or existing) variables

You may compute either new variables or update existing old variables.

Hot Tip. If you add new data to your file, the derived variables will need to be recalculated. However, in this case, if you click on the Paste button in the dialog box, you may save the formula as a syntax file and reuse it later.

Ex. Compute the Body Mass Index (BMI), where $BMI = WT / (HT * HT)$



Procedure:

1. From the **T**ransform Menu, select **C**ompute... Complete the dialog box.
2. In the **T**arget Variable text box, type in BMI
3. Build up the equation in the Numeric **E**xpression text area.
 - (a) Click on the WT variable within the source variable list.
 - (b) Click the | button to transfer the variable into the Numeric **E**xpression area.
 - (c) Click the / divide symbol adjacent to the numeric pad and then () for brackets.
 - (d) Insert the mouse cursor between the brackets.
 - (e) Click on the HT variable and transfer it into the Numeric **E**xpression area.
 - (f) Click on the * multiply symbol.
 - (g) Click on the HT variable and repeat step (e).
 - (h) When complete, click on the OK button.

Extra Exercise for Quick Workers.

Compute the AGE where $AGE = (TODAY - DOB) / (3600 * 24 * 365.25)$
(N.B. SPSS counts in seconds from the start of the Gregorian Calendar.)

5 Help

If there is time, please do have a look at the extensive Help menu. As mentioned in Appendix A, Help consists of several components, such as, Topics, Tutorial and Statistics Coach. Please DO use these facilities.

SPSS, Facebook, Twitter and other web links.

<http://www.spss.com>

For technical support for which you need to register first.

<http://support.spss.com>

Information on the latest version of SPSS

http://www.spss.com/spss/whats_new.htm

Comparison between SPSS versions

http://www.spss.com/software_version/index.cfm?product=base

Social Networking Sites, such as Facebook and Twitter

<http://www.facebook.com/pages/SPSS-Inc/87247783371?ref=s>

<http://twitter.com/spssglobal>

The search engine Google is good at acquiring useful information

<http://www.google.com>

The author is (very slowly) developing a local web based resource

<http://webpace.qmul.ac.uk/pdbrowne/>

Acknowledgements

Material for the Appendices A and B were collated and adapted from:

The SPSS Primer

David Nye,
IT Services.

Using SPSS 10.0 For Windows

Instructional and Research Computing,
University of North Carolina at Greensboro.

Getting Started With SPSS for Windows

Information Systems Services,
University of Leeds.

Comments about these notes and this course are always welcome.

Please do not hesitate to contact me. Thank you.

Peter Browne

December 2009

Appendix A

Menu Bar

SPSS has at least ten main menus available when you are viewing the Data Editor window. The list of menu options changes when you are in different windows. The menus follow the standard Windows format of having the File and Edit menus to the left-hand side of the screen, and the Window and Help menus on the right. The menus found in the Data Editor window are described in more detail below:

File - The File menu deals with all the file-handling aspects of SPSS. Such as opening an existing file, opening a new file, saving a file or printing a file. There is an Apply Data Dictionary command that stores information about your data. This is especially useful if you have received data from elsewhere and you need to know what it is and how it is organised. The command to exit from SPSS when you have finished is in this menu.

Edit - This menu contains commands for undoing the last action, cut, copy and paste actions, and the command for find.

View - The View menu changes some display attributes, such as the appearance of grid lines and whether value labels are displayed instead of codes. The chosen commands are ticked.

Data - The Data menu is used to make global changes to SPSS data files, such as merging files, transposing variables and cases, or creating subsets of cases for analysis. You can insert new variables and new cases, or you can sort your cases according to a particular variable, transpose the cases or weight the cases. These changes are temporary unless you specifically save the file with the changes.

Transform - The Transform menu is used to make changes to selected variables in the data file and to compute new variables based on the values of existing ones. Random numbers can be generated and ranking or recoding of data can also be done. Changes are only temporary unless you specifically save the data file

Analyze –The Analyze (formerly the Statistics) menu contains all the categories of statistics that SPSS is capable of performing. The arrows to the right of all the menu options indicate that there is a submenu of more specific tests or groups of tests.

Graphs - The Graph menu enables you to plot graphs from the data. This may be achieved either via the Chart Builder or the (older) legacy dialogue boxes. Some of these graphs can also be plotted as part of a wider analysis. SPSS supports the five main chart types - Bar (or Column), Line, Area, Pie and High-Low. It can also produce real histograms, quality control graphs, scatterplots in up to 3 dimensions, and several other specialist graph types (version 9.0 can also produce an ROC curve). One extra graph SPSS supports is the Boxplot, used mainly in (small sample) non-parametric analysis. See section 3.5.

Utilities - The Utilities menu contains a list of variables currently available from the data sheet; it allows you to define sets of variables that need to be analysed together. Finally the Utilities menu allows you to run a script or change the menus using the menu editor.

Add-ons

New version 16+, Add-on features, such as Neural Networks and online Statistics Guides.

Window - The Window menu allows you to select and resize the windows used in SPSS. Most of the commands here can also be done with the mouse.

Help - Contains the commands for the SPSS Help system.

Consists of :- Topics, Tutorial, Statistics Coach, Syntax guide & SPSS Home Page link.

Appendix B

The (Power) Tool Bar



File Buttons

The first three buttons give the three most common commands from the File menu – Open an Existing File, Save the Current File and Print the Current File respectively.



Dialog Recall

This button gives you immediate access to the last 12 dialog boxes you were working with. This is extremely useful if you are building up an analysis and constantly going back to the same box to change or modify an option.



Go to Chart

In older versions only, this button allows you to go to an open Chart Editor.



Go to Data

When in a window other than the Data Editor window, this button will return you to the Data Editor.



Go to Case

This button allows you to go directly to a specific case in the data editor. It is useful if you need to edit your data, or if something unusual has cropped up in your analysis and you want to check out the source data.



Go to Variable

Clicking on this button produces a dialog box containing a list of all the variables defined in the data file. Selecting a variable from this list displays the variables properties - its name, label, type, information about missing values and the value labels. This box can be kept open whilst you work with the data file so that you can examine a variable's information as you examine the results of an analysis.



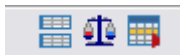
Find

This button allows you to perform a simple search to find a value.



Insert Case / Insert Variable

Data entry is fraught with errors. Consequently it is not unusual to find yourself wanting to add a case or a variable in the middle of ones already created. These two buttons will add a blank row or column in your data set to allow you to do that.



Split File / Weight Cases / Select Cases

These buttons allow you to do three of the Data menu commands with ease.



Value Labels

In SPSS you can create meaningful labels for non-meaningful numerical data. This button allows you to display the labels in the data editor to enable easier data checking.



Use Sets

You can group variables together into sets so that the variables can be analysed together. This button allows you to specify what sets of the ones you have defined you want to use.



or for older versions Designate Window Button

SPSS allows you to have more than one Output or Syntax window open at any one time so that the results you create can be sorted into different files. One of the Output windows has to be the Designated Window. This is indicated by a green cross (or in older versions a red exclamation mark) in the toolbar. If you want an Output or Syntax window to become the designated window, make it the active window and click this button on the Tool Bar.

Note: The Designated Window is not the same as the Active Window. The Active Window is the window at the forefront of your screen. Anything you type, or any menus you use should happen in the Active window. Of all the windows you have open on your desktop, only one will be the Active Window. The term Designated Window only applies to Output and Syntax windows which act as a destination for SPSS produced text. There is one designated Output window at any one point in time plus one designated Syntax window.



Run Commands

This button lets you to run commands from the Paste / Syntax window. SPSS will begin from the current insertion point and go through the commands until it finishes. See section 3.4.

Appendix C

Running (and installing) a networked version of SPSS

First of all, you need to be logged into the Novell Netware system.

How you run SPSS depends on which Novell Netware Network Service you are using. There are two services:- the SMD Staff Service and the Student Teaching Service.

Hit Tip. Please, note that SPSS version 10 does NOT work with Windows XP. Conversely versions 11 and above do NOT work with Windows 95. For further information about system requirements for SPSS, visit:- <http://www.spss.com/statistics/>

The School of Medicine and Dentistry Service

The Installation Procedure (excluding Windows 95 PCs)

SPSS version 12.0.1 has been installed in the Apps volume on three SMD Application Novell Netware servers QMWCHS, QMWWSF and QMWWHT for the Charterhouse, West Smithfield and Whitechapel sites respectively.

SPSS files are located within the SPSS 12.0.1 sub-folder within SMD folder on the APPS volume and are mapped up to the I: drive. To install, navigate to the Setup sub folder with SPSS, find the Setup.exe program and double click on it to start the process. Follow the instructions as required.

The Running Procedure (excluding Windows 95 PCs)

1. Click on the Start Button.
2. Click on [All] Programs
3. Click on SPSS for Windows
4. Click on SPSS 12.0.1 for Windows
5. SPSS should start on your PC.


Hot Tip. Windows Vista works best on SPSS 16+ (or SPSS 15 with a 'hot patch'). However, Vista does work even with SPSS 12.0.1. Just ignore the OLE error message.

Hot Tip. Windows 7 works with SPSS 18. Older versions 17, 16 and 15 (the last with the Vista 'hot patch') should work with Windows 7 (subject to confirmation).

The Teaching Service (SPSS12.0.1)

This applies to registered (undergraduates and M.Sc) students and some teaching staff, who use the teaching laboratory PCs.

The Run Procedure

1. Click on the Taskbar Options flyout - labelled Public 
2. Find the '11 - Maths and Stats' menu
3. Locate and double click on the 'SPSS (statistics)' item
4. SPSS should start on your Windows PC.

Appendix D

A quick method for Contingency Tables

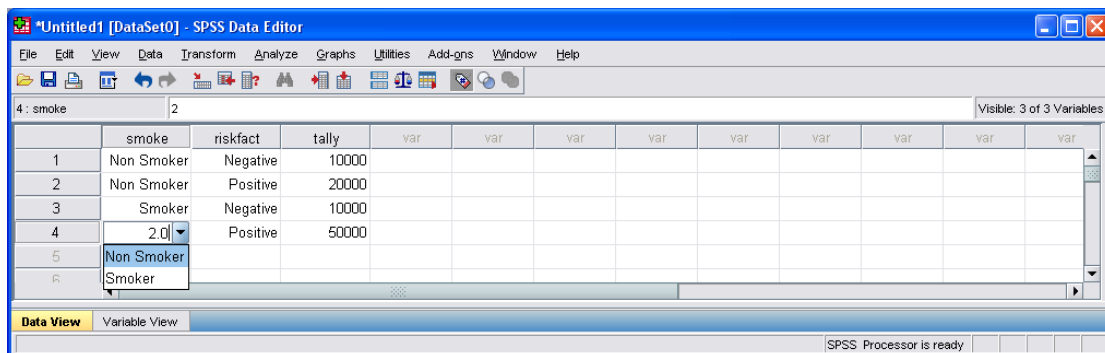
If you have a final Contingency Table(s) with large totals, the data entry effort required to perform the final Cross Tab test(s) is horrendous. However, using Weighted data, it is easy.

Example

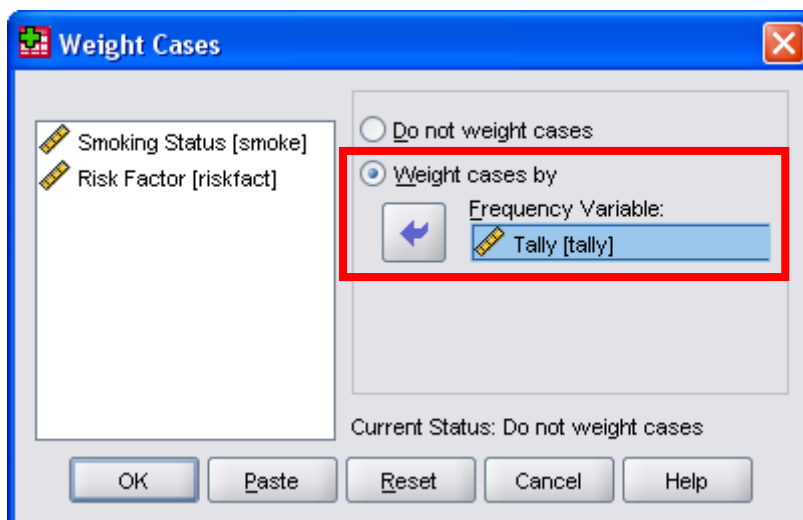
A fictitious simple 2 X 2 Contingency Table with large totals and SPSS Coding Scheme.

| Table | | Smoking Status (Smoke) | | |
|---------------------------|---------------|------------------------|-------------|-----------|
| | | Non Smoker (=1) | Smoker (=2) | Sub total |
| Risk Factor (RiskFact) | Negative (=1) | 10000 | 10000 | 20000 |
| | Positive (=2) | 20000 | 50000 | 70000 |
| Sub total | | 30000 | 60000 | 90000 |

Without Weighting, 90000 cases are required; with Weighting ONLY 4 cases are required.



Prior to analysis, you need to set up data weighting.
Within SPSS, go to the Data Menu and the Weight Cases... dialog box.



Complete the dialog box.
Choose the Weight cases radio button and Tally as the Frequency Variable, then OK.

Finally, to analyse the data, do the following sequence. Within SPSS, go to the Analyze Menu, then Descriptive Statistics and the Crosstabs... dialog box. Complete as required.