

TEESSIDE UNIVERSITY
SCHOOL OF HEALTH & SOCIAL CARE

SPSS Workbook 1 – Data Entry

Research Audit & Data

RMH 2023-n

Prepared by: Sylvia Storey
s.storey@tees.ac.uk

SPSS – data entry

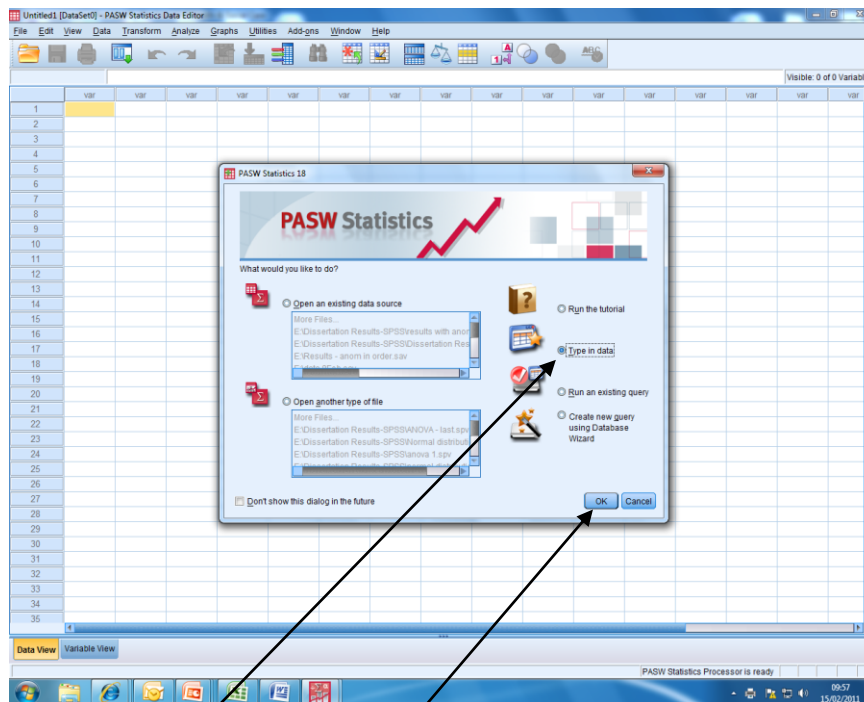
This workbook is designed to introduce you to the statistical software package SPSS (Statistical Package for Social Sciences). This package enables you to enter data, perform descriptive statistics and execute inferential statistical tests. This workbook focuses on data entry.

SPSS is constantly being updated, as are all software packages, so there are a variety of versions in use, this workbook has been written for SPSS 18. Some of you may have the same or other versions on your home or work place computers. The variations in the versions do not alter the package in a dramatic manner and you should be able to transfer your skills from one package to another without too much difficulty. Some of the images may vary depending on the version of SPSS you are using.

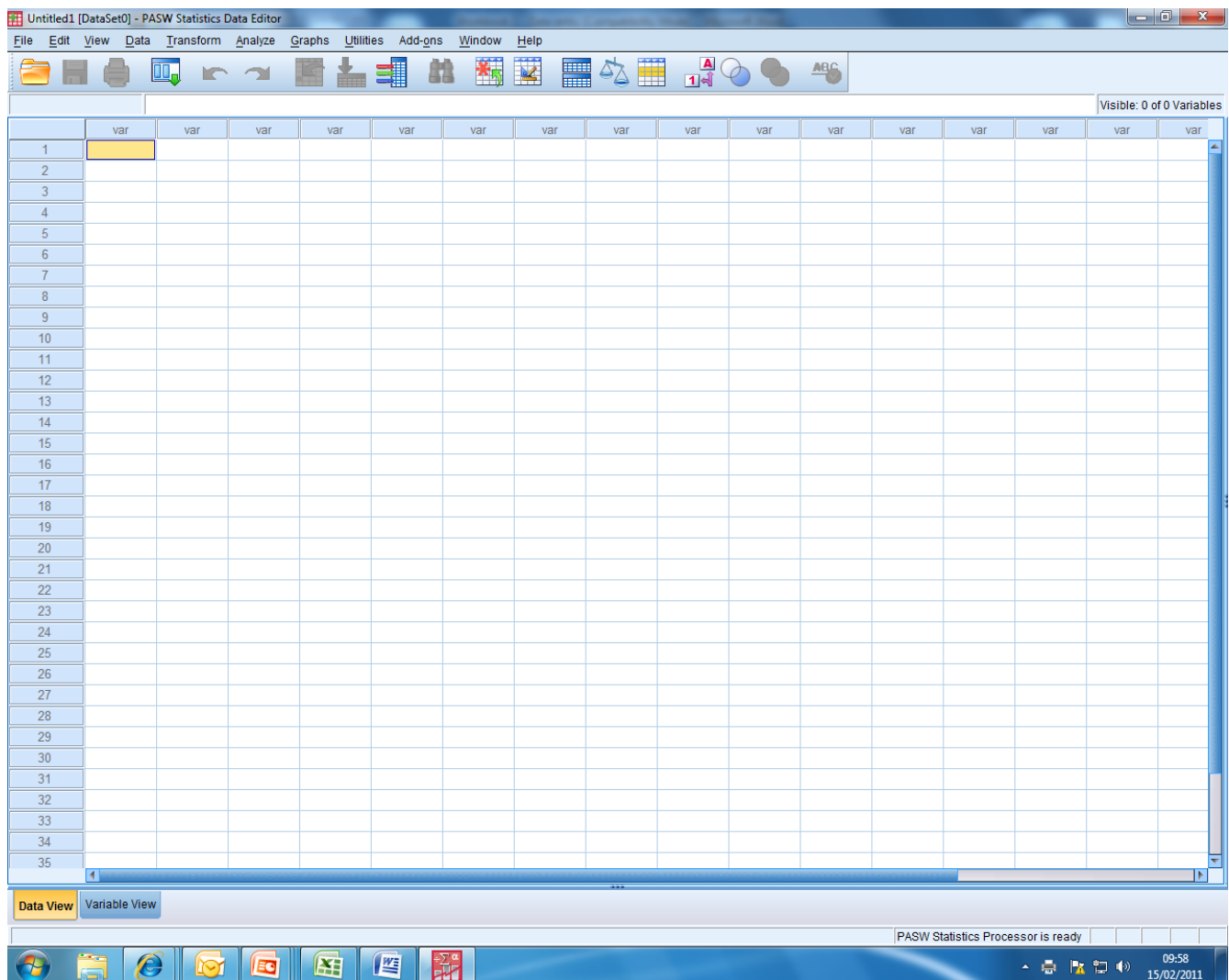
Part one: Finding your way around the package.

Before you can use the package you obviously need to turn on the computer. Once the computer is turned on you need to access the SPSS by clicking on the SPSS icon or going through the programmes menu. (Note: there is a tutorial package included in SPSS that it might be worth you working through at a later stage)

Once you have started the SPSS package you will be presented with an image similar to the one below.



At this screen click on 'Type in data' and then 'OK'. This will take you to the next screen that is the Data Editor.



This is the data editor. The data editor is basically a spread-sheet in which data is entered; this is the first stage in using SPSS to analyse data. Notice that the data editor screen has a row of drop down menus at the top of the screen. The menus, as with all Windows® applications, begin with **File** and end with **Help**. The help menu is useful and can take you through the procedures. Try clicking on the menus and familiarise yourself with the options they offer. Some are self explanatory, for example in the **File** menu you can open a file, print a file etc.

The Data Editor

Before you begin to enter data into the data editor it is useful to remember the nature of data. Researchers collect or generate data from a variety of sources and there are many types of data. It is the analysis of the data that produces information; data analysis aims to make sense of what is known as the 'raw data'. Thus, it is vital that data are collected and recorded meticulously and that every stage in the process of data entry and analysis is accurate, this avoids misrepresentation of the raw data and enhances the validity and reliability of the study. All entered data needs to be checked for mistaken entries.

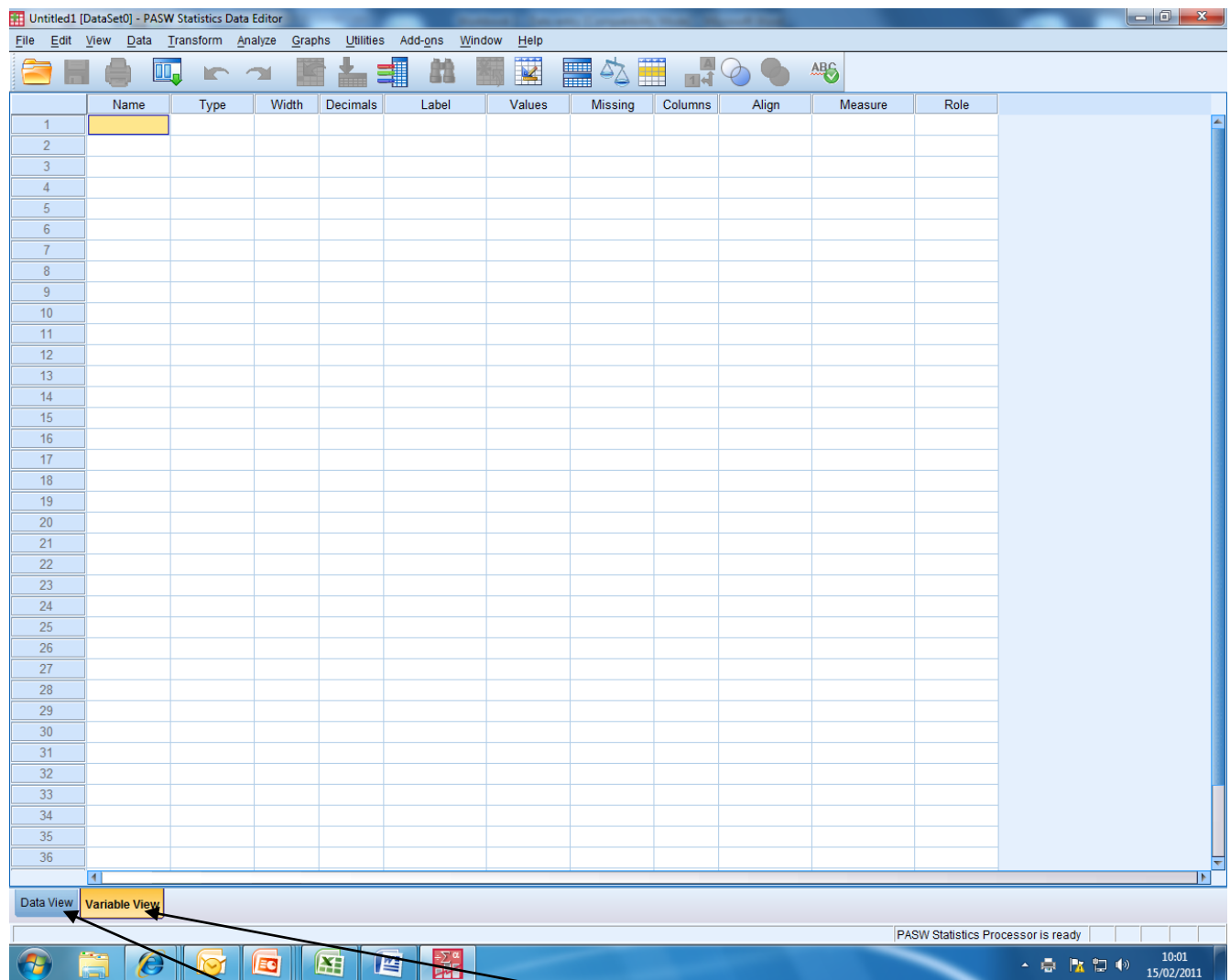
Remember the data editor is basically a spread-sheet. Down the left-hand side of the data editor you will see a column of numbers beginning with the number 1. Each number represents a row in the data editor. We use each row to represent a single case in our research. So, if we were entering data collected during a research project concerned with length of hospital stay following hip replacements each row could represent a single patient, a single ward or a single hospital. If we had collected data for a study on the number of goals scored in the premier league, each row or case could represent an individual player or team. So remember, the rows represent cases, a case is the unit of analysis that can be anything from cars to people.

Naming Variables

Now notice that across the top of the data editor, below the menus, is a set of columns. Each column represents a particular variable that we have collected data on. Each column has **var** in the top cell, **var** means variable. A

variable is the characteristic we are studying which may vary from case to case. So in the example of length of stay we could collect data on various variables that may help us have more information about the length of stay, and on what affects length of stay. We could collect data on age of the patients, reason for the hip replacement i.e. arthritis or trauma. We could also collect data on the patient's sex, weight, body - mass index, consultant, ward etc. In the example of goals scored in the premier league we may collect data on variables such as home or away matches, age of the player, position played, cost of player, type of boots worn, the type of playing surface etc. We can change the title in the top cell of each column from **var** to something that represents the variable of concern, so we can name the variable. We do this by clicking on the '**Variable View**' tab at the bottom left hand corner of the window (see above).

This will bring you to a different spread-sheet that allows you to define each of the variables that you intend to enter data for. It looks very similar to the Data View sheet but it serves a different purpose.

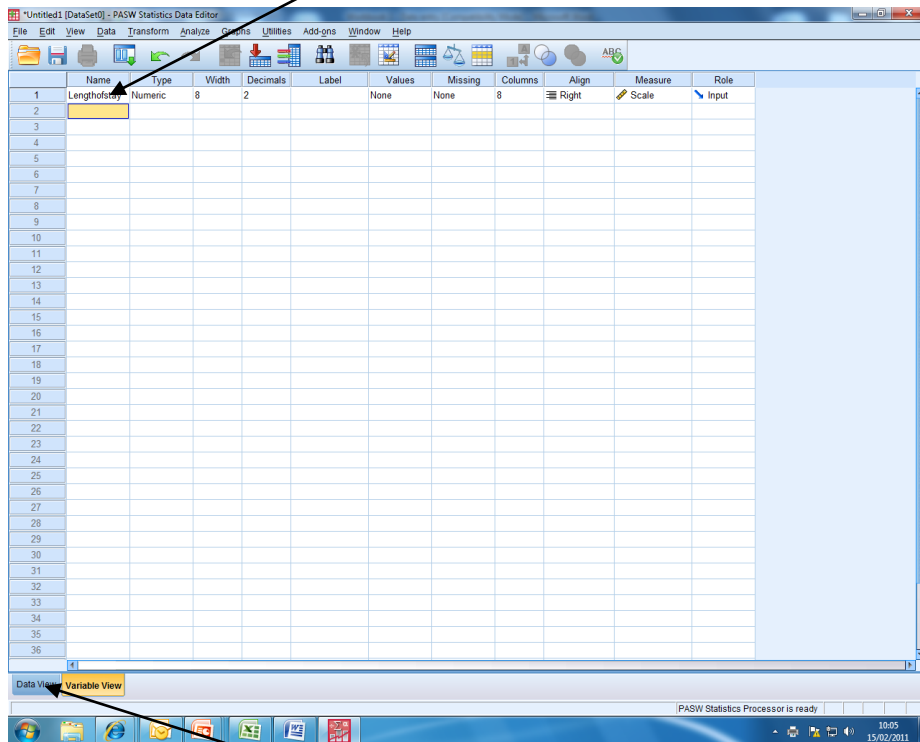


You **enter** the actual data in **Data View** and **define** each variable in **Variable View**.

Notice that in **Variable View** the top of each column no longer says '**VAR**' but has Name, Type Width etc. In this window the rows 1, 2, 3 etc refer to each variable and the columns (Name, Type, Width) allow you to specify what type of data is being entered.

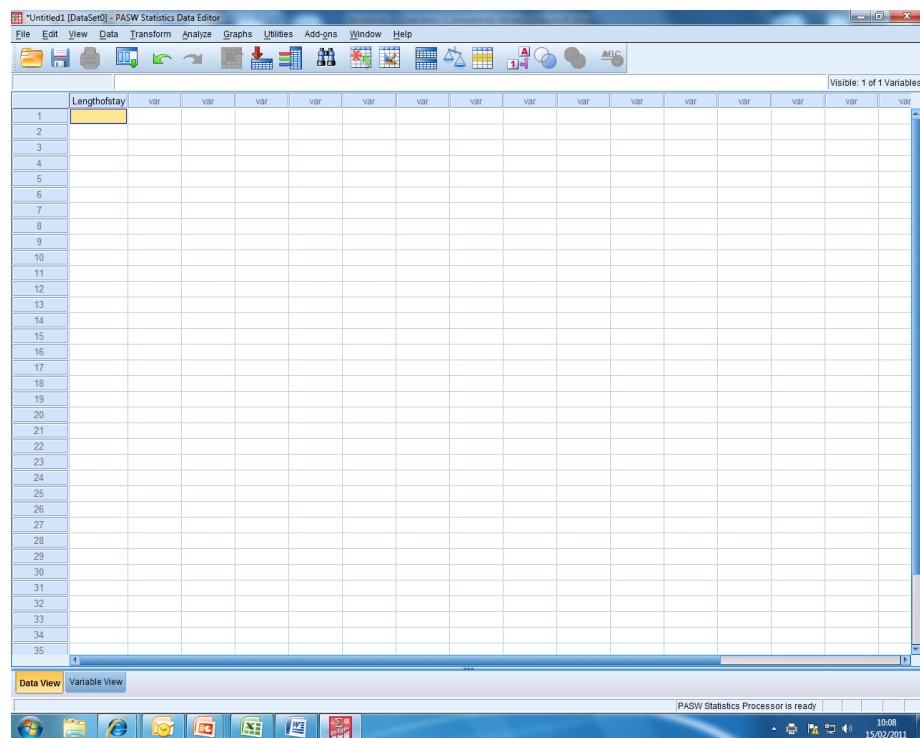
The first variable we are going to define (in row 1) is about our hip replacement patients and gives the number of days that they stayed in hospital following their surgery (i.e. their length of stay in hospital). So in row 1 under Name you can now type a variable name. Try it, type in a name that reflects days spent in hospital (**Lengthofstay**). We will ignore the other options for now. Try to use a name that reflects the variable. Blank

spaces or full stops are not allowed. Variable names should be unique, as they are the names by which the computer sorts all the information. When you have typed in the variable name you want, press the enter key on your keyboard and some other information will appear under each column. (We are going to ignore this for now). The name you typed should appear along with some other information, as below.



Entering data on one variable

Now that you have defined the first variable you can enter the data that you have on length of stay in hospital. To do this return to the **Data View** using the tab in the bottom left corner. The name of the variable you typed should now appear at the top of the first column (replacing 'var').



The raw data (see copy at back of Workbook) we have tells us that:

Patient 1 stayed in for 9 days, patient 2 for 10 days, patient 3 for 18 days... and so on for each of our 40 patients.

Length of stay data:

9,10,18,14,20,15,11,15,18,16,17,11,17,18,11,19,11,20,30,20,
10,9,23,10,15,12,16,10,11,24,35,28,41,11,10,13,13,12,20,13.

Remember each row of the spread-sheet represents a patient or case. Each column represents a variable. To enter data the process is simple.

1. The cursor needs to be in the first cell in the first row of the first column. This cell will be highlighted in bold: this denotes the active cell. You can use the mouse or the cursor direction keys (the keys with arrows on them) to move around the data editor.
2. Type in the number of days in hospital for patient 1. What you are entering, remember, is a true number.
3. Press the down arrow key on the key board, this will move the cursor to the next cell down in the column. Alternatively you can use the enter or return key. Then enter data for patient 2 and so on until you have entered the data for all 40 patients
4. If you make a mistake move back to the cell and re enter the correct number then move to the next cell. You can delete an entry by using the backspace key or the delete key.

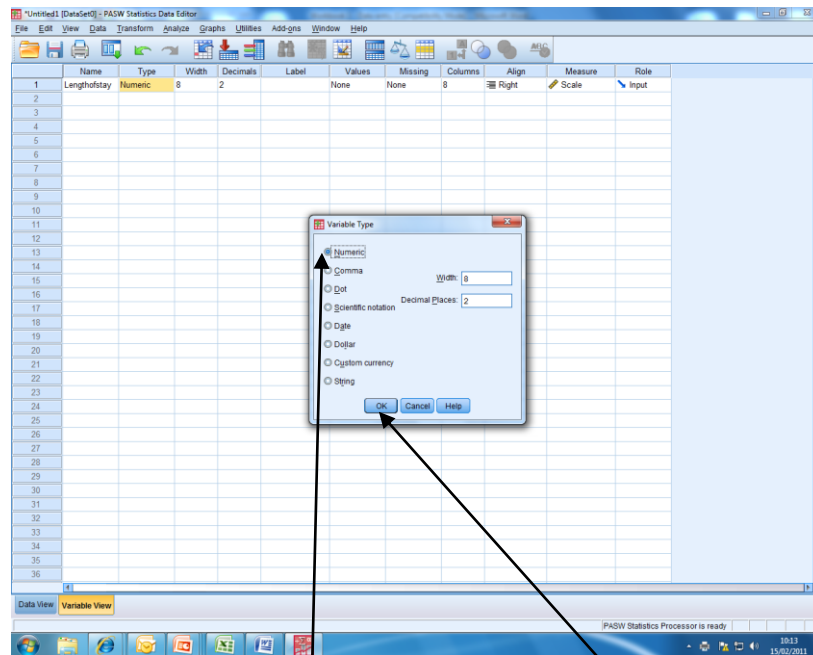
40 : Lengthofstay	13.00	Visible: 1 of 1 Variables														
Lengthofstay	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var
7	11.00															
8	15.00															
9	18.00															
10	16.00															
11	17.00															
12	11.00															
13	17.00															
14	18.00															
15	11.00															
16	19.00															
17	11.00															
18	20.00															
19	30.00															
20	20.00															
21	10.00															
22	9.00															
23	23.00															
24	10.00															
25	15.00															
26	12.00															
27	16.00															
28	10.00															
29	11.00															
30	24.00															
31	35.00															
32	28.00															
33	41.00															
34	11.00															
35	10.00															
36	13.00															
37	13.00															
38	12.00															
39	20.00															
40	13.00															
41																

After entering all of the data you should have reached row 40 (see above). If you are not at 40 then something has been entered wrongly. You might have mistyped a number, entered something twice, or not entered data for one patient. Therefore make sure all the data has been entered correctly. This can be time consuming, especially for large data sets, but if you enter data incorrectly your analysis will also be affected.

Defining Variables

It is necessary to enter other information about the data that you are intending to analyse in order to describe the variables more accurately. SPSS allows you to use variable labels; these are descriptions of the variables. This allows you to describe in detail what a variable name stands for. This is helpful, especially if you have given numeric codes to non-numeric data. This is easy to do:

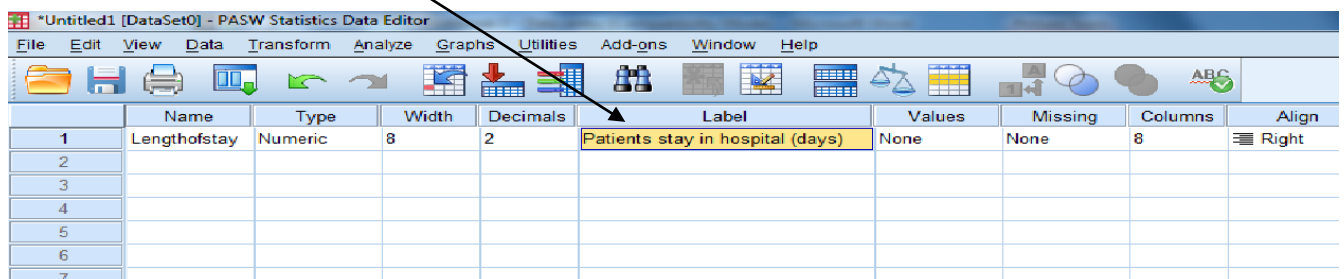
1. Use the tab button at the bottom to return to Variable View.
2. You need to tell SPSS the type of data you are entering. In most cases this will be numeric. To do this go to the column headed 'Type' for row 1 (length of stay). If you click on the small grey box (containing 3 small dots) this brings up another menu that allows you to specify the type of data being entered. This is what you should see:



As you can see you are presented with several different options. String refers to a “string of letters” so data can be entered as a name. We will leave the other options for now. The options available vary depending on the type of data you want to enter. Make sure ‘Numeric’ is chosen and then click ‘OK’ to return to the Variable View.

Notice when the numeric option is highlighted, this allows you to enter numeric data that can consist of 8 digits of which 2 can be decimal places. These can be changed, if for example you are entering data that has no decimals and only consists of three digits. These options also get added to the appropriate column in data view.

3. In the column headed Label you can type in your description of the variable. This can give a fuller label to the variable that helps you understand what it is measuring. You do not have to complete this section.



Entering Coded Data

Using our example of length of stay following hip replacement we may wish to enter data collected on each respondent's consultant. You would not enter the consultant's name into SPSS; instead you would code the data so that you entered a number instead of their name. If there were three consultants, Smith, Jones and Wilder we could code them as 1 (=Smith), 2 (=Jones), 3 (=Wilder).

If we had data on how the patients rated their length of stay we could code it:

1= shorter than I expected, 2=what I expected, 3=longer than I expected. Similarly you could code sex 1=female, 2=male. These value labels can be entered into SPSS.

Now we will enter data for our second variable (patient's consultant). We need to specify details of this variable in Variable View. This time we will be using 1, 2 and 3 to represent Smith, Jones and Wilder (the consultants) and will need to specify these as value labels.

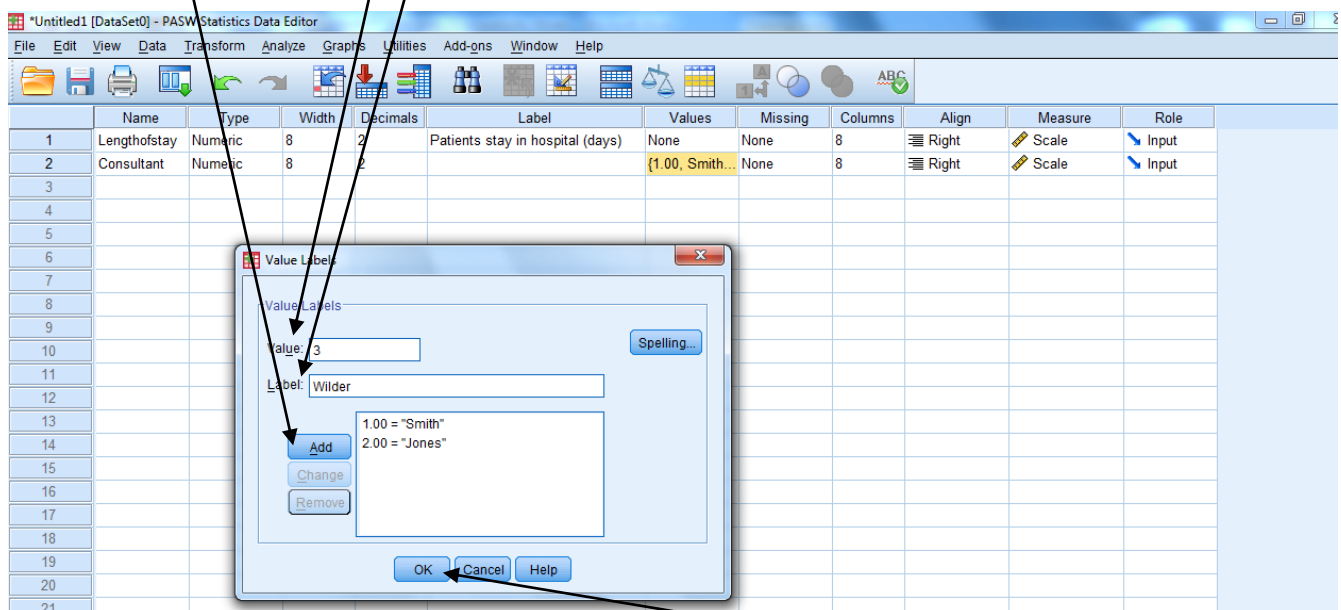
1. In order to enter coded data you first of all need to give the second variable a name. This time the name should mean something that reminds you that it is about the consultant the patient saw. The new name should be entered in the second row, underneath the variable definition for length of stay. (See the section above on Naming Variables)

Because we are going to enter a number instead of the consultant's name we need to tell SPSS what each of the numbers mean. To do this we use Value Labels.

Value Labels

2. In the Variable View spreadsheet you need to define value labels. Make sure you stay on Row 2 (our variable for consultant) and go to the cell under the column headed 'Values', and click on the little grey box (with 3 dots) this will bring up the window below:

3. Place the cursor in the box "**Value**" and type in the number one.
4. Place the cursor in the box "**Label**" and type in Smith.
5. Click on "**Add**" you should see that 1 equates with **Smith** in the lower box. Repeat this for the other two values. (2=Jones, 3=Wilder)



6. When you have entered the labels for all three consultants click the 'OK' button to return to Variable View.

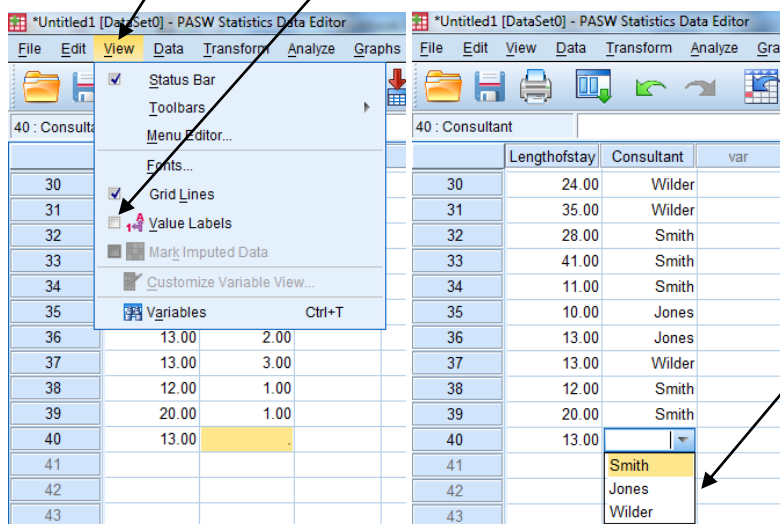
You can now go back to Data View to enter data on the second variable “Consultant”. Again there are 40 patients so you need to enter the correct code (1, 2 or 3) depending on the consultant each patient saw.

Here is the raw data:

1.Smith	11. Wilder	21. Jones	31. Wilder
2.Smith	12. Wilder	22. Smith	32. Smith
3.Jones	13. Smith	23. Smith	33. Smith
4.Smith	14. Smith	24. Smith	34. Smith
5.Jones	15. Jones	25. Wilder	35. Jones
6.Jones	16. Wilder	26. Jones	36. Jones
7.Wilder	17. Smith	27. Wilder	37. Wilder
8.Wilder	18. Jones	28. Jones	38. Smith
9.Jones	19. Wilder	29. Jones	39. Smith
10.Jones	20. Jones	30. Wilder	40. Jones

Note that we assigned numbers to code the data but you must take care **not** to treat the data on consultant as if they were a true number. The number merely represents a category “consultant”. It has no numeric meaning. So what type of data is it? (Nominal Ordinal, interval or Ratio)

By selecting **View Value Labels** it is possible to select the consultants name from a drop down list instead of typing in their nominal values.



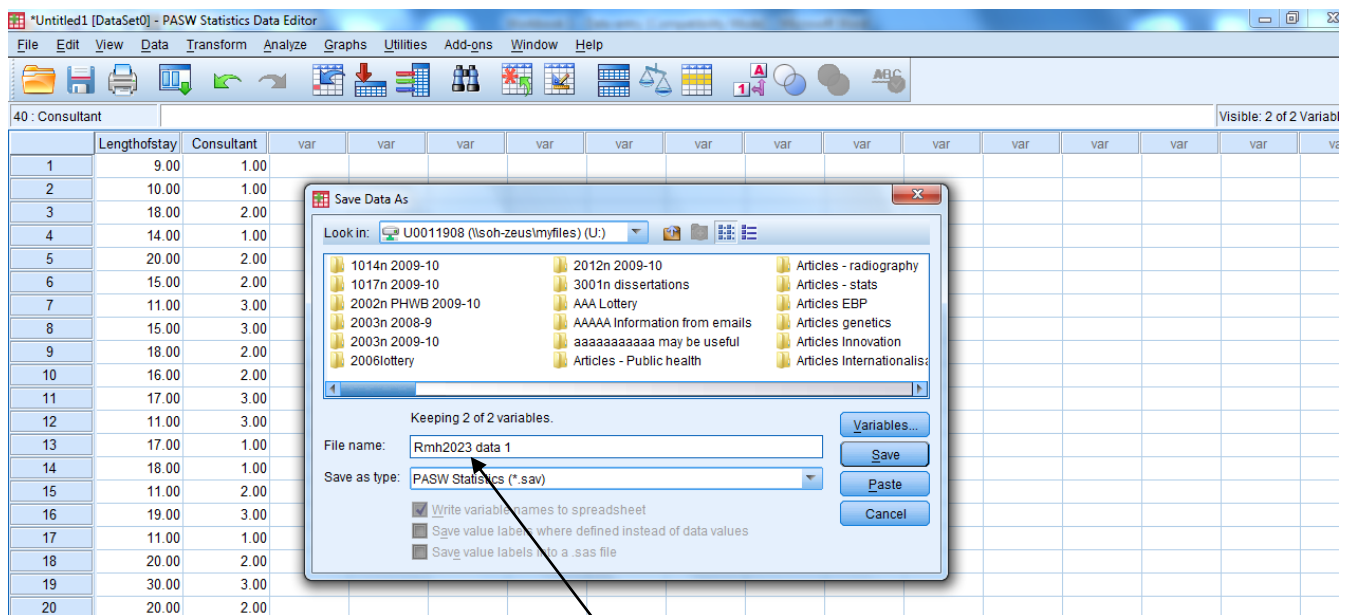
Of course the data on length of stay were true numbers. This is a higher level of measurement, that is, ratio data. We know that if a patient ‘A’ is in hospital for 5 nights and patient ‘B’ is in for 10 nights that patient B’s stay was twice as long as patient A. Also, remember that you can make no other assumptions about the reason for the length of stay unless you have other data to test your assumptions out. It could be that patient A had a very speedy recovery or it could mean reason for discharge was “death”. Thus, to collect a comprehensive data set on length of stay we would need to collect data on many variables, reason for discharge being one.

Saving Data

Having entered some data it may be an idea to make sure how to save it.

1. Click on the **File** menu.
2. **Click on** Save As and save file to an appropriate location (recommend your homespace (u: drive)).

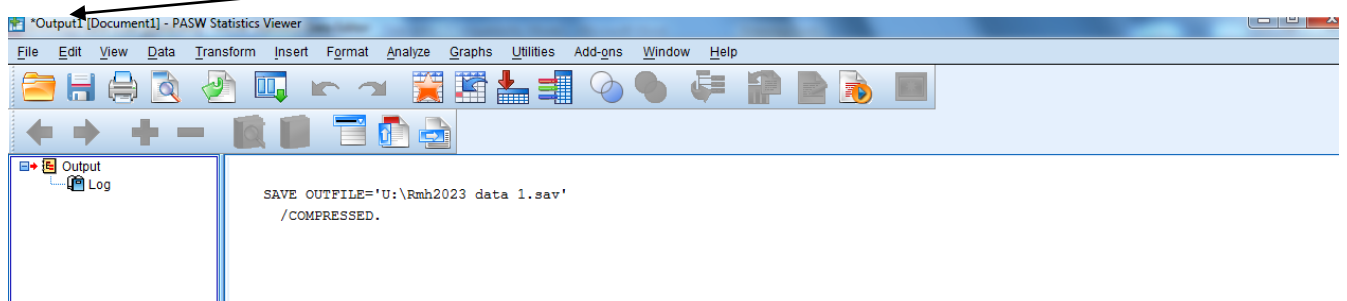
This image on the screen will vary slightly from the one below depending on your version of SPSS:



Give the file a name by placing the cursor in the **File name** box. The file will automatically be given a type **.sav** after the name. The file extension **.sav** is used so that you can recognise files that have SPSS data in them from other files. You will come across the other file extensions when you perform data analysis. Do not click on **save** until you have selected the place that you wish the data to be saved in. To do this click on the arrow next to the dialogue box **Look in:** and select where you would like to save it to.

It is sensible to regularly save data entered so that you do not lose anything.

When you save your file SPSS automatically opens an Output window that details where you have just saved the data file (you do not need to ~~save this file~~)



Now that you understand some of the basics of setting up variables and entering data you need to try doing it on something more real

When you have collected your own data you are most likely to enter the data directly from some questionnaires. Rather than give you a whole lot of questionnaires to look through we have provided a spreadsheet containing all of the raw data (see end of the workbook). You should now focus on defining each of the variables, give them a name, define them and define the value labels if you are going to enter coded data for that variable.

***You will notice that the dataset only contains information for 20 patients but you have begun to enter data for 40 patients. You will need to remove cases 21-40 before you begin your analysis sections.

SPSS DATA SET FOR PATIENTS ADMITTED TO HOSPITAL FOR HIP REPLACEMENT SURGERY.

Part No	Length of Stay	Consultant	Age	Gender	Diagnosis	Weight OA	Smoking	Waterlow	Exercise	Weight DG	Blood Loss	Income
1	9	1	60	2	2	68	2	5	2	65	250	1
2	10	1	52	2	2	56	2	4	2	53	270	2
3	18	2	65	1	1	70	2	8	1	68	300	3
4	14	1	58	1	1	76	2	7	2	73	220	2
5	20	2	72	2	2	53	1	15	1	49	320	1
6	15	2	78	2	1	98	1	10	1	94	210	2
7	18	3	70	2	1	54	2	4	2	53	220	2
8	19	3	65	1	1	65	2	7	1	64	290	2
9	18	2	81	1	1	70	1	16	1	66	180	2
10	16	2	70	2	1	71	1	10	2	70	350	1
11	17	3	77	1	2	72	1	15	1	68	220	1
12	15	3	76	1	1	64	2	12	2	62	270	1
13	17	1	82	2	2	68	2	17	1	65	400	3
14	18	1	68	2	1	90	1	16	1	87	320	2
15	11	2	72	2	1	76	2	8	2	74	260	3
16	19	3	79	1	1	58	1	15	1	55	290	3
17	11	1	84	1	2	58	2	8	2	56	240	1
18	20	2	69	1	1	83	1	16	1	80	250	1
19	30	3	70	2	1	85	1	18	2	79	300	2
20	20	2	80	2	1	80	2	17	2	78	190	2

Key: Variables (and levels, where applicable)

Length of Stay: Length of Stay in Days	Diagnosis: 1 = Chronic Illness, 2 = Trauma	Exercise Regime: 1 = Traditional, 2 = New
Consultant: 1 = Smith, 2 = Jones, 3 = Wilder	Weight OA: Weight in Kg (on admission)	Weight DG: Weight in Kg (on discharge)
Age: Age in Years.	Smoking: 1 = Smoker, 2 = Non-Smoker	Bloss: Blood loss in Surgery in millilitres
Gender: 1 = Female, 2 = Male.	Waterlow: Waterlow Score	Income: 1 = below national average, 2 = national average, 3 = above national average

Levels of Measurement for the Variables in the SPSS Dataset on Hip Replacement Surgery

Variable (and levels where applicable)	Level of Measurement
Length of Stay: Length of Stay in Days	Ratio
Consultant: 1 = Smith, 2 = Jones, 3 = Wilder	Nominal
Age: Age in Years	Ratio
Gender: 1 = Female, 2 = Male.	Nominal
Diagnosis: 1 = Chronic Illness, 2 = Trauma	Nominal (at a push ordinal if conceptualised in terms of severity)
Weight OA: Weight in Kg (on admission)	Ratio
Smoking: 1 = Smoker, 2 = Non-Smoker	Nominal
Waterlow: Waterlow Score	Interval/ratio (but at least interval)
Exercise Regime: 1 = Traditional, 2 = New	Nominal
Weight DG: Weight in Kg (on discharge)	Ratio
Bloss: Blood loss in Surgery in millilitres	Ratio
Income: 1 = below national average, 2 = national average, 3 = above national average	Ordinal