Laboratory Tutorial 1-1: Basic data processing tasks in SPSS

In this laboratory tutorial you will:

1. Begin to learn how to use a data analysis application called SPSS
2. Load an existing data file into SPSS
3. Perform some basic data processing tasks using SPSS

This is a mandatory tutorial. In order to pass the coursework, you must achieve a score of 50% or higher on the associated Blackboard quiz (Lab Quiz 1-1).

Preamble to the Laboratory Tutorial series

Laboratory tutorials are self-contained sets of exercises and questions that will help you to progress towards achieving the learning outcomes for this module. You can measure your progress at any time by taking the associated test on Blackboard. Many of these tutorials are mandatory and must be completed to a satisfactory standard, by taking the quiz on Blackboard, before you can pass the coursework assessment of this module. You can take a quiz maximum 3 times and in order to get a total score recorded in the grade centre, you must complete the whole quiz in a single session. For this reason, it is best to write your answers down as you go through the tutorial.

Each tutorial document has a descriptive title and numerical code. The first digit in the code indicates which learning outcome a tutorial addresses and the second indicates the order in which you should attempt it. For instance this tutorial is codes as 1-1, which means it addresses learning outcome one and is the first in the series. Once you have completed this tutorial you should move on to 1-2, then 1-3 and so on. From time to time it will be necessary to switch to tutorials relating to other learning outcomes. For instance, later in Term 1 we will cover structured query language (SQL), which relates to learning outcome two. The first tutorial in this series will be coded 2-1.

Tutorials are available for learning outcomes one, two and three (but not four).

Exercise 1: Learning the basics of SPSS

In this tutorial you will be using data generated from a survey study conducted to explore the prevalence and impact of sleep problems on aspects of people's lives (see Pallant, 2007). This is real data, although individual cases (rows) are distinguished by an identification number (ID) rather than their name, address or any other data that might reveal their identity. Making data 'anonymous' in this way is common practice and satisfies the ethical requirement to preserve participant confidentiality[1]. This serves to protect their privacy whilst maximising the likelihood that the responses they provide are truthful.

[1] We will cover this issue in the lectures on Ethics and also Research Methods (see schedule in Study Guide)

Find the data file

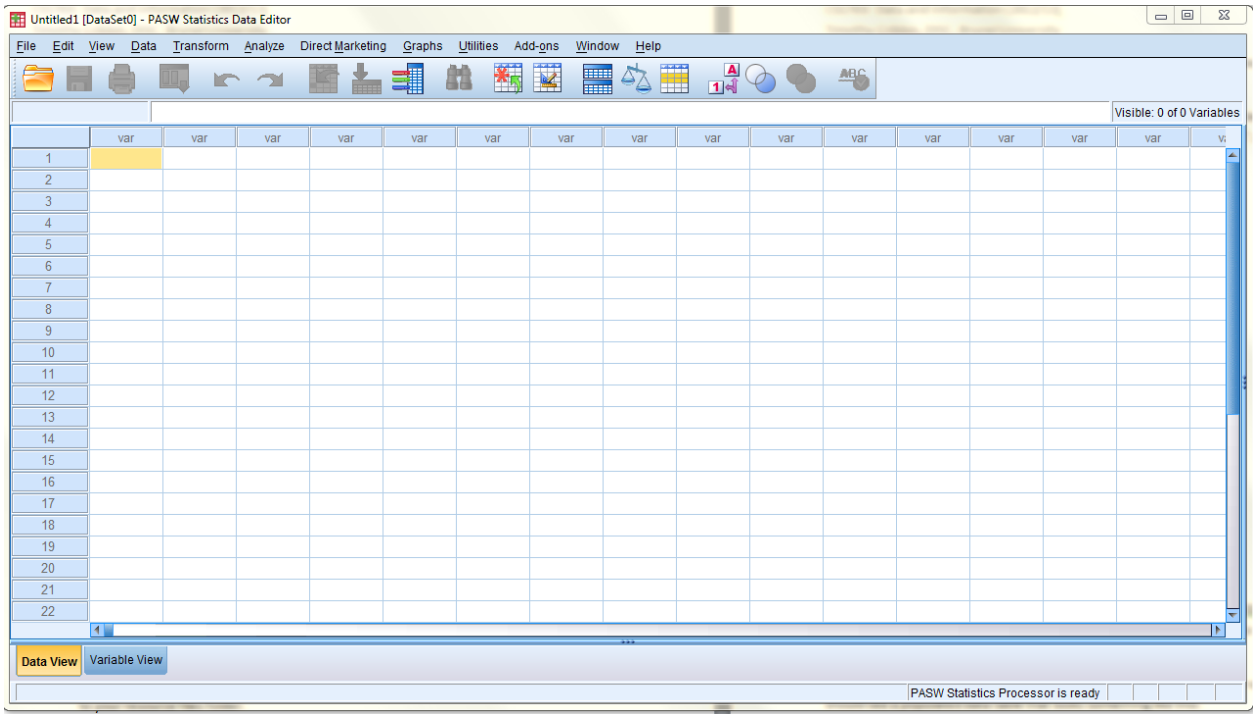
Before you begin the following exercises you need to:

- **Create a folder** in a location that you have read and write access to. This can be your network file area (H: drive), a pen-drive or some other device. Name the folder "CS1703 Lab Files" or something similar.
- **Download the data** from the CS1703 section on Blackboard. Go to the folder "Laboratory → Resources" folder and select the file named "Lab1-1_sleep3ED.sav" (Pallant, 2007). Save it to your resource files folder.
- **Download the questionnaire and codebook** that provides a description of the data source and how the collected data has been coded in the SPSS data file. This is in the same location as the data and is called "sleep_qaire_codebook.pdf" (Pallant, 2007).

Start SPSS

SPSS can be found on the University lab PCs by typing "SPSS" into the search box on the Windows Start menu. You may get more than one link returned, choose the latest version. If you can't locate the link, ask a tutor for help. You can also obtain a copy of SPSS for your own PC (or Mac) by logging into the Connect Portal (https://connect.brunel.ac.uk). Select 'My Apps' and you should see links to the SPSS software download and current license key.

On loading, you should initially see the main Data view as shown below. You may see a menu asking you "What would you like to do?". Select "Type in data", then click the OK button. You should see a window like that shown below – essentially an empty data table that looks a bit like a spreadsheet.



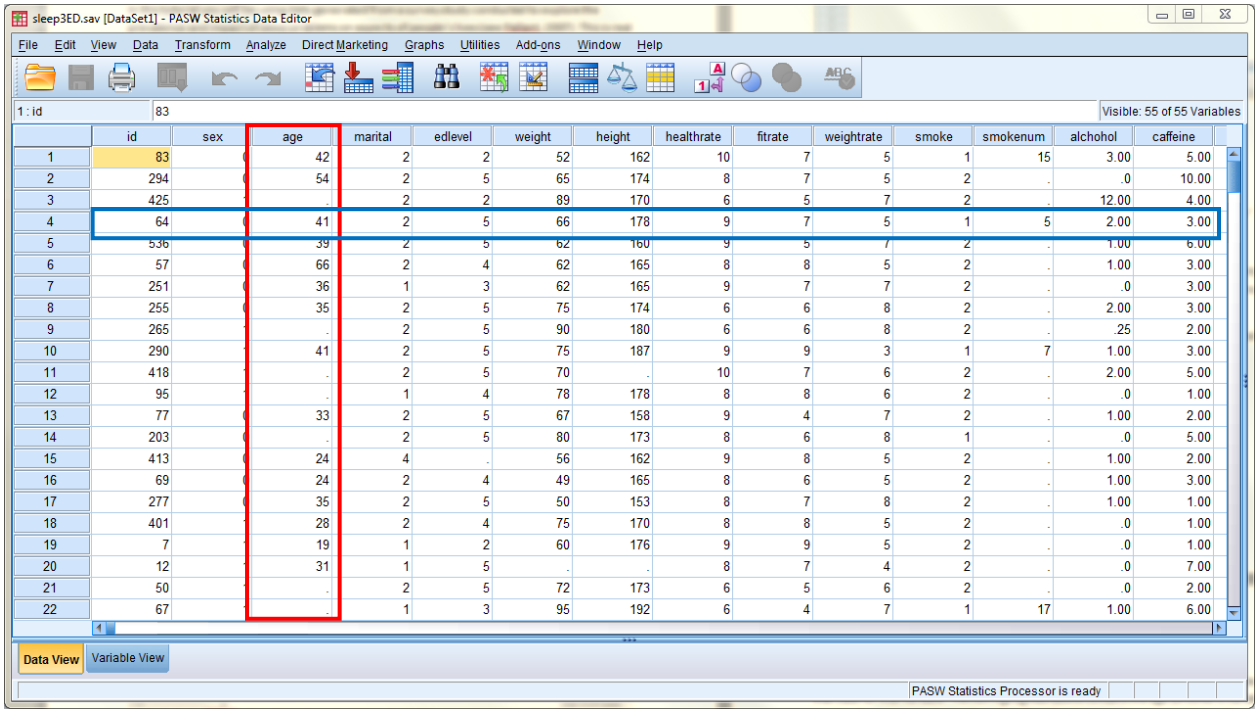
Load the data file

To load the data file you just downloaded:

- Select the menu path: “File → Open → Data” from the SPSS
- Navigate to the lab folder you created earlier and select “Lab1-1_sleep3ED.sav”
- Click “Open”

The file should have loaded into the data view (entitled: “Lab1-1_sleep3ED.sav[DataSet1]”). SPSS will also open a second window called Output1. The purpose of this window is to display the results of any data operations that you perform. We will come back to this later.

For now, bring the data view back into focus by selecting the data editor window (DataSet1). You should see a populated data table that looks something like this:



	id	sex	age	marital	edlevel	weight	height	healthrate	fitrate	weightrate	smoke	smokenum	alchohol	caffeine
1	83		42	2	2	52	162	10	7	5	1	15	3.00	5.00
2	294		54	2	5	65	174	8	7	5	2	.	.0	10.00
3	425		.	2	2	89	170	6	5	7	2	.	12.00	4.00
4	64		41	2	5	66	178	9	7	5	1	5	2.00	3.00
5	536		39	2	5	62	160	9	5	7	2	.	1.00	6.00
6	57		66	2	4	62	165	8	8	5	2	.	1.00	3.00
7	251		36	1	3	62	165	9	7	7	2	.	.0	3.00
8	255		35	2	5	75	174	6	6	8	2	.	2.00	3.00
9	265		.	2	5	90	180	6	6	8	2	.	.25	2.00
10	290		41	2	5	75	187	9	9	3	1	7	1.00	3.00
11	418		.	2	5	70	.	10	7	6	2	.	2.00	5.00
12	95		.	1	4	78	178	8	8	6	2	.	.0	1.00
13	77		33	2	5	67	158	9	4	7	2	.	1.00	2.00
14	203		.	2	5	80	173	8	6	8	1	.	.0	5.00
15	413		24	4	.	56	162	9	8	5	2	.	1.00	2.00
16	69		24	2	4	49	165	8	6	5	2	.	1.00	3.00
17	277		35	2	5	50	153	8	7	8	2	.	1.00	1.00
18	401		28	2	4	75	170	8	8	5	2	.	.0	1.00
19	7		19	1	2	60	176	9	9	5	2	.	.0	1.00
20	12		31	1	5	.	.	8	7	4	2	.	.0	7.00
21	50		.	2	5	72	173	6	5	6	2	.	.0	2.00
22	67		.	1	3	95	192	6	4	7	1	17	1.00	6.00

Anatomy of the data editor

The data view may seem familiar to you. If you have used a spreadsheet before (e.g. Excel) then you will already understand many principles that are relevant to viewing and editing data in SPSS. However, SPSS is not the same as spreadsheet program. For instance you cannot insert formulas directly into a cell. However, SPSS has many powerful functions for data processing and analysis that are not possible (or at least so easy) to perform in a spreadsheet program.

Starting with the basics, the area highlighted in blue, on the screenshot above, is a row of data. In line with convention for data tables[2], a row represents a distinct record. However, SPSS uses the term "**case**" instead of record to refer to a row. A case might refer to a person, a product or some other subject (sample member) for which data has been collected. In the screenshot, case number 4 is highlighted.

The area highlighted in red is a column of data. In line with data table convention, a column represents a distinct field or attribute. However, SPSS refers to a column as a "**variable**" - i.e. a value on which cases will vary. The column highlighted here relates to the variable called "age". Each case will generally possess a value for some or all of the defined variables in the data table.

The intersection between a particular column and row is called a **cell** and displays the value of that case for that variable. The cell highlighted above contains the age (a numeric variable) of case number four: 41 years. Should you wish, you could change the value of a cell in the same way you would do with any other spreadsheet by clicking on the cell and typing the new value. However, you would not normally change a value in a set of collected data such as this unless you had a valid and ethical reason for doing so. One reason might be where the value is obviously an error - e.g. an age of 243 years would probably be a typo – and you know the correct value to replace it with.

You might notice that there are some variables that look like they should be represented by words or phrases, yet the column contains numbers all the same. Sex is one example and Marital (Status) is another. The reason for this is that SPSS is primarily a quantitative analysis tool that works with numbers. Categorical variables like Sex and Marital Status should have their values translated or **coded** into a numeric scheme. For instance we could use a 1 to refer to females and a 2 to refer to males. It doesn't matter what numbers we assign to each category, so long as we are consistent (i.e. 1=males, 2=females would work just as well. SPSS has a feature called **value labels** that makes it easier to understand such codes. You can toggle between the labels and actual category numbers by clicking on the "Value labels" icon  or checking "Value Labels" in the "View" menu.

This is a good point to open and read the document containing the questionnaire and codebook ("sleep_qaire_codebook.pdf"). If you scroll down to the table at the end of this document, you should be able to see that the code for the gender variable is: 0=female and 1=male. Take a look through the whole table, noting which variables are categorical (represented by types or classes) and which are quantitative (represented by numbers). You will also find it helpful to cross-reference each variable in the codebook with the questionnaire item that is derived from. If you're not sure about a particular variable, ask a GTA or lecturer (if present).

Any variable for which a case has no value is represented as a missing (null) value, shown as a dot within the cell. It is important to remember that a zero should not be used to record a missing numerical value – in SPSS zero means a measure was made but it just happened to be zero, whilst a missing value means that no data was available for that variable. Missing values are excluded from any analyses, whilst zeros will be included.

[2] See the first lecture titled "Data and Information"

Variable metadata

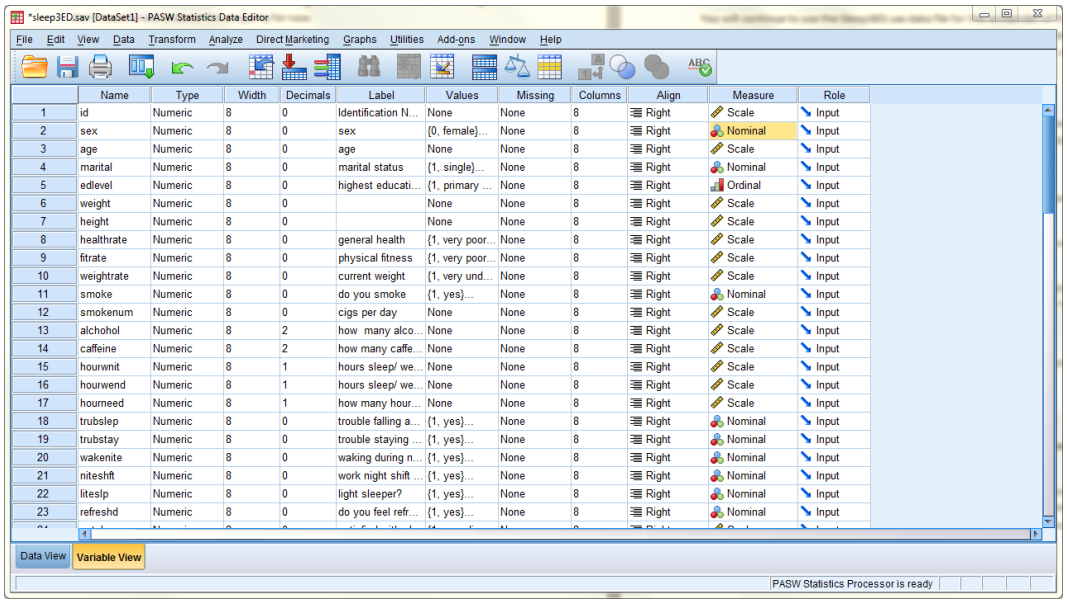
You may have noticed two tabs at the bottom left of the data table. So far, you have only been looking at the "Data View". The second tab, "Variable View", provides a different view of the data: its metadata.

Select the "Variable View" tab. You'll see a very different view of the data. Now variables are represented by rows instead of cases. Variables are ordered from top to bottom as they appear from left to right in the "Data View". Each column then describes a common attribute of each variable. Each variable has the follow attributes or metadata:

- Name – this is the unique identifier of the variable which appears on the top row of the "Data View". Names should be as short as possible and not include spaces or symbols.
- Type – SPSS supports a variety of data "types" but in most cases you will be dealing with "Numeric" data. "String" data is free text (e.g. a word or phrase). We may make use of this type of variable in later exercises.
- Width – this specifies the number of characters to display. The default width is 8 characters. This is normally sufficient for numeric data, but larger numbers or string variables may require a higher value.
- Decimals – this specifies the number of decimal places to display for a continuous numeric variable. By default this is zero, but if you are working with continuous data, you may need to set this to non-zero number.
- Label – this allows you to specify a longer description, with spaces, of your variable name. SPSS will display the label in any processing output, so a clear label makes output easier to read.
- Values – this property allows you to associate labels with particular category values (codes). As discussed above, this is most important if the variable is a categorical variable that is used to separate cases into pre-defined groups or types (e.g. male/female, young/old).
- Missing – sometimes there might be a variety of known reasons why a value is missing. This property allows the user to specify one or a range of values to be treated as missing. We don't need to worry about this property for now.
- Columns – this specifies the width of the column display. This is set to 8 characters by default.
- Align – this is "Right" by default". We don't need to worry about this property.
- Measure – you'll learn later that this property is important when it comes to selecting variables for statistical tests. There are three different levels of measure: **Scale, Ordinal and Nominal**. In short, quantitative data are considered as scale variables, ranked data (e.g. position in a league table) are considered as ordinal variables, whilst categorical data are

treated as Nominal variables. We will discuss the differences in more detail later in the lecture "Preparing Data for Analysis".

- Role – this relates to the likely role that the variable will play in analyses. For instance, an input variable means that it will be used as an independent variable, whilst a target variable will be used as a dependent variable. You'll come to understand what this means later in the term. You can safely ignore this for now



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	8	0	Identification N...	None	None	8	Right	Scale	Input
2	sex	Numeric	8	0	sex	{0, female}...	None	8	Right	Nominal	Input
3	age	Numeric	8	0	age	None	None	8	Right	Scale	Input
4	marital	Numeric	8	0	marital status	{1, single}...	None	8	Right	Nominal	Input
5	edlevel	Numeric	8	0	highest educati...	{1, primary ...	None	8	Right	Ordinal	Input
6	weight	Numeric	8	0		None	None	8	Right	Scale	Input
7	height	Numeric	8	0		None	None	8	Right	Scale	Input
8	healthrate	Numeric	8	0	general health	{1, very poor...	None	8	Right	Scale	Input
9	fitrate	Numeric	8	0	physical fitness	{1, very poor...	None	8	Right	Scale	Input
10	weightrate	Numeric	8	0	current weight	{1, very und...	None	8	Right	Scale	Input
11	smoke	Numeric	8	0	do you smoke	{1, yes}...	None	8	Right	Nominal	Input
12	smokenum	Numeric	8	0	cigs per day	None	None	8	Right	Scale	Input
13	alchohol	Numeric	8	2	how many alco...	None	None	8	Right	Scale	Input
14	caffeine	Numeric	8	2	how many caffe...	None	None	8	Right	Scale	Input
15	hourwnit	Numeric	8	1	hours sleep/ we...	None	None	8	Right	Scale	Input
16	hourwend	Numeric	8	1	hours sleep/ we...	None	None	8	Right	Scale	Input
17	hourneed	Numeric	8	1	how many hour...	None	None	8	Right	Scale	Input
18	trubslep	Numeric	8	0	trouble falling a...	{1, yes}...	None	8	Right	Nominal	Input
19	trubstay	Numeric	8	0	trouble staying ...	{1, yes}...	None	8	Right	Nominal	Input
20	wakenite	Numeric	8	0	waking during n...	{1, yes}...	None	8	Right	Nominal	Input
21	niteshft	Numeric	8	0	work night shift ...	{1, yes}...	None	8	Right	Nominal	Input
22	liteslp	Numeric	8	0	light sleeper?	{1, yes}...	None	8	Right	Nominal	Input
23	refreshd	Numeric	8	0	do you feel refr...	{1, yes}...	None	8	Right	Nominal	Input

Now go to **Lab Quiz 1-1** in Blackboard (Laboratory → Quizzes) and enter your answers to the following questions:

Q1: Using the questionnaire-codebook document, determine which SPSS variable name represents the data collected from question: "Would you describe yourself as a 'light sleeper' (easily awoken)"

Q2: What is the size of the sample recorded in this dataset?

Q3: How tall (in cm) is the person coded as id=417?

Q4: What is the level of measurement of the variable cigsgp3?

Q5: How many distinct values are associated with the variable qualslp?

Exercise 2: Basic data processing in SPSS

Some basic data processing tasks were introduced in the first lecture (Data and Information). In this section, you will learn how to perform the following tasks in SPSS:

- Sorting
- Classification
- Summarisation
- Selection

You will continue to use the Sleep3ED.sav data file for the remainder of this tutorial. You will learn how to perform the techniques by answering questions relating to this data. For a description of the original questionnaire and resulting data set (variable names and codes) read the PDF document that was referred to earlier.

Sorting Cases

Questions of the form what is the *nth* highest/lowest case according some variable(s) are easily answered after sorting the rows of the data table. Let's work with the following question:

Question: How old were the oldest participants in the study?

- Go to the menu and follow the path Data → Sort Cases. A dialogue box (new window) should appear.
- In the "Sort Cases" dialogue window, select the variable "age" from the list on the left, then click on the arrow button to move this into the "Sort by" list
- As we want to know the oldest participant in the study, we need the data table to be sorted by age in descending order. Select the appropriate option in the "Sort Order" frame.
- Click OK.

Q6. What age is the oldest participant in the study?

Q7. What is the age of the 10th oldest participant in the study?

Q8. What is the ID number of the youngest (where age is known), married participant in the study?

Classification

Task: Classify participants into age ranges

This is a task that you need to complete as preparation for the **Summarising Groups** questions in the next section. In order to look at differences between younger and older adults you need to define age categories or groups and assign each case to the correct group. You'll notice that there is an age

group variable specified in the codebook already (**agegp3**). However, this has been deleted from the dataset you are working with, so you will need to recreate it.

You'll see that the groups are defined as follows:

1. <=37 yrs
2. 38 to 50 yrs
3. 51+ yrs

Before you create the age group variable, try to work out the correct answers to the following questions just by studying the sorted data table:

Q9. How many cases belong to the <=37 age group?

Q10. How many cases belong to the 38 to 50 age group?

Q11. How many cases have no recorded age data?

How did you do it? You might have used arithmetic or just simple row counting. Your answers may be correct, but you can probably see that either method is prone to error. It would be much better just to classify cases according to the age group criteria and then let SPSS compute frequencies for each group. Moreover, the creation of grouping variables is vital in order to perform a range of more detailed kinds of analysis. Just jot your provisional answers down for now and you can correct your answers (if necessary) later.

Recoding age

You will now use the **Recode** command to get SPSS to create a new variable that assigns each case to the appropriate group:

- Go to the menu and follow the path: Transform → Recode into different variables
- Select "age" from the variable list on the left of the dialogue box. Click on the adjacent arrow button. You should see the expression "age --> ?" appear in the box labelled "Numeric Variable - > Output Variable". To complete the recoding expression:
    - Type in the name, as defined in the code book, for the Output Variable. Note that a variable name must be one word only (no spaces). Click on the "Change" button to confirm the entry
    - You should see the expression has changed to "age --> agegp3"
    - The label is optional. You can leave this blank for now (you can edit this property in the variable view) or enter something more meaningful like "Age Group".
- Finally you need to specify the group value ranges. To do this click on "Old and New Values". A new dialogue box should appear. Each group will be assigned a number or 'code': 1 for <=37, 2 for 38-50, 3 for 51+. Here's how to specify the first category:

- In the frame labelled “New Value” select the “Value” option and type the number “1” into the adjacent box (omitting the quotations of course). You need to assign numbers to each group initially. You can assign variable labels, to aid readability, later.
- Now look at the options in the “Old Value” frame. There are seven different options (selectable using radio buttons) for specifying the criteria for a group. The most useful ones for this task are:
  - Range – assign cases where the input value falls between specified lower and upper bounds
  - Range, LOWEST through value – assign cases where the value falls at or below some upper bound
  - Range, value through HIGHEST – assign cases where value falls at or above some lower bound
- As this category takes all values equal to or less than 37 years, choose “Range, LOWEST through value”. When selected, an adjacent text box should change to white and become editable. Enter 37 into the box.
- Now click “Add”. You should see the expression “Lowest thru 37 -> 1” appear in the box labelled “Old --> New”.
- Note that we could have defined the range using the two-valued “Range” option, specifying the lower bound as zero and the upper bound as 37, but the former is less verbose.

- Repeat the procedure to define the remaining two categories:
- When you have completed the procedure you should see the following in the “Old --> New” box:
  - Lowest thru 37 --> 1
  - 38 thru 50 --> 2
  - 51 thru Highest --> 3
- If all is well, click “Continue”. This should take you back to the main Recode dialogue.
- Click “OK” and the recoding operation will execute.

Adding Value Labels

Inspect the results in the "Data View". The new variable "agegp3" should be the last column (far right). In order to make things easier to read and help us to remember what the category numbers refer to, we need to label the code numbers that we have used.

- Go into "Variable view" of the data table and scroll down to the bottom of the table to find the "age_grp" row.
- Find the "Values" column and click on the cell that intersects "age_grp" and "Values". A small button should appear on the right of the cell. Click on this button.
- You should see a dialogue box called "Value Labels"
- To label category one:
    - Type the number "1" into the "Value" text box
    - Type "<=37" into the "Label" text box
    - Click "Add" to add the expression to the list below
- Repeat this procedure as required for categories "2" and "3".
- Click "OK" to confirm the new value labels.

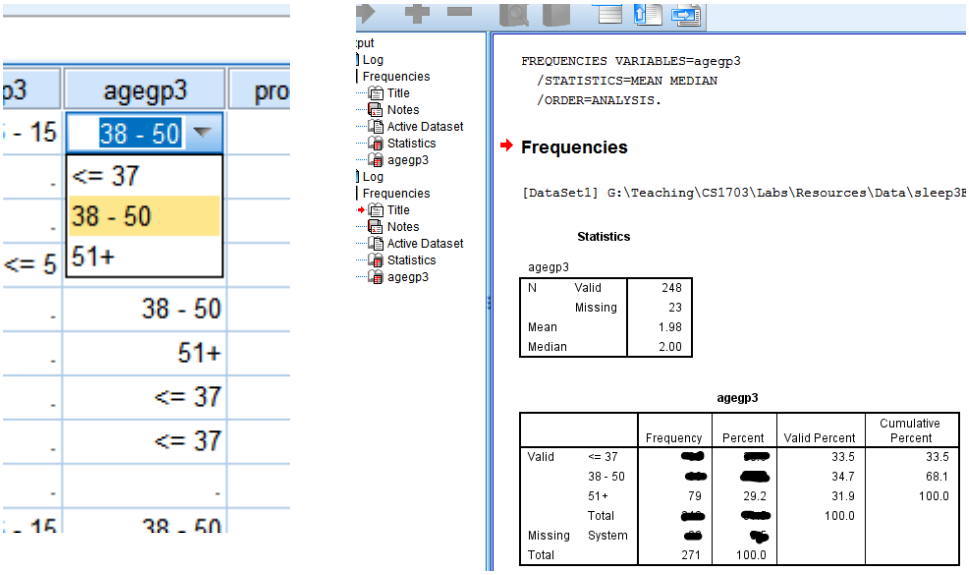
When you return to the "Data view" you should find that all cells in the "agegp3" column are labelled appropriately. If you cannot see the value labels then return to the previous section "Anatomy of the data editor" and re-read it to find out how to switch this feature on. If all is well, you will also see that all defined categories for our new variable are now easily selected from a drop down menu when you click on a cell (see left-hand figure below), which is convenient when modifying cases or adding new cases.

Computing group frequencies

You can now use the **Frequencies** function to check your answers to questions 9, 10 and 11. To compute a frequency table follow the following steps:

- Go to the menu and follow the path Analyze → Descriptive Statistics → Frequencies
- Find and select the variable age_grp within the left hand list
- Use the arrow button to move this into the right hand list
- Make sure "Display Frequency Tables" is ticked (it should be already)
- Click OK

If you look in the output window, you should see a table like the one shown below. Check that the Frequency values for the visible rows contain the same numbers. If this is the case, then proceed to double check your answers to the earlier questions. Were your original answers correct or not?



Summarising groups

Question: Do older people experience more daytime 'sleepiness' than young adults?

For the purpose of this question we will define young adults as those falling into group 1 of agegp3 and older adults as those in group 3. The questionnaire used to collect this data set contained a number of different measures of sleepiness and perceived sleep quality, ranging from questions like "How satisfied are you with the amount of sleep you get?" ("satsleep"), "How tired have you felt over the last week?" ("tired") to more sensitive measures like the variable "totsas" is the overall score for a scientifically developed measure of sleepiness called the "Sleepiness and Associated Sensations Scale". Make yourself aware of the full range of relevant measures, the response scales, and how they appear in the SPSS data table by reading through the document containing the original questionnaire and SPSS coding scheme which you downloaded earlier on.

For the purpose of this basic tutorial we will focus on the variable "sleepy" which was simply ratings in response to the question "*How sleepy have you felt over the last month?*". However, once you have learnt this technique feel free to apply it to find the group means for other relevant variables.

We will now compare the mean average rating on "sleepy" between the youngest and oldest age groups:

- Go to the menu and follow the path Analyze → Compare Means → Means. A dialogue box called "Means" should appear.
- Select the variable "sleepy" from the variable list on the left hand side. Sleepy is our measure or **dependent variable** so we need to move it to the "Dependent list"
- Select the variable "agegp3" from the variable list. This is our **independent variable** so we need to move it into the "Independent list". When comparing groups, the independent

variable is chosen because the analyst believes that it will affect the dependent variable in some way.

- The "Options" button takes you into another dialogue that allows you to specify various statistics to compare groups by. Don't worry about this as the default statistics will suit us fine for now. We will explore this in more detail in later lab sessions
- Click "OK" to execute the operation.

Means

[DataSet1] G:\Teaching\CS1703\Labs\Resources\Data\sleep3ED.sav

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
sleepy * agegp3	[redacted]	[redacted]	[redacted]	[redacted]	271	100.0%

Report

sleepy

agegp3	Mean	N	Std. Deviation
<= 37	[redacted]	[redacted]	2.222
38 - 50	5.55	[redacted]	2.327
51+	[redacted]	[redacted]	2.073
Total	5.61	[redacted]	2.221

In order to see the result of the test, we need to go to the "Output" window. You should see something like the output shown above. The main part of the output appears below under the heading "Means". The "case processing summary" describes the number of cases that were included in the analysis (cases with valid values for both variables) and the number that were excluded. The main table of interest for us, however, is the "Report" which shows the mean response rating on the chosen question for each group, along with number of valid cases and the standard deviation (a measure of variation of scores around the mean).

I've blacked out the important results in the "Report" so you need to check your own results to answer the following questions. Bear in mind that this question had a response scale of 1="not at all" to 10="to a great extent" so the mean of each group should fall somewhere within this range.

Q12: Do older adults report feeling more sleepy than young adults?

Q13: To one decimal place (1 dp), what is the mean sleepiness score for the older adult group?

Bear in mind, however, that just because there is a difference between group means, it **does not** mean that this difference is meaningful or what statisticians call *significant*. It could be down to just chance variation resulting from a small number of extreme/erroneous cases in the data which distort the mean average, especially when the sample size is quite small. We will look at statistical significance, and how to measure it, in later tutorials.

NOTE ABOUT SPSS SYNTAX: Notice that the first section of the previous output seems to describe a formal command:

```
MEANS TABLES=sleepy BY agegp3
  /CELLS MEAN COUNT STDDEV.
```

This is an example of SPSS syntax language. The executed command is output when you compute most SPSS functions. If you are going to be repeating the same or similar analysis many times, you can copy and paste this code into a Syntax window. Follow the path: New → File → Syntax to open an instance of this window. Command lines are terminated with a full-stop. Selecting the whole line and clicking on the run. You can edit the command to change variable names and other parameters just like any other programming language. We will focus in more detail on using SPSS syntax in a later tutorial as you will be making use of these commands during your coursework. However, you are encouraged to experiment with editing and re-running commands as you explore this dataset further.

Selection

Question: If we focus only on the older adult group (51+ yrs), is there a difference between males and females in terms of number of hours they sleep on week nights?

To answer this question we will use the same statistical procedure as before, replacing “agegp3” with “sex” as the independent variable and replacing “sleepy” with “hourwnit” as the dependent variable. However, note that the question specifies that we are only interested in the older age group (i.e. 51+), so we need to filter out all cases that fall into the other categories. To do this we use the following procedure:

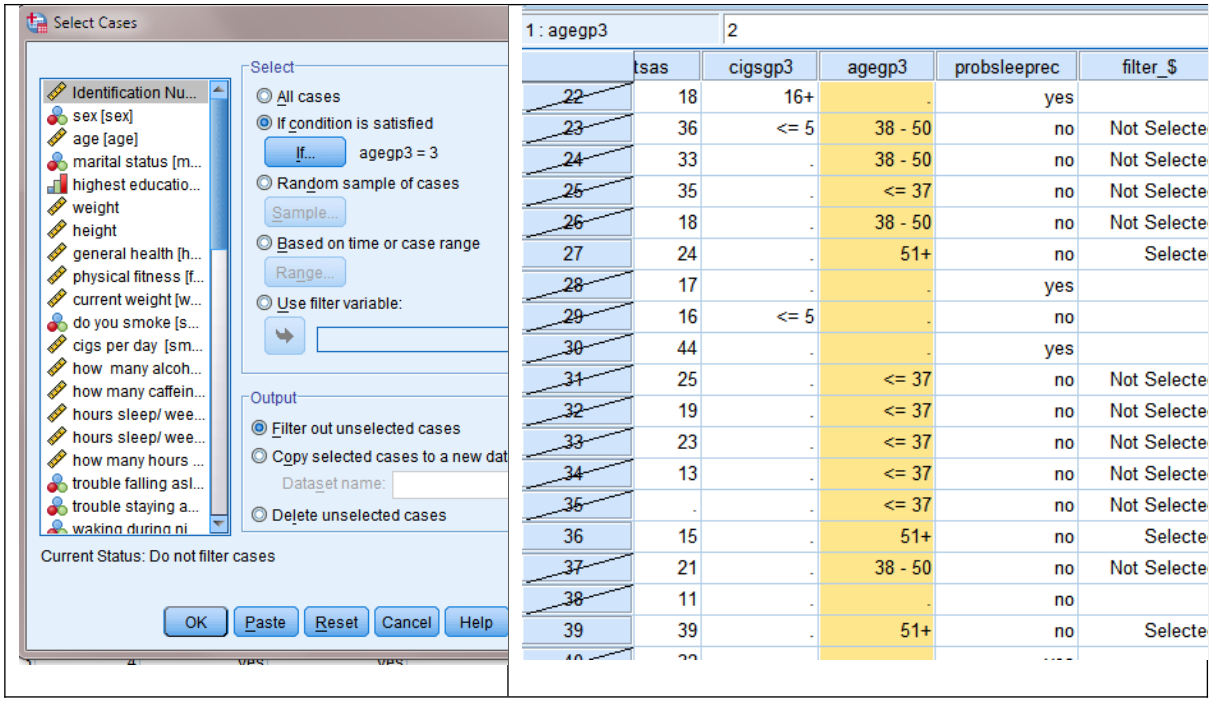
- Bring the Data Editor window to the front.
- Go to the menu and follow the path Data → Select Cases. A dialogue box called “Select Cases” should appear.
- In the “Select” frame, there are several options available. Select the option “If condition is satisfied”, then click the adjacent “If” button. Another dialogue box called “Select Cases: If” should appear.
- Select “agegp3” from the variable list on the left and transfer it into the expression box on the top right. Alternatively you can just type the name of the variable into the expression box (but make sure you spell it correctly).

- The condition we have is that all valid cases must members of the 51+ age group, which we coded as “3”. To complete the conditional expression we append the variable name with “=3“ to give “agegp3 = 3”.

- Click “Continue”. This will close the dialogue box and take you back to the main “Select Cases” dialogue box.

- You’ll see that there is a frame called “Output”. You can simply filter (temporarily disable) the invalid cases, transfer the selected subset of cases to a new data table or delete the unselected cases. We may need the unselected cases later, so it is recommended that you leave this setting on the default for now: “Filter out unselected cases”.

- Click “OK” to execute the selection operation.



You should see that the data view has changed so that all unselected (invalid) cases are filtered out (but still visible). Filtered cases are marked with a strike through the row label cell. For good measure, a new nominal variable has also been created called “filter_$” to show the status of each case. You can use this variable to manually change the selection status of specific case, but we won’t need to take advantage of this functionality for now. Filtered cases will not be included in any further analysis until they are re-selected. You can remove the selection filter completely by simply deleting the variable “filter_$”[3].

Having done this you are now able to use the “Means” procedure again to answer the following questions:

[3] To delete any variable, just click on the column header and hit the delete key

Q14: Amongst older (51+ yrs) adults only, who sleeps longer, on average, on week nights: men or women?

Q15: How many hours per weeknight (to 1 dp) do older males sleep on average?

Summary

In this tutorial, you have experienced an introduction to SPSS. You have learnt about loading a data file, the most important aspects of the data (data table), variable (metadata) and output views and how to perform some basic data processing and analysis tasks, as introduced in the first lecture. You are strongly encouraged to experiment beyond these exercises using the methods you have learnt.

The next tutorial in the series, 1-2, contains some more advanced exercises based on the techniques you have learned today. Before you move on to these exercises, make sure you have correctly answered the 15 questions from this tutorial and recorded your score successfully in the Grade Centre on Blackboard. If there are any aspects of this tutorial that you are not full confident about, please do not hesitate to discuss with a member of the teaching team.

Don't forget to save your data file before you close SPSS along with any output and syntax files that you want to return to later (these must be saved separately).

Further Reading

Pallant, J. (2013) SPSS survival manual : a step by step guide to data analysis using IBM SPSS, Chapter 3, 6