top-2colsgreen-community.eps

mathcentre community project

encouraging academics to share maths support resources

All mccp resources are released under a Creative Commons licence

|  |
| --- |
| **Getting Started with Data Handling in SPSS – Audio Transcript**  **Title slide**  Welcome to the ‘Getting started with data handling in SPSS’ video tutorial.  **Overview**  In this tutorial, you will learn about the different windows in SPSS and how to navigate them, as well as how to complete some basic computations and statistical calculations.  It is recommended that you complete the steps in SPSS yourself alongside this video. To do this, please download the accompanying SPSS data document file entitled “Getting Started With SPSS”.  The video demonstration that follows was created using SPSS version 22. Other versions hold the same functionality, though older versions may look slightly different.  This tutorial is aimed at complete beginners, so if you have never used SPSS before, then this is the tutorial for you.  **Navigation**  Moving to SPSS, we can see a small dataset that we will be using for our examples. It contains data relating to ten online purchase transactions.  There are six pieces of information recorded for each transaction, ranging from the date to the customer feedback score.  Data must always be stored in this format in SPSS; where each row represents an observation or case – here, an online transaction, while each column represents a variable – which here is information about the transaction.  **Data View and Variable View**  There are two tabs in SPSS relating to how we can view our data.  The *Data View* tab, currently selected, gives us a display of our actual data. It is also where we can type in new data, delete data, or change existing data.  *Variable View* gives us the metadata. Metadata is the data about our data, so it describes the data we have. In *Variable View*, we can see the list of variables, alongside various other pieces of information.  **Exploring Variable View**  The *Name* is simply the name of the variable.  The *Type* describes the category of variable. Clicking on one of them, followed by the button containing three dots, allows us to see the different types of variable available, as well as different formatting options.  The *Width* is the maximum number of digits permitted for cases within that variable. For example, Department has a width of eight. This is fine here, as our departments are Books, Clothing and Homeware, but if we had a new department name containing more than eight letters, then we would need to change this number.  *Labels* are useful if your variable names are not self-explanatory. Also, variable names cannot contain spaces, but labels can. If I expand the *Name* column we can see the fifth variable is the delivery method. We may want to label this without the underscore. This will re-label the variable in our output, which we will see later. Let’s also change the name of the variable simply to DM.  The *Values* column allows us to assign individual labels to each value that a variable can take. Delivery method was originally recorded as a number, which is quick to input but not so easy to interpret. We can use this option to assign descriptions to the numbers. If we click in the relevant box and then on the smaller box containing three dots, we can see the four labels that are already defined. Within this box, we have the opportunity to change labels, add new labels, or remove existing labels.  The final column that I wish to mention here is the *Measure* column. *Measure* gives us extra information on the variable type, in addition to the *Type* variable that we have already seen.  **Differences between Measures**  It is important to understand the differences between measures, as this can inform and influence the type of analyses that are available to us.  Nominal variables are classified qualitatively. This means that there is no order to them. Examples include a person’s gender, the city they live in, or ethnic origin.  Ordinal variables do have an order, but on a defined scale. Examples include Likert scales, grouping of scale variables, and opinion questions in a questionnaire.  Scale variables are continuous. Order is certainly important here, and variables do not have a predefined scale.  Make sure you look back at our data set and understand why the variables have the measures shown.  **Navigating in Data View**  Returning back to *Data View*, we’ll now explore some of the options available in the main menu.  The majority of the options available in the top tool bar are self-explanatory. For example, if we select *View*, and then *Fonts*, we change the font, the font style or the size.  The *Analyze* menu contains all of our statistical analysis options, while the *Graphs* menu contains our graphics options.  We’ll now quickly look at a few of the shortcut icons on the main toolbar. Again these are mostly self-explanatory. For example, *Go to case* allows us to go to a particular case within our data set. Remember that a case refers to an observation, or a row, and is particularly useful if you have a large number of cases to save you scrolling through all of your data set.  The next button is similar but instead we are selecting a variable. If we click *Go to variable*, we can select from a dropdown menu, and that variable will then be selected. Again this is more useful for larger data sets, where you may have to scroll through many variables to find the one that you are looking for.  The next icon we’ll look at is *Find*, which is the pair of binoculars. Selecting cases was easy here as we have a small dataset, but with larger datasets again the *Find* option may be useful. Click within the variable you wish to search, and type the search string. Here I’m using the Department variable and searching for Books.  The final icon that we’ll look at is *Value Labels*. Clicking this switches between displaying the numeric or string versions of variables. Remember that the values for Delivery Method were defined also as strings. We click the *Value Labels* button to see them. Click it again to convert back to the numerical value.  Finally, we’ll look at sorting the data. To sort by a variable, right click on the variable and sort ascending or descending.  **Computing variables**  You may need to calculate the values of a new variable, based upon existing values within the data set. For example, we are going to use *Cost* to calculate how much of the total cost was Value Added Tax, or VAT.  To do this, select *Transform* and then *Compute Variable…*.  The *Target Variable* is the name of our new variable; in this case, VAT.  We write our formula for this new variable in the *Numeric Expression* box.  To use another variable, we can either type the name directly, drag the variable from the list or simply double-click on the variable.  We then continue to type the rest of the expression, or we can use the buttons below. VAT is currently 20%, so we need to multiply *Cost* by 0.2.  There are a number of other functions which may also come in useful on the right hand side.  Click *OK* to perform the computation, and close down any output document that may appear. There is no need to save it.  We can see that a new variable called *VAT* has been created, which has calculated 20% of the *Cost* variable.  **Simple data analysis**  Statistical analysis options come under the *Analyze* toolbar.  We’re going to perform one simple analysis here, but it is advised that you explore these options in your own time to fully appreciate what is available.  We’re going to calculate some summary statistics on customer feedback score, grouped by delivery method.  Select *Analyze*, *Compare Means* and then *Means*.  The *Dependent List* is the list of variables to perform the analysis on, while the *Independent List* is for grouping variables. We want statistics on customer feedback score, so we drag that to the *Dependent List*. Alternatively we can use the arrows.  We want to group by delivery method, so we drag this to the *Independent List*. Note that delivery method has both the original variable name, and the label that we gave it earlier.  Clicking *Options…* allows us to choose which statistics we want to see. Again we can use the arrows or drag and drop. I’m going to remove Standard Deviation, and add in the Minimum and Maximum.  Click *OK* to perform the analysis.  An output window will appear automatically. We’re given a summary at the top, and then our requested statistical analysis underneath. This output can be saved in the usual way, while the tables can be copied or amended by right-clicking on them.  Note that for delivery method, instead of the variable name which is DM, the delivery method label is given. Additionally we see the descriptions of each numeric value, instead of the numeric value itself, so this helps us with interpreting the data.  **Word of warning!**  Most statistical tests have assumptions associated with them. It is up to you to check these assumptions to ensure that the test you perform is valid.  It is important to note that SPSS will let you perform statistical analysis that is not necessarily statistically sound.  **Additional help and resources**  That concludes this ‘Getting Started With Data Handling in SPSS’ tutorial. Thank you for watching and I hope you have found it useful.  For additional support and resources, see the mathcentre and statstutor websites, and help files provided by SPSS. |