Digital literacy training

SPSS

Introduction

2018
# Table of Contents

To start SPSS .................................................................................................................. 1  
Import Excel files ............................................................................................................. 1  
The Data View window ...................................................................................................... 2  
The Variable View window ............................................................................................... 3  
Create new variables ........................................................................................................ 4  
Create graphs ................................................................................................................... 4  
  Edit histograms .................................................................................................................. 5  
  Edit scatterplots ............................................................................................................... 5  
  Edit bar charts .................................................................................................................. 6  
Descriptive statistics ........................................................................................................ 7  
Use a Syntax file ................................................................................................................. 9  
Export graphs and tables to Word ..................................................................................... 10  
Data manipulation ............................................................................................................. 10  
Practice ............................................................................................................................ 11  
Other resources ............................................................................................................... 12
SPSS Introduction

Presenter: Nyree Mason nyree.mason@gmail.com

The IBM Statistical Package for the Social Sciences (SPSS) was developed as a data management and analysis tool. This course will show you how to:

- Open and import data.
- Set-up data files using the Variable View window.
- Manage data using the Data Editor window.
- Create graphs and tables and export them to different file formats.
- Obtain descriptive statistics (i.e., percentages, means and standard deviations).
- Use the Syntax Editor to keep records of data manipulation & analyses and to run routine analyses on multiple data files.
- Compute new variables and recode existing ones.
- Sort variables.
- Select cases and split files to analyse subgroups separately.
- Merge files.

The data files

goo.gl/4ncQnw

Save these 5 files:

Employee_Data.sav
Anxiety_2.sav
Car_sales.sav
Employee_2.sav
Employee_3.sav

* These files will be used for the Advanced Significance Testing SPSS course.
To start SPSS

1. Log on to the computer using university ID (e.g., ‘u1234567’) and password.
2. On Windows Desktop slide down to find IBM SPSS Statistics.
3. SPSS File → Open → Data look in university ID folder (e.g., ‘u1234567’) open Employee_data.sav data file and click OK (all data files have the *.sav extension in SPSS).

**Note:** You can open a file any time within SPSS using File → Open → Data. SPSS allows you to have multiple data files open at once, and has reminders in the Output file to let you know which data file was used for each analysis you perform. You can also open any file from another spreadsheet style application (e.g., Excel) following the SPSS prompts.

Import Excel files

**Note:** You will need to have your variable names in the first row of the spreadsheet so that SPSS recognises them as labels. Do not have any annotations etc. underneath your data in the Excel spreadsheet, as this will confuse SPSS about your variable types during the import.

1. Go to File → Open → Data and in the drop-down menu Files of Type choose Excel and then find your file.
2. Select read variable names from the first row of data if necessary, and if there are multiple sheets in the Excel file select the relevant one. Then click OK.
3. You will have to set up the data file afterwards in the variable view (see details below).
The Data View window

Click on the **Data View** tab at the bottom left corner of the screen. This is where you can view and enter your data in a similar way to any other spreadsheet-style program.

- There is one column per variable (e.g., date of birth, gender, salary).
- There should be only one participant/case per row.
- Columns can be rearranged: highlight one column by clicking on the variable name, and then click and drag the name to another position (this can also be done in the **variable view**).
- Extra columns can be inserted into the data set by clicking on the name of the variable you want the new column to go next to, and then going through the menu: **Data → Insert Variable**.
- You can also delete whole columns by clicking on the name of the variable you want to delete (thus highlighting the whole column) and then pressing the **delete** key.

**Note:** The **Undo** option in the **Edit** menu may come in handy if you make a mistake.
The Variable View window

Click on the **Variable View** tab at the bottom left corner of the screen. This is where you tell SPSS about your data (i.e., names, labels, type, levels of measurement). The information you provide in the **Variable View** can help you avoid making common mistakes in data analysis: SPSS will limit the types of analyses you are allowed to perform to those which are appropriate to the statistical properties of your data.

- **NAME** determines what the columns will be called in the **Data View**.
- **TYPE** determines what form the data in each column takes (e.g., **jobcat** is **numeric** because numbers are entered, **gender** is **STRING** because letters are entered, **bdate** is **DATE** because dates are entered).
- **WIDTH** determines how many digits/letters can go in each column (Note: NOT how many are displayed). **WARNING:** Do not change the width after the data is entered unless you are 100% sure you want to. Old versions of SPSS will delete any characters that are already entered that go beyond the new width limit without warning! (i.e., if the width is set to 4 a number such as 12,345 will be changed to 1,234).
- **DECIMALS** determines how many decimal places are displayed (i.e., 3 \( \rightarrow \) 12.345).
- **LABEL** gives the variable a more descriptive label in the **Output Viewer**.
- **VALUES** labels the coding values entered in the column (i.e., m = male, 1 = Strongly Disagree), and will present these in the **Output. WARNING: SPSS prefers numerical codes, and some statistical analyses cannot be performed unless the data is coded with numbers.**
- **MISSING** determines which coding values will be considered missing and as such will be excluded from statistical analysis (e.g., 0 = “Other” Response, 77 = Invalid Answer, 88 = Not Applicable).
- **COLUMNS** determines how wide the columns will be displayed in the **Data View**. You can change this manually by clicking and dragging the columns wider in the data view also.
- **ALIGN** changes the alignment of the data characters (Left, Centre, and Right).
- **MEASURE** determines the level of measurement of each column, either: **nominal** (e.g., types of running shoes), **ordinal** (e.g., university grades: Pass, Credit etc.), or what SPSS terms **scale**, which includes both **interval** and **ratio** levels of measurement (e.g., real numbers: temperature, weight, time, frequency etc.).
- **ROLE** is designed to help you choose the right variables for analysis (Input = Independent variables, Target = Dependent Variables etc.). It is not fully functional in SPSS as yet.

**Handy Hint:** In the **Data View** you can see the labels associated with the raw data by going to **View \( \rightarrow \) Data Labels**. When entering data you can also use this function: click on an empty cell, use the drop-down menu to select the appropriate label for your data and SPSS will fill in the number for you automatically (behind the scenes).
Create new variables

Try creating a new variable for job satisfaction called jobsat, which represents levels of agreement to the statement: “I am satisfied with my job”. It is measured on a 5-point scale with these value labels:

1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree
99. Not applicable

Any “Not applicable” values will be coded as a missing value of 99.

Handy Hint: You can copy one value and paste it to multiple cells at once in the Variable View: copy one missing value, select the cells you want to paste it to and then paste.

Create graphs

HISTOGRAMS describe scale data best, but is also used for ordinal data. It can display what the normal curve would look like given the mean and standard deviation of your data. Graphs → Legacy → Histogram select a continuous variable (e.g., current salary) and tick display normal curve then OK or Paste. These can be split into groups (e.g., by gender) using Panel by Rows (graphs one above the other) or Panel by Columns (graphs side by side). This has the advantage of maintaining identical X and Y axis values.
Edit histograms

- Double-click on the graph to open the Chart Editor and Properties windows. If the Properties window doesn’t open automatically, just click on the graph in the Chart Editor and it will appear.

- Whatever you want to change, the easiest option is to click on it and the Properties window will give you all the appropriate formatting options for that feature in separate tabs.

- For example, to change the range and increments used on either axis, click on one of the values, select the Scale tab in the Properties window and change the values you wish.

- To change the colour of bars for example, click on the bars and Properties window will give you a tab for Fill & Border options. This allows you to change the look of the bars – even add a pattern. You can also change the binning here (width of the bars).

SCATTERPLOTS describe the relationship between two scale or ordinal variables. Graphs ➔ Legacy ➔ Scatter ➔ Simple select one scale variable for the x-axis (usually the IV, e.g., education level) and another for the y-axis (usually the dependent variable, e.g., beginning salary), then OK or Paste. These can also be panelled if you wish.

![Scatterplot graph]

Edit scatterplots

- Double-click on the graph to open the Chart Editor

- To get the linear regression line, go to Chart ➔ Elements ➔ Fit Line at Total. The R squared value will be displayed in the top right-hand corner to indicate the amount of variance explained by the line. With the line highlighted you can change the type you want in the Properties Window in the Fit Line tab (in this example a Quadratic equation was used as it provided a better fit to the data).

- Other elements can be changed as per the Histogram example above. To find the case number of potential outliers, go to Chart ➔ Elements ➔ and tick Data Label mode. This changes your cursor to a target, and click on the data point in question to see which case number (row number in the data view) it is.

- Close the Chart Editor window to return to the output.
**SPSS Introduction**

**BAR CHARTS** are useful for describing the frequencies/percentages of nominal and ordinal data. They look like histograms but have fixed bars which can be ordered in any way you wish. To produce a Clustered Bar Chart go to **Graphs  Legacy  Bar  Clustered** with a Nominal/Ordinal variable as the **Category Axis** (e.g., job category) and a Nominal/Ordinal variable (e.g., minority) in the **Define Clusters by** box. Check the **% of cases** radio button to display percentages on the Y-axis rather than counts.

![Clustered Bar Chart Example](image)

**Edit bar charts**

- Double-click on the graph to open the **Chart Editor**
- You can change the order of the categories on the X-axis by clicking on a category label, then in the **Properties Window** go to the **Categories** tab and rearrange the order by clicking on a category and using the up and down arrows.
- You can do the same for the order of the bars by clicking on a bar, selecting the **Categories** tab and rearranging the order the same way.
- Other elements can be changed as per the Histogram example above.
- Close the **Chart Editor** window to return to the output.
**Descriptive statistics**

**Note on Terminology:** Dependent variables (a.k.a. Response) are those with which you expect to measure some effect. Independent variables (sometimes called Factors in SPSS) are those that you expect to have an effect on the Dependent Variables. For example, if you think males will be taller on average than females, sex will be the Independent Variable and height will be the Dependent Variable.

**DESCRIPTIVES** are used to primarily to produce means and standard deviations for scale data. For example, to find the mean and standard deviation for education level: Analyze → Descriptive Statistics → Descriptives. Choose education level and move it into the Variable(s) box with the arrow button. In the Options area you can also choose to display sum, variance, standard errors, range, skewness and kurtosis.

**Note:** The information appears in an Output Viewer Window, which can be saved separately with the extension *.spv. Please note that frequently output files from older versions of SPSS cannot be viewed properly in newer versions. Exporting as another file type (described later) will avoid any problems. Also, the material in the output viewer is completely independent of the data. If you change the data in any way, all analyses will have to be run again.

**FREQUENCIES** Used to describe nominal/ordinal data. To find the frequencies and percentages of the sample in each minority classification: Analyze → Descriptive Statistics → Frequencies. Choose jobcat. You can also produce multiple bar charts, histograms or pie chart using this option. In Charts select Bar and tick percentages. For scale or ordinal data you can produce a histogram with normal curve but be careful to untick 'display tables' for scale data before you run the analysis or you will probably get very large tables as well.

**Note:** The Percent column calculates the percentage out of the total number of cases including missing cases. The Valid Percent column calculates the percentage out of the number of valid cases which excludes missing cases.

**Frequencies** can be used to produce the same “Descriptives” for scale data also, but provides extra information particularly useful for ordinal data: quartiles, percentiles, median and mode. Select a scale variable, untick “display tables”, click on Statistics and select the statistics you require.
EXPLORE is useful when you want descriptive statistics for scale data broken into groups (e.g., means for males vs females). To find the difference in current salary between males and females: Analyze → Descriptive Statistics → Explore. Put the dependent variable (which must be scale) beginning salary into dependent list and the independent variable(s) (which must be nominal/ordinal) gender into factor list box. In Statistics choose Descriptives and you’ll note that 95% confidence interval is the default. You can also select outliers if you want SPSS to show you the cases that have values greater than 3 standard deviations away from the mean. In plots unselect stem-and-leaf and select box plots (so you can see the outliers, and assess normality).

### Editing Tables

- Double-click on the table so that it gets a dotted line around it. Move your mouse onto the right line of the statistic box. When you get the double-headed arrow, click and hold out the mouse while dragging it towards the right. This should expand the box so that more figures are displayed. You can minimise and maximise the width of all columns this way.

- To change the text, double-click on a label type the new text. You also have a Formatting Toolbar which you can use to change the font etc..

- To delete a value from a table click once on a value and press the delete button. If you delete the corresponding statistic for all categories, the whole line will disappear. You can delete all unnecessary/unwanted information in the table this way.

- To change the number of decimals displayed, highlight the appropriate values and right-click on the selection, and select Cell Properties from the menu. In the Format Value tab there is a decimals option.
CROSSTABS can be used for nominal/ordinal data only. To find the relationship between employment category and gender: Analyze → Descriptive Statistics → Crosstabs. In the Rows box enter a Nominal/Ordinal variable (e.g., jobcat) that you want represented in the rows of the contingency table. In the Columns box enter the Nominal/Ordinal variable you want in the columns (e.g., gender). Click the Cells button and note than Observed is the default in order to display the actual count in each cell. You can also choose to display Row and/or Column Percentages here. Format again, is not necessary.

<table>
<thead>
<tr>
<th>Employment Category</th>
<th>Gender</th>
<th>Count</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>208</td>
<td>363</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>157</td>
<td></td>
</tr>
<tr>
<td>% within Employment Category</td>
<td>56.7%</td>
<td>43.3%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% within Gender</td>
<td>95.4%</td>
<td>60.9%</td>
<td>76.6%</td>
</tr>
<tr>
<td>Custodial</td>
<td>Count</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>% within Employment Category</td>
<td>0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% within Gender</td>
<td>0%</td>
<td>10.5%</td>
<td>5.7%</td>
</tr>
<tr>
<td>Manager</td>
<td>Count</td>
<td>10</td>
<td>74</td>
</tr>
<tr>
<td>% within Employment Category</td>
<td>11.9%</td>
<td>88.1%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% within Gender</td>
<td>4.6%</td>
<td>28.7%</td>
<td>17.7%</td>
</tr>
<tr>
<td>Total</td>
<td>Count</td>
<td>218</td>
<td>258</td>
</tr>
<tr>
<td>% within Employment Category</td>
<td>45.6%</td>
<td>54.4%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% within Gender</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Use a Syntax file

- Repeat the previous Crosstabs analysis but click Paste instead of OK this time.
- PASTE places the commands that SPSS uses to execute its tasks to a Syntax File (*.sps).
- Useful for when you want to run the same analyses again on different variables or data sets. It is also a useful record of what you have done to the data file.
- In the Crosstabs syntax, try changing the variable name gender to minority and then highlight the syntax and then Run → Selection.

CROSSTABS
/TABLES= jobcat BY minority
/FORMAT= AVALUE TABLES
/CELLS= COUNT ROW COLUMN
/COUNT ROUND CELL.
Export graphs and tables to Word

**CUT-AND-PASTE METHOD** Open a Word document (Start → Programs → Word). When you are happy with how the table/graph looks in SPSS, in the Output Viewer select the table, then Copy. In Word go to Edit → Paste.

**EXPORTING METHOD** You can export the whole output document in one go, or select certain objects in the Output to export by using CTRL-Click method. In the Output window, go to File → Export, and you have the option to save the file as a Word, HTML, pdf, PowerPoint, Text (no graphs), Excel (no graphs) or “None” (separate picture files for graphs only) file types. It’s usually best to export visible objects only otherwise you get a lot of non-visible SPSS syntax in the new file.

Please note that graphs will be exported as picture files and can no longer be edited (apart from size).

Data manipulation

**COMPUTING NEW VARIABLES** is useful for creating new variables based on the data you have or transforming non-normal data, etc. To transform skewed variable (e.g., salary) for further analysis, go to Transform → Compute then type log_salary in the target variable box. This will be the name of the new variable in your data file. In the Function Group box click on Arithmetic, and this will display all the arithmetical options in the Functions and Special Variables box. Here, select Ln (for Natural Log) and click on the up arrow to move it into the Numeric Expression Box. Replace the “?” symbol with the variable salary (type or select from the list and move it into the box). You should have the formula: LN(salary). Then click OK or Paste and Run, and a new variable will be displayed at the end of your data set. Save the file again, so that this change will not be lost.

**Please Note:** To raise a number to a power, use the symbols ** (NOT ^ as you would in Excel).

**RECODING VARIABLES** is useful for grouping or regrouping variables (e.g., if you want to turn education in years into education levels <=12, 13-15, and >=16). Transform → Recode → Into Different Variables (so you don’t overwrite an existing one) and in Name type edulevel and click Change. In Label, type Education Level to give it a name in the output. Then go to Old and New Values and click the Range radio button. Select Lowest through ? radio button and type 12, then in New value type 1, then click Add. Select the ? through ? radio button, and type 13 through 15. Give this range the New value 2, then Add. Finally, select the ? through highest radio button, type in 16 and give the range the New value of 3, then Add. Then Press Continue, and then OK (or Paste and Run). You should have a new variable called edulevel in your Data View. If you go into the Variable View by clicking the tab at the bottom left of the screen, you can label these values (when a Values cell is selected click on the blue square and choose the appropriate options (e.g., type in value “1”, and label it “8-12” then Add and so on).
Practice

See what changes in the **Frequency Histogram** graph and the **Explore** analysis for the new **log_salary** variable.

**SELECTING CASES** is useful when you want to perform statistics on only a subsample of your data set. For example, to make sure that any analysis performed on the data is only performed on those cases that have been in the job for more than or equal to 72 months (6 years), follow these steps: **Data → Select Cases** then choose the **If condition is satisfied** radio button. Then click **If...** and move Months since hire (jobtime) to the box. Then select the greater than or equal to symbol (≥) and type “72” (so you get this formula: jobtime ≥ 72), then **Continue**, then **OK**. Note that in order to perform statistics on the whole data set from now on requires going back into **Data > Select Cases** and checking the **All Cases** radio button, then **OK**. SPSS shows you that it has done what you asked by putting diagonal lines through the case labels being excluded. In the bottom right hand corner it should also say **Filter On**.

**SPLITTING FILES** is useful when you want to analyse two or more groups within the data set separately. Split the files into those groups by **Data → Split File** and select the **Compare groups** radio button. This option splits tables by the variable selected. You could also use **Organize output by groups** and this gives you the same information, just in separate tables in the output. Move the grouping variable into the **Groups based on** box (e.g., gender), make sure the **Sort file by grouping variables** is checked, then press **OK**.

**Example of the split file output:**

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>216</td>
<td>$15,750</td>
<td>$58,125</td>
<td>$26,031.92</td>
<td>$7,558.02</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>216</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>258</td>
<td>$19,650</td>
<td>$135,000</td>
<td>$41,441.78</td>
<td>$19,499.21</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>258</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**MERGING FILES** is useful when you have two data files you would like to merge (e.g., when two people have been entering data into different spreadsheets with the same set-up), SPSS allows you to merge them.

Please note the files must have identical properties in the Variable View to work seamlessly.

- **TO ADD CASES TO A DATA SET** (also allows you to add variables) **Data → Merge Files → Add Cases**. Then choose the relevant data file. SPSS will automatically place the common variables in the list of variables for the new file. If you want it to include variables that are specific to one file then highlight them and transfer them to the variables box. You can also choose to create a new variable that shows you which data file each case came from by checking the **indicate case source as variable** box.

- **TO ADD VARIABLES TO A DATA SET** Data → Merge Files → Add Variables. Use the same procedure as above to include or exclude variables. You can match up the data for cases/participants by choosing key variables (e.g., participant ID number). Just check the appropriate box and move the key variable(s) into the key variable box. The key variable must be present in both of the files you are going to merge. **Note:** make sure both files have been sorted by the key variable before you merge them.

- E.g., open Employee_2.sav and try adding cases from Employee_3.sav.
Other resources

Training notes
To access training notes, visit the Research & learn webpage anulib.anu.edu.au/research-learn and select the skill area followed by the relevant course. You can register for a workshop and find other information.

Research & learn how-to guides
Explore and learn with the ANU Library’s how to guides (ql.anu.edu.au/howto). Topics covered are:

- Citations & abstracts
- Data Management
- E-books
- EndNote
- Finding books and more
- Finding journal articles and more
- Finding theses
- Increasing your research impact
- NVivo
- ORCID (Open Researcher and Contributor ID)
- Topic analysis
- Using Google scholar from off-campus

Subject guides
Find subject-specific guides (ql.anu.edu.au/subjectguides) and resources on broad range of disciplines. Such as:

- History, indigenous studies, linguistics and philosophy
- Criminal, human rights and taxation law
- Biochemistry and molecular biology, neurosciences and psychology
- Asia Pacific, Southeast Asia and East Asian studies
- Engineering
- Astronomy and astrophysics, earth sciences, mathematical sciences and natural hazards

Online learning
Online learning is available through ANU Pulse (ql.anu.edu.au/pulse), which can be accessed from both on and off campus by all ANU staff and students.

IT skills development modules available in ANU Pulse
- Microsoft Office (Access, Excel, OneNote, Outlook, PowerPoint, Project, Visio, Word)
- Microsoft Office (Mac)
- Adobe suite (Illustrator, Photoshop)
- Type IT

Training calendar
Select Events » near the bottom of the Library homepage to access our events calendar with upcoming training opportunities displayed day by day (ql.anu.edu.au/cal).

Feedback
Please provide feedback about today’s workshop via an online feedback form (ql.anu.edu.au/survey)