stcp-marquier-categoricalS

The following resources are associated:
SPSS dataset 'Titanic.sav', Chi-squared test for association in SPSS

Summarising categorical variables in SPSS

Dependent variable: Categorical

Independent variable: Categorical

Data: On April 14th 1912 the ship the Titanic sank. Information on 1309 of those on board will be used to demonstrate summarising categorical variables. The dataset *Titanic.sav* includes the variables such as class of passenger and whether or not they survived.

Variable name	pclass	survived
Name	**Class of passenger**	**Survived 0 = died**
Abbing, Anthony	3	0
Abbott, Rosa	3	1

Research question: Did class affect survival?

To break down survival by class, a cross tabulation (contingency table) is needed.

To produce a basic contingency table go to: *Analyse→ Descriptive statistics → Crosstabs*

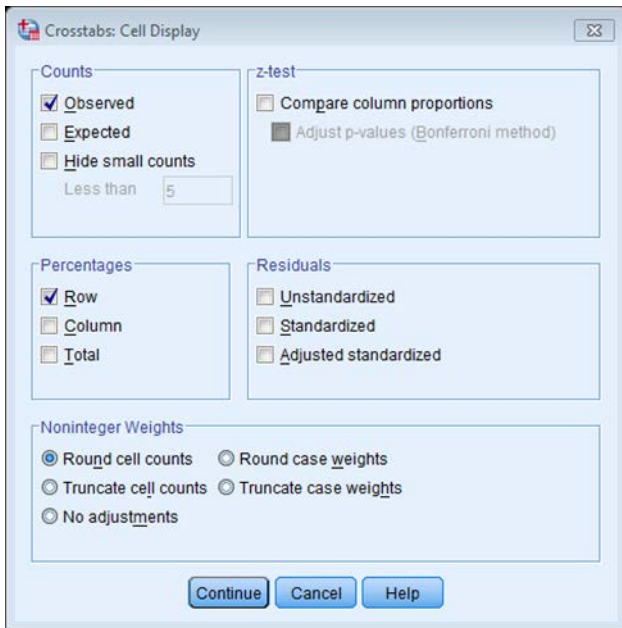Move *Class* to the 'Row(s)' box and *Survived* to the 'Column(s)' box.' and click **OK**.

The resulting table is called a 3x2 contingency table as there are 3 categories in the row variable (class) and 2 categories in the column variable (survived).

		Survived	Survived	Total
		Died	Survived	
Class	1st	123	200	323
	2nd	158	119	277
	3rd	528	181	709
Total		809	500	1309

Unless you have very small sample sizes, percentages are usually preferable to frequencies as they are easier to compare especially when group sizes differ.

To produce percentages repeat the procedure above but then click on the **Cells** button in the Crosstabs menu and choose the required percentages from row, column, or total. Choose which percentages best answers your research question. Here, we are interested in the percentage dying within each class so we need row percentages. It would be misleading to use column percentages as there were more people in 3rd class.

Click 'Row' percentages, **Continue** and then **OK**.

It's clear that class and survival are related as only 38% of people in 1st class died compared to 75% in 3rd class.

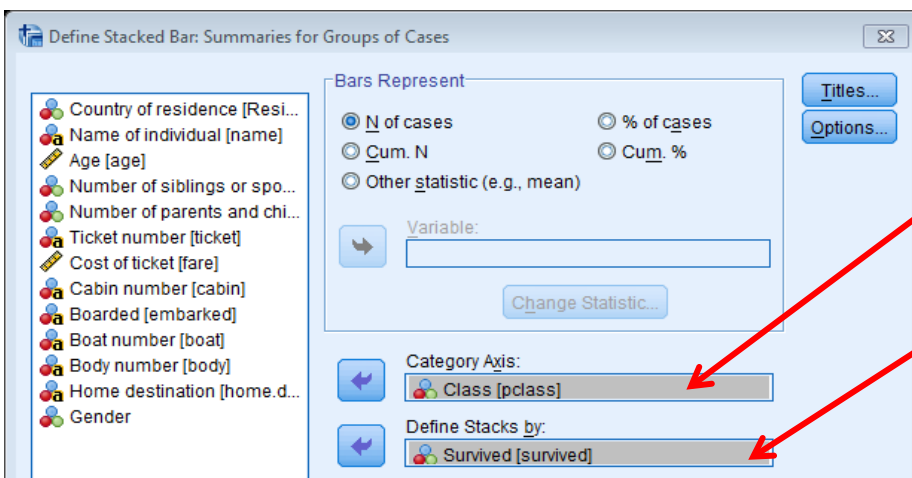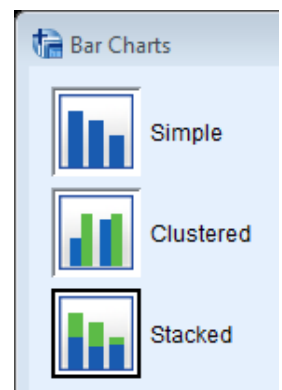If the table is included in your report, do not report all percentages in the write-up.

Class * Survived? Crosstabulation

| | | | Survived? | | Total |
			Died	Survived	
Class	1st	Count	123	200	323
		% within Class	38.1%	61.9%	100.0%
	2nd	Count	158	119	277
		% within Class	57.0%	43.0%	100.0%
	3rd	Count	528	181	709
		% within Class	74.5%	25.5%	100.0%
Total		Count	809	500	1309
		% within Class	61.8%	38.2%	100.0%

Bar Charts

Bar charts are the most common way to display categorical data. To produce a bar chart, go to *Graph → Legacy Dialogs → Bar*
Choose from three types of bar chart in the first screen. A simple bar chart is for one categorical variable. To display the information from the cross-tabulation graphically, use either a stacked or clustered bar chart. Select 'Stacked' to access the following screen.
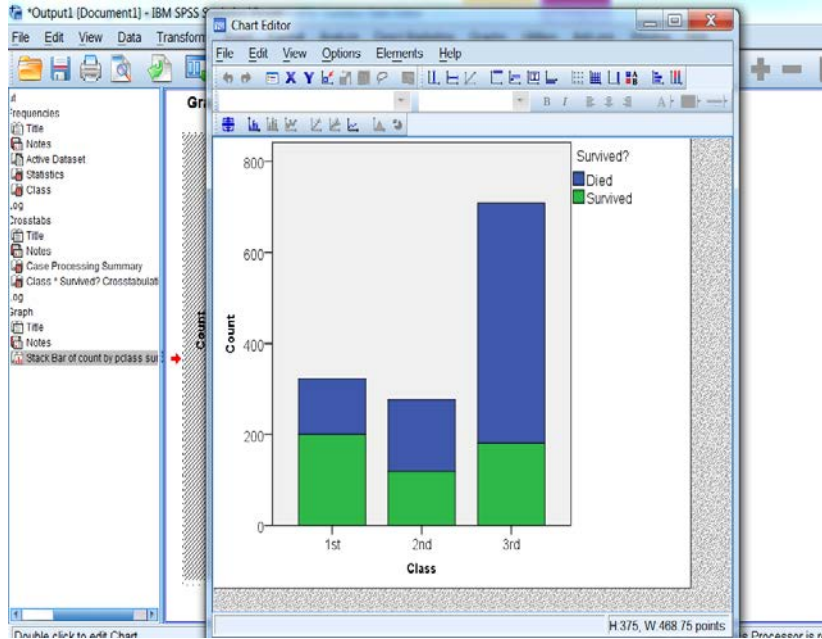

Move *Class* to the 'Category Axis' box and *Survived* to the 'Define stacks by' box.

Note: It is possible to produce bar charts to display means of continuous variables by group by selecting 'Other statistic (e.g. mean)' and moving the continuous variable to the 'Variable' box.
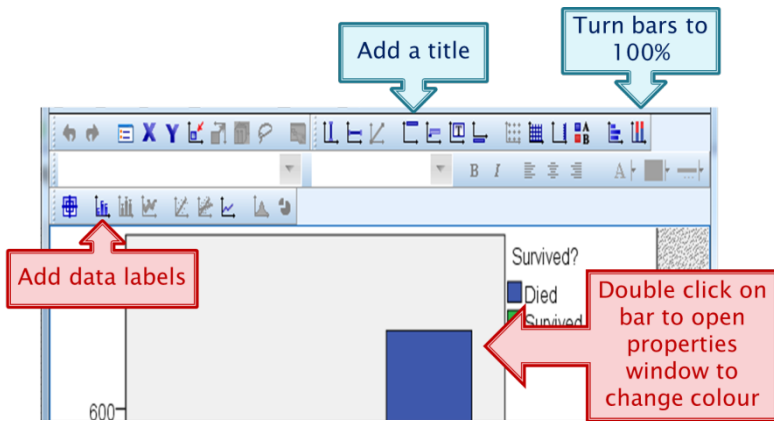
The chart shows the frequencies of those dying and surviving within each class. It looks like class and survival are related but it's not always clear with different frequencies within each group. Comparing percentages is often better.

Double click on the chart in the output window to open an editing window.

Tidying up a bar chart

Add a title

Turn bars to 100%

Add data labels

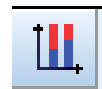Double click on bar to open properties window to change colour

There are many adjustment options in the chart editor window which are accessed through the toolbar or by double clicking on the part of the chart to be edited e.g. bar.
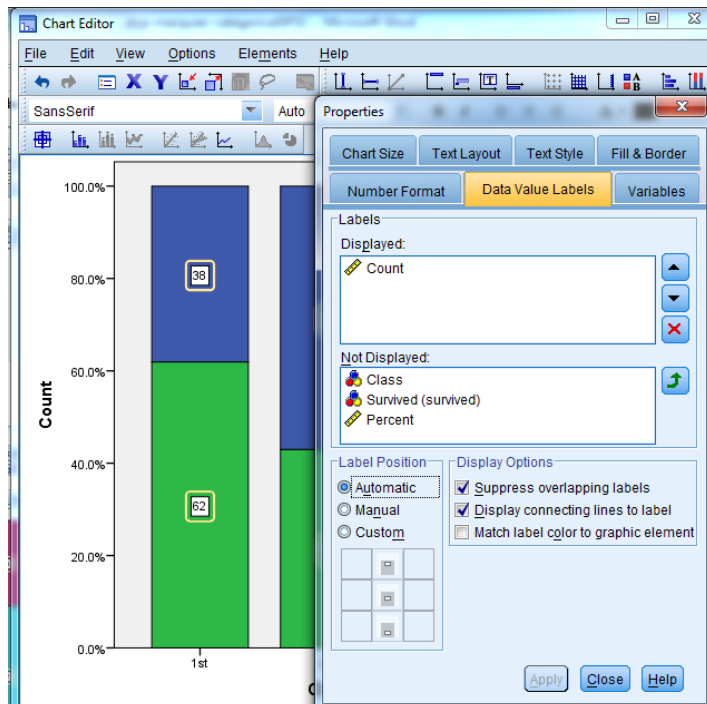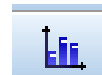
Add a title to the graph

Change the bars to display percentages within groups

Add labels to the bars

Editing Labels

Once the labels have been added to the bars, double clicking on one label will open the label properties window. Percentages are usually more useful so move 'Percentage' up to the 'Displayed' box and remove 'count'.

Use the '**Number Format**' tab to reduce the number of decimal places to 0. Whole numbers are clearer when reporting. Finally, use the '**Text Style**' options to increase the preferred font size to 12.
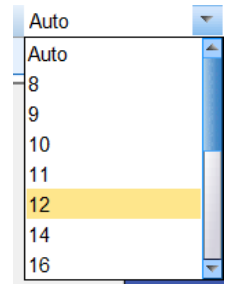
Click **Close** to return to the main editor window.

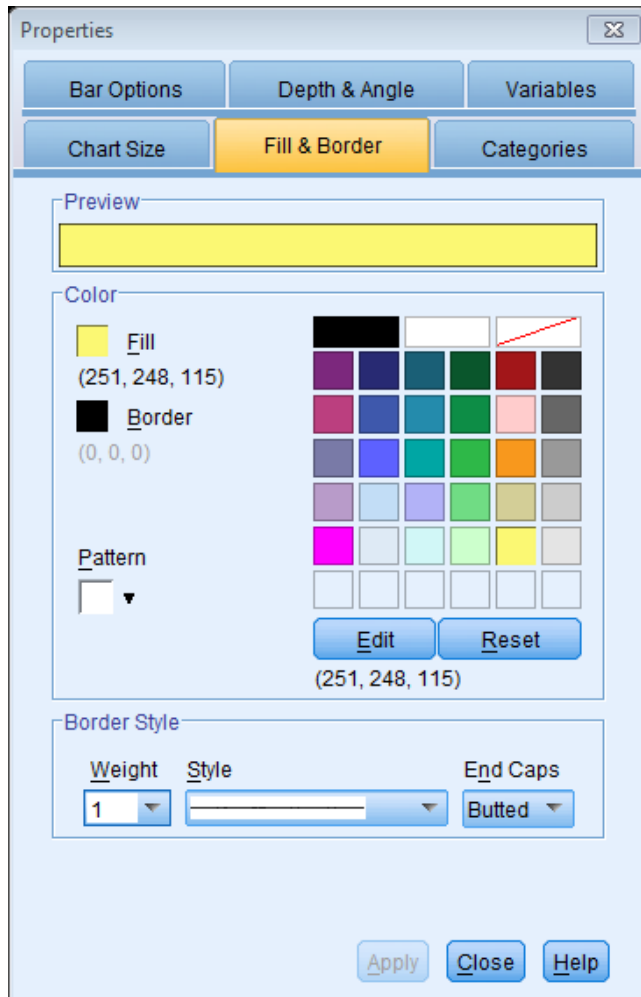Then change the label on the y axis from 'Count' to 'Percentage'.

Adding titles and adjusting font

SPSS labels automatically appear in a small font which is hard to read in Word so adjusting the size of the font is recommended for titles, axes labels and the legend.

Select each axis and change the font size to 12 from the main toolbar. The axis titles and percentages displayed on the bars can also be changed in this way.

Changing the colour of bars

Double-clicking on any part of the chart will open the **Properties** window for that element. The section being edited will be surrounded by a yellow line.

When changing one of the colours in the bar chart, the first click highlights the whole bar (surrounds each bar with a yellow line), then a second click activates one category only. Double click again to open this properties window. To change the colour, click on the box next to the word **Fill** and then select your chosen colour from the options. You have to click on **Apply** after each adjustment to apply the changes.

If you are printing a report in black and white, choose a range of colours from light to dark so the different categories are clear.

Clicking Close takes you back to the chart editor window.

When you have finished all the editing, close the chart editor window to return to the main output window.

Right click on the chart in the output window, copy and paste into word. Sometimes you may need to select 'Copy Special' to move charts. Pasting as a picture enables easy resizing of graphs/output in Word.

Tips on reporting
Do not include every possible chart and frequency.
Think back to the key question of interest and answer this question.
Briefly talk about every chart and table you include but don't discuss every number if the table is included.

Percentages should be rounded to whole numbers unless you are dealing with very small numbers e.g. 0.01%.