Frequency Test

Before we begin using **s1gcseptsnew** in any analysis, we should run a frequency statistic to see how the data in **s1gcseptsnew** looks. The reason for running an initial frequency check is that sometimes, mistakes can be made in entering data into datasets or missing data can appear as usable data. Any time you start working with a new variable, it's good practice to run frequency statistics to make sure the variable looks good.

To check variable frequencies, go to **Analyze**, **Descriptive Statistics**, and **Frequencies**. In the **Frequencies** dialogue box that appears, find **s1gcseptsnew** in the variable list on the left. In a large dataset, you can find variables more easily by right-clicking on the variable list and selecting **Display Variable Names** and **Sort Alphabetically**, which organizes variables and makes them easier to find.
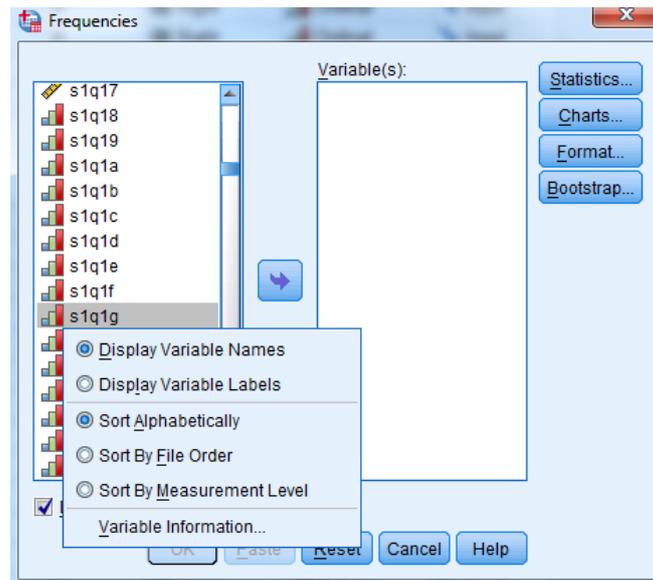
Once you have found **s1gcseptsnew** in the variable list, click on it to highlight it and use the blue arrow to move **s1gcseptsnew** to the **Variable(s)** column on the right side of the dialogue box.

Click **OK** to run the frequency statistic.

Your SPSS output should be a very long table with GCSE scores ranging from ".00" to "857.00." This tells you that the variable **s1gcseptsnew** is continuous (and not categorical), as each data point is a numerical value that falls within a range.

ks4 pts score on new basis not capped

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	233	1.7	1.7	1.7
	8.00	1	.0	.0	1.7
	11.00	2	.0	.0	1.7
	14.00	3	.0	.0	1.7
	16.00	22	.2	.2	1.9
	20.00	1	.0	.0	1.9
	22.00	11	.1	.1	1.9
	28.00	5	.0	.0	2.0
	32.00	20	.1	.1	2.1
	33.00	2	.0	.0	2.1
	34.00	6	.0	.0	2.2
	38.00	15	.1	.1	2.3

However, interestingly, in the YCS, missing data is coded as ".00." This means that the 233 respondents in the ".00" category above didn't actually score zero total points on their GCSEs. This dataset just does not have any data for those individuals in this variable. To prevent these missing data values from being analysed (and potentially skewing our results), we need to tell SPSS that these data points are missing values.

To do so, find **s1gcseptsnew** in the **Variable View** window of SPSS. (You can do this easily by clicking on a variable name in the far left hand column, hitting **Ctrl + F**, and typing in **s1gcseptsnew**.) When you have found **s1gcseptsnew**, simply click to open the **Missing** cell in the **s1gcseptsnew** row, select **Discrete missing values**, and enter .00 as a missing value for **s1gcseptsnew**. Click **OK** to save this information.

Next, run another frequency check on **s1gcseptsnew**, using the steps outlined above. You should see that the values labelled ".00" have been removed from the data and are now listed as "Missing" at the bottom of the table.

Now that we've made sure our data is ready for use, we can begin other univariate analyses.

Summary

In this section you have studied the variable related to your research question – total GCSE score – and investigated if there was missing data. You have recoded the missing data in s1gcseptsnew so that it doesn't skew the results of further analyses. Now you're ready to move on to other univariate analyses!

***Note: as we are making changes to a dataset we'll continue using for the rest of this section, please make sure to save your changes before you close down SPSS. This will save you having to repeat sections you've already completed!**