

Graphing

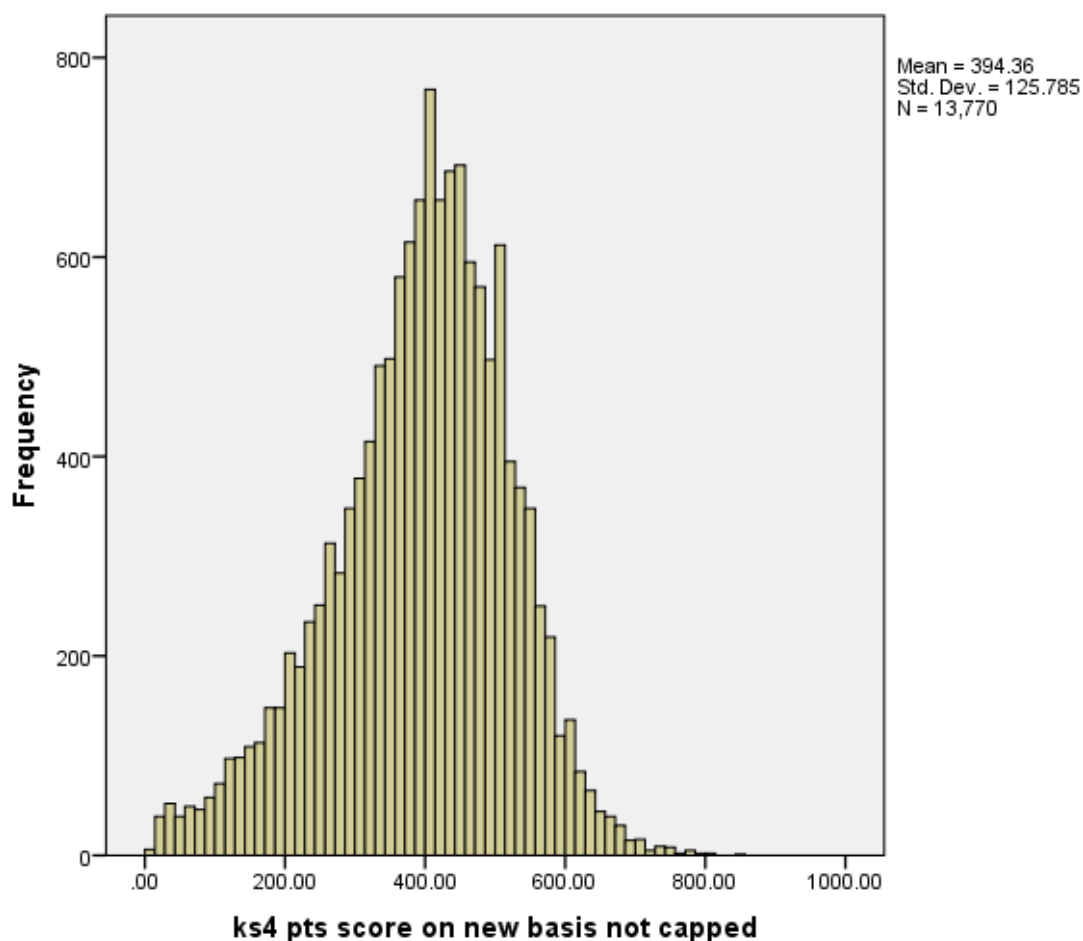
Are the s1gcseptsnew scores normally distributed?

We now know the mean value of our variable, and we can look at the distribution of values around the mean. We can check on the distribution of our variable by using SPSS to create a histogram. Histograms are a way of displaying the distribution of a continuous variable, such as the one we have.

To create a histogram, go to **Graphs, Legacy Dialogs**, and then **Histogram**.

Move **s1gcseptsnew** from the variable list on the left to the **Variable** box at the top of the dialogue box.

The SPSS Output window should open. Your histogram will look like this:



A histogram is said to display a normal distribution when the values are concentrated near the mean, and fall off symmetrically to both the left and right of the graph.

Notice that our histogram is not symmetrical, with the highest numbers of respondents near the left hand side. A distribution of this sort is called a positively skewed distribution. However most of the value points are near the mean of 394.35, with the number of data points decreasing as they move farther from the mean in either direction. This shows us that the values in our variable **s1gcseptsnew** are more likely to be towards the bottom of the range of potential values than towards the top. This is good to know because it is important that you understand your dependent variable before you analyse it.

You can create various types of graphs using the **Graph → Legacy Dialogue** function. Simply select your graph of choice (bar, line, etc.) and then move the variable you wish to illustrate to the **Category Axis** text box in the centre of the graph dialogue box. Other graphs that may be helpful to exploring a continuous variable like **s1gcseptsnew** are stem and leaf plots and box plots, which both help to illustrate the data variability.

Summary

You have studied the distribution of s1gcseptsnew in order to better understand this dependent variable before using it in statistical analyses. You've seen that the data in s1gcseptsnew is positively skewed, with most of the data trending toward the low end of the range of values. What could this imply about performance on the GCSEs?