RSC | Advancing the Chemical Sciences

(Activity to be used in conjunction with Central Limit Theorem section)

(You can either do this exercise yourself or within a group by asking each member to pick a tile sequentially.)

Cut up a piece of card into 100 small tiles and write on them the following numbers; ten 10's, ten 9's, ten 8's, ten 7's, ten 6's, ten 5's, ten 4's, ten 3's, ten 2's and ten 1's. Plot a frequency distribution of the population. It should be a rectangular shape with equal numbers of every value, hence not Gaussian. Calculate the mean and standard deviation (using a calculator).

Place the tiles into a large box and shake them around.

Draw out a batch of 4 tiles (**n** = 4), calculate the average of the four values on the tiles, record the average, replace the tiles into the box. Shake the box and repeat by drawing a fresh batch of four tiles. Repeat the drawing of batches of four tiles until you have had enough, preferably at least 50 times! Plot the average values onto a frequency distribution graph. You should see a rough Gaussian distribution. Calculate the mean and standard deviation of the batches and compare to the mean and standard deviation of the population. Are the means similar? And is the batch standard deviation smaller that the population standard deviation? Using the formula, calculate the standard error of means using the standard deviation of the population and **n** = 4. Compare the batch standard deviation to the calculated standard error of means, are they similar? What would be the effect of withdrawing larger batch sizes hence increasing the value of **n** on the accuracy of the results?

From a commercial point of view what would be the advantages and disadvantages of selecting a particular value of **n** (high or low)?