Stratified Sampling

An example using the 'real data' from *CensusAtSchool* 2004/05

When a population is divided into strata (subgroups/categories) it is important that any sample to be taken from the population is representative of the strata. For example, if the population contains three times as many Year 9s as Year 7s then the sample taken should reflect this. You must also make sure you choose a big enough sample size. So:

- separate the population into your strata;

- find out what proportion of your population is in each stratum;

- take a random sample from each stratum so the proportion of each stratum in the sample is the same as that in the population.

Example:

From 500 pupils who completed the *CensusAtSchool* questionnaire 2004/05:

73 said they lived in a city;
242 said they lived in a town;
155 said they lived in a village;
30 said 'other'.

Using a stratified sample of 50, estimate the mean height of pupils who completed the questionnaire.

The sample size must be in proportion to where the pupils live.

City: 73/500 x 50 = 7.3 (7 pupils)

Town: 242/500 x 50 = 24.2 (24 pupils)

Village: 155/500 x 50 = 15.5 (16 pupils)

Other: 30/500 x 50 = 3 (3 pupils)

You can use Excel to sort and find your random sample using the RAND() function. Then use the AVERAGE function to find the mean height for the four categories.

Stratified Sampling

The sample as shown in the Excel data file Stratified.xls calculated:

Mean height for City   = 168.3 cm
Mean height for Town   = 160.2 cm
Mean height for Village   = 163.3 cm
Mean height for Other   = 168.7 cm

Use these mean heights to find the estimated height for the population of 500 by:

$$\frac{(\text{Mean City} \times 7) + (\text{Mean Town} \times 24) + (\text{Mean Village} \times 16) + (\text{Mean Other} \times 3)}{50}$$

= 162.62 cm

This will be equal to the mean of the whole stratified sample and can be calculated by =AVERAGE(all the values in this sample). There is no need to calculate the mean of each stratum unless you need these for your analysis.

The mean height for all 500 pupils, without stratification, is easily calculated using Excel and produced the figure of 160.46 cm. This is different from the result gained by the stratified sample. It appears that pupils who live in a 'City' or 'Other' (a fifth of the total sample) are taller than the average by about 8 cm. This has probably accounted for the difference with the overall mean. This highlights the importance of having a large enough sample size. These calculations were done using real data, provided by pupils. The nature of real data often shows outliers and small samples can produce skewed results. However, the benefits are that meaningful conclusions can be drawn and good reasons for any anomalies can be easily found. It allows pupils to understand the need for refining their techniques in order that they can answer hypotheses accurately.

TASK A
1. The spreadsheet Stratified Sampling.xls contains a random sample of 500 children's data. Select a stratified sample of 50 with the following strata for the area that you live in:
   City
   Town
   Village
   Other

2. Find the mean of each strata and then the overall mean for this sample.

3. Compare this to the results in the example above.