Processing large data sets

A lot of research relies on collecting lots of data and then looking for trends in the data. In this activity students will have the chance to choose results within the large datasets to focus on, and find out if there are any links between the data.

Outcomes

Students will be able to:

- study large volumes of data and look for patterns
- put forward hypotheses based on the raw data
- choose sections of relevant data from the large volume to analyse and present it appropriately.

Time required

Allow 1 hour lesson time plus homework.

Can also be an ICT lesson.

Outline of the activity

Give all students a copy of the briefing sheet. The dataset to be used is on the Stats for schools website Smoking dataset:
http://webarchive.nationalarchives.gov.uk/2010 0406121152/http://www.stats4schools.gov.uk/la rge_datasets/smoking/default.asp

The data consists of questions asked in the survey, participant responses and compiled data. Whilst all instructions are given on the briefing sheet, students may need guidance on what they should be looking for i.e. they should be encouraged to look at trends in the data and whether they want to investigate a particular relationship in greater detail. How would they go about doing this?

Tips and strategies

Students should be allowed to work on the data in Microsoft Excel spreadsheets rather than printouts. This will allow them to convert spreadsheet data into graphs and tables easily if they wish to do so.

Processing large data sets: briefing sheet

Have a look at all the data. Is there anything interesting within the data that you would like to investigate further?

What trends in the data do you wish to look at more closely? And how are you going to do this? Make a plan of how you intend to analyse the data.

What is your hypothesis about the trends or patterns shown by the data?

Now test your hypothesis, and present the data in a visual (graph or chart).

Share your data with a classmate who analysed the same data set. Did you choose the same results to focus on or not? Discuss what you have both discovered.

Write down any points you have learned about looking at trends in large datasets.