UNIT 8 Data Handling *NC: Handling Data 2a, 2b, 2c*

		St	Ac	Ex	Sp
TOPICS (Text and Practice Books)					
8.1	*Tables and Timetables*	✓	✓	✓	✓
8.1	*Pictograms and Bar Charts*	✓	✓	-	-
8.3	*Pie Charts*	✓	✓	-	-
8.4	*Line Graphs*	✓	✓	-	-
8.5	*Questionnaires and Surveys*	✓	✓	✓	✓
8.6	*Frequency Graphs*	✓	✓	✓	✓
8.7	*Histograms with Unequal Class Intervals*	×	✓	✓	✓
8.8	*Sampling*	×	×	✓	✓
Activities (* particularly suitable for coursework tasks)					
8.1*	*Stem and Leaf Plots*	✓	✓	✓	✓
8.2	*Misuse of Statistics*	✓	✓	✓	✓
8.3*	*Vehicle Registrations*	✓	✓	✓	✓
8.4	*Testing Physical Fitness*	✓	✓	✓	✓
8.5	*Sampling Fish*	×	×	✓	✓
8.6*	*Sampling in a Village*	×	×	✓	✓
OH Slides					
8.1	*Distance Charts*	✓	✓	✓	✓
8.2	*Pie Charts*	✓	✓	-	-
8.3	*Birthdays*	✓	✓	-	-
8.4	*Biased Questions*	✓	✓	✓	✓
8.5	*Histograms*	×	✓	✓	✓
8.6	*Table of Random Numbers*	×	✓	✓	✓
Revision Tests					
8.1		✓	-	-	-
8.2		×	✓	-	-
8.3		×	×	✓	✓

UNIT 8 *Data Handling* Teaching Notes

Background and Preparatory Work

The British statesman, *Benjamin Disraeli*, once stated, "There are lies, damned lies and statistics." This might have been a bright remark to make in a debate, but the status of statistics was given a quite underrated currency which persists, even today! Disraeli's statement is in fact the reverse of the truth, and a more apt expression would be "There are lies, damned lies and numerical statements made by people ignorant of statistics!"

Statistical analysis makes possible the testing of numerical data for relevance, reliability and validity. Statisticians must present data in such a form that others can utilise the relevant information to enable them to make judgements.

The study of Statistics is reported to have started with the Englishman, *John Graunt* (1620 – 1674), who collected and studied the death records in various cities of Britain. He was fascinated by the patterns he found in the whole population.

Much of current day statistical analysis is of quite recent development, the availability of cheap computing power acting as a catalyst for the development of appropriate ways of presenting and analysing data. In fact, the more advanced statistical analyses and tests are based on probability theory, developed over the past few centuries, but put into a more modern context by mathematical statisticians such as

Karl Pearson (1857 – 1936)

Sir Ronald Fisher (1890 – 1962)

Jerzy Neyman (1894 – 1981).

You can find interesting bibliographies of these people on the internet.

For this Project we have divided statistical work into two Units, namely

Unit 8 *Data Handling*

Unit 9 *Data Analysis*

As with statistical surveys, you can divide the work involved into stages such as

- understanding the problem: postulating hypotheses
- collecting the relevant data
- presenting the data
- analysing the data
- making conclusions related to the original hypotheses.

In this Unit deal with the first three of these stages.

You might find it useful to have plenty of sources of data available for pupils to use, particularly if they will be undertaking coursework. There is an abundance of data in newspapers, particularly

For further background information, see

"How to lie with statistics"
by Darrell Huff, Pelican,
ISBN 0 14 021300 7

"Use and abuse of statistics"
by W. Reichmann, Pelican
ISBN 0 14 020707 4

"Figuring and society"
by Ronald Meek, Fontana
ISBN 0 00 632560

See web site
http://www-groups.dcs.
st-and.ac.uk:80/~history/

- financial data
- sports-related data.

You might also find it useful to use the internet to obtain various data, although the actual finding of data is an important aspect of the whole of this area of work, and at times students should be encouraged to use their ingenuity and research skills to obtain data for themselves.

See, for example, the 'data' pages on web site
http://www.ex.ac.uk/cimt/

Teaching Points

Introduction

Pupils should already be aware of the use made of data, which is to be found in abundance in newspapers, television and other media. For example, nearly all daily newspapers include statistical data on

- weather
- financial markets
- sports (football, cricket, tennis, etc.).

In some cases it might be helpful, particularly if you intend to set coursework, for you or your pupils to have saved complete sets of newspapers covering a 2- to 4-week period.

For all routes, it is important to stress that

- the optimum way of presenting the data depends on the purpose for which the data has been collected,
- the method of presentation must be fair and unbiased. (There are many examples in the press and on TV of badly presented data – sometimes, for example, as with much of recent General Election coverage on television, because too much is being attempted.)

A8.2

For the *Express/Special* route, there is an important (for statistical purposes) section on sampling (8.8). In January 1997, ITV organised a massive live discussion on the future of the monarchy – over 2 million people voted by telephone in response to the question,

"Do you support the continuation of the monarchy?",

voting *Yes* or *No*. The result (about $\frac{2}{3}$ voting *Yes* and $\frac{1}{3}$ voting *No)* was hailed as showing massive support for the monarchy. Is this a valid conclusion for the whole UK population? Statistically, the answer is a definite *No*!, since the sample who watched and voted by telephone could be very unrepresentative. There were also minor irregularities such as people registering 9 or 10 votes by telephone!

This is a very fruitful area for discussion, as are General Elections (very topical at the time this is being written), in which possible projects could include

- monitoring and checking how well the opinion polls predict the final outcome,
- undertaking your own local opinion poll,
- on the day, conducting an 'exit poll' by asking voters willing to give the information, which candidate they voted for.

OS82 and OS8.3

Some of the OH slides have been designed to help you analyse and present data for whole-class teaching. The 'Birthdays' sheet (OS8.3) can be used as a whole-class activity with each pupil putting in their data.

Language/Notation

You should try to be consistent with the language and notation used. For example, tally charts should use the notation 𝍸 |||, etc. both for accuracy and for ease of adding up.

A possible source of confusion arises when using grouped data, and deciding where the group boundaries actually lie. For example, a height (or weight) is usually given to a specified degree of accuracy, e.g. nearest cm; 172 cm actually means that the value lies in the interval

$$171.5 \text{ cm} \leq \text{height} < 172.5 \text{ cm}$$ (we conventionally round up this end point)

So heights may be grouped into 'every 5 cm' but the 170 – 174 cm grouping is in reality,

169.5 cm – 174.5 cm

and this is what should be used in the display.

To really confuse matters, the same is not true for ages where, for example, 'age 10' means the age grouping

10 – 11 years

(even someone who is age 10 years, 364 days is still recorded as age 10!). So great care must be taken, and common sense used.

Other key language used includes

- *frequency*
- *biased (data, question, sample)*
- *sample*
- *population*
- *random sample*

} *Express*/*Special* route

- *hypothesis*

Key Points

In general, the presentation of data should be

- clear
- straightforward
- unambiguous
- not biased.

For histograms of unequal width, you should use

$$\text{Frequency Density} = \frac{\text{Frequency}}{\text{Class Interval}}$$

to determine the appropriate heights, and stress that it is the *area* that represents the quantity of the data.

When designing questionnaires, it is crucially important to

- decide what you are trying to test for;
- decide who you are testing, and if your sample is a reasonable representation of the complete population under study;
- pilot the questionnaire with just a couple of people in order to discover any possible difficulties in its interpretation or implementation.

For sampling, it is important to know and understand the differences between

- random
- systematic
- stratified

samples and to appreciate that, for most purposes, a stratified sample is likely to be of most use, and to be more accurate than a simple random sample.

A8.6

OS8.4 and A8.2

Misconceptions

Apart from the obvious misconceptions that can arise from poor or biased presentation, for histograms if should be stressed that it is the area that is the measure of the quantity of data. For equal width histograms this means that the height is proportional to the frequency, but when dealing with unequal width histograms this is no longer true.

OS8.5