UNIT 9   Data Analysis   *NC: Handling Data 1b, 2c, 2e*

		St	Ac	Ex	Sp
TOPICS (Text and Practice Books)					
9.1	*Mean, Median, Mode and Range*	✓	-	-	-
9.2	*Finding the Mean from Tables and Tally Charts*	✓	✓	-	-
9.3	*Calculations with the Mean*	✓	✓	✓	-
9.4	*Mean, Median and Mode for Grouped Data*	✓	✓	✓	✓
9.5	*Cumulative Frequency*	✗	✓	✓	✓
9.6	*Standard Deviation*	✗	✗	✓	✓

Activities					
9.1	*Averages*	✓	✓	✓	✓
9.2	*Inter-quartile Range*	✓	✓	✓	✓
9.3	*Box and Whisker Plots*	-	✓	✓	✓
9.4	*National Lottery*	✗	✗	✓	✓
9.5	*Olympic Swimming Records*	✓	✓	✓	✓
9.6	*Correlation*	✓	✓	✓	✓
9.7	*Estimation of Mean Distances*	✓	✓	✓	✓
9.8	*Normal Distribution 1*	✗	✗	✓	✓
9.9	*Normal Distribution 2*	✗	✗	✓	✓
9.10	*Standardised Normal Distribution*	✗	✗	✗	✓

OH Slides					
9.1	*Mean, Median and Mode*	✓	✓	-	-
9.2	*Mean Value from Frequency Tables*	✓	✓	-	-
9.3	*Estimation of the Mean*	✓	✓	✓	✓
9.4	*Cumulative Frequency*	✗	✓	✓	✓
9.5	*Comparing Data*	✗	✓	✓	✓
9.6	*Standard Deviation*	✗	✗	✓	✓
9.7	*Normal Distribution 1*	✗	✗	✓	✓
9.8	*Normal Distribution 2*	✗	✗	✓	✓
9.9	*Normal Distribution 3*	✗	✗	✓	✓

Mental Tests					
9.1		✓	✓	-	-
9.2		✓	✓	-	-
9.3		✗	✗	✓	✓
9.4		✗	✗	✓	✓

Revision Tests					
9.1		✓	-	-	-
9.2		✗	✓	-	-
9.3		✗	✗	✓	✓

UNIT 9 *Data Analysis*   Teaching Notes

Background and Preparatory Work

This has been covered in the corresponding section in *Unit 8*. However, it is worth restating that in these two units there is an obvious link to reality and that whenever possible, real data should be used and analysed.

This unit covers some key methods of statistical analysis, including

> *Mean*, *Median*, *Mode*   Measures of *central tendency*

and

> *Standard Deviation*, *Inter-quartile Range*.   Measures of *variation*

After completion of this unit, students will be well prepared for practical coursework, except that it should be noted that formal coverage of the concepts of

> *correlation* and *regression* (line of best fit)   *Coursework* section

have been left to a later algebraic unit. In case these concepts are needed now in coursework tasks, activities have been included in order to give an introduction to these topics.   A9.5 and A9.6

Teaching Points

Introduction

Much of our analysis here deals with measures of central tendency:

> *mean*, *median* and *mode*.

Pupils should understand fully the differences between the measures but should also appreciate that these measures can *all* be classified as an 'average' for a set of data. The important point to make is that the most appropriate measure of average should be used and this depends on the context of the data being studied.

The use of *grouped data* is also an important aspect of the unit but estimating the mean, and particularly the median, may cause problems on the *Standard/Academic* routes.   OS 9.2

The idea of the *spread* or *varation* in data sets is another important concept introduced. For the *Academic* route, the *inter-quartile range* is one measure of spread about the median, whilst the *Express/Special* route also covers *standard deviation* which is another measure of spread but this time about the mean value. These measures of spread are especially important when comparing two (or more) data sets.   OS 9.5 and 9.6

With most students having scientific calculators with 'STATS' mode, it is a relatively easy task to find the mean and standard deviation of a set of data – all that is required is accuracy in inputting the data! Although it is important that you make sure that your pupils can do this (GCSE exam boards will expect their candidates to use this method), it is even more important that the topic is introduced through simple examples which can be calculated easily.

You might also like to encourage pupils to use the alternative formula

$$\text{s.d.}^2 = \frac{\sum x_i^2}{n} - \bar{x}^2$$

T 9.6

which is often easier to calculate – but its derivation is rather tedious!

Another interesting and important method of comparing data is

Box and Whisker Plots

A 9.3

which uses the median, upper and lower quartiles and the inter-quartile range. This method is outside the National Curriculum syllabus but we have included it as an optional activity.

Two other concepts which you may wish to introduce, although they are also technically outside the National Curriculum, are

• *Percentiles* – you could use *OS9.4* and *9.5* to introduce this briefly;

OS 9.4 and 9.5

• *Normal Distribution* – this is covered by using *OS9.7–9.9* and *A9.8–9.10* and is a useful extension to the work on standard deviation.

OS 9.7 – 9.9
A9.8–9.10

Neither of these topics has been covered formally in the text or practice books, although *percentiles* are mentioned in an 'information' and *normal distribution* is mentioned in a MEG exam question.

Text, bottom of page 161
Text, page 183, Ex. 14

Language / Notation

There are a number of key terms used and pupils should be familiar with their definition and use:

• mean
• median   } all measures of average
• mode

• range

• upper quartile

• lower quartile

• inter-quartile range

• cumulative frequency

• standard deviation.

Key Points

- The median is the $\left(\dfrac{n+1}{2}\right)$th data point.

- The lower quartile is the $\left(\dfrac{n+1}{4}\right)$th data point.

- The upper quartile is the $\left(\dfrac{3(n+1)}{4}\right)$th data point.

- For grouped data with large n, say $n \geq 100$, the quartiles can be approximated by the $\dfrac{n}{4}$, $\dfrac{n}{2}$ and $\dfrac{3n}{4}$ data points.

- A cumulative frequency graph is a graph of cumulative frequency (*y*-axis) against variable (*x*-axis).

- For cumulative frequency graphs of grouped data, points are plotted at the *end* of each class boundary.

- A smooth curve drawn through the points is more accurate than joining the points with straight lines.

Misconceptions

- When finding the mean value of grouped data, you must divide by the total frequency, *not* the total sum of *x*'s.

- In finding the median, some students will want to use the

$$\left(\dfrac{n}{2}\right)\text{th term,}$$

but examples using small odd values of n should clarify that what is needed is the

$$\left(\dfrac{n+1}{2}\right)\text{th term.}$$

Coursework

As mentioned earlier, students now have a range of statistical techniques which can be used to test hypotheses by collecting and analysing data and making conclusions.

A 8.3, 8.6 and 9.4

It is also possible to use techniques not directly in the National Curriculum such as

- *Stem and Leaf Plots*   A 8.1
- *Box and Whisker Plots*   A 9.3

to enhance the presentation of data using 'new' mathematical skills.

In the activities for Units 8 and 9, we have included some possible coursework starters but what is needed for a high level statistical project is to develop and implement a plan of work along the following lines:

- *Hypothesis* – the statement or conjecture to be tested.

- *Data Collection* – either 'primary' data which is actually collected by the student, or 'secondary' data which is found from other sources.

- *Data Presentation* – using the techniques of Unit 8 to present, neatly and succinctly, the main characteristics of the data.

- *Data Analysis* – using the techniques of Unit 9 (and possibly cumulative scatter diagrams, or lines of best fit (regression) to analyse the data.

- *Interpretation* – indicating what the data are illustrating, relevant to the critical hypothesis.

- *Conclusion* – summarising the evidence to reject or accept (you cannot 'prove') the hypothesis.

Here are some ideas for practical coursework which will require an analysis as outlined above.

National Lottery – which numbers do people choose most often:

    (i)   numbers in the centre of the card, or

    (ii) low numbers (0–31) which correspond to birthdays?

A source of all data concerning the lottery is to be found on the World Wide Web at

http://lottery.merseyworld.com/

Cars – how long do they last and are some cars more reliable than others?

Fitness – how much effect does weight, height, exercise, diet, smoking, drinking, etc. have on how fit a person is?

Football – does a smaller pitch (the pitch size can vary between certain constraints) encourage more or fewer goals?

Hair Strength or
Tooth Decay – does this significantly depend on sex, age, diet, hygiene, etc?

I.Q. Scores – does practice increase IQ scores?

Weather – does the South West have a warmer but wetter climate than the rest of the country?

The meteorological office for the UK has a Web site at:

   http://www.meto.govt.uk/

but a much wider range of data can be found through

   http://www.met.rdg.ac.uk/
     ~brugge/index.html

It is very important for your pupils to be interested in the topic being analysed and although having 30 or more pupils all doing different statistical projects may cause you much grief, it will be of great motivational benefit to pupils if they are investigating something they genuinely want to know more about.

An excellent source of statistics relating to track and field athletics can be found on the Web at

   http://www.uta.fi/~csmipe/
     sport/eng/mwt.html