

Chapter notes: 16 Summarising data

Overview

This chapter looks at ways of summarising and communicating aspects of both monivariate and bivariate data. It requires approximately 12 hours of teaching time. There are many opportunities to make links with real-world situations. For example, Supplementary sheets 22 'Measuring risk' and 20 'Misleading statistics' look at some ways in which statistics can be misleading.

Introductory problem

The introductory problem should encourage the understanding that what might be considered a 'small' variation depends on the spread of data. This should lead to the introduction of measures of spread. The worked solution is given at the end of the chapter, page 502; the idea being that students should be able to answer the question using the methods covered in the chapter.

16A Measures of the centre of data, p466

The calculations in this section are covered in Prior learning section Y. However, the purpose of this section is to cover slightly more difficult examination-style questions, such as questions 7 and 8.

The other types of mean mentioned in the 'Research explorer' box on page 467 are (for two data items, x and y):

$$\text{Geometric: } \sqrt{xy} \qquad \text{Harmonic: } \frac{2}{\frac{1}{x} + \frac{1}{y}} \qquad \text{Quadratic: } \sqrt{\frac{x^2 + y^2}{2}}$$

An example of a set of scores which satisfy the requirement of question 9 could be:

	Level 1	Level 2
Amy	1, 1	100
Bob	0	99, 99

Hints for the grade 7 questions:

8. There must be over 15 students with a score greater than 2 if the median is 3.

16B Measures of spread, p470

Hints for grade 7 questions:

5. The information from the median tells us that $b = 12$. The information from the mean and range set up two simultaneous equations.

16C Frequency tables and grouped data, p472

The 'Theory of knowledge issues' box on page 474 encourages students to realise that there are disadvantages to having too much detail.

16D-E

There are no specific teacher notes for these sections.

16F Constant changes to data, p489

In question 7, students should realise that the measures of spread are multiplied by $|x|$.

Hints for the grade 7 questions:

6. The new median is $20a - b$ and the new interquartile range is $10a$.

16G Correlation, p491

This is a major new topic to the Standard Level course. In examinations, it is unlikely that ambiguous values of r will appear or need to be calculated. Question 6 encourages students to think about the effect a linear relationship (such as those in section 16F) has on data.

16H Linear Regression, p497

Hints for the grade 7 questions:

6. Create a new column for x^2 , and then treat it as a new variable.