|  |  |
| --- | --- |
| **UNIT 7: Statistics, sampling and the averages** | **Teaching time**  **6-8 hours** |

[Return to Overview](#Overview)

**SPECIFICATION REFERENCES**

S1 infer properties of populations or distributions from a sample, while knowing the limitations of sampling

S2 interpret and construct tables, charts and diagrams, including frequency tables, bar charts, pie charts and pictograms for categorical data, vertical line charts for ungrouped discrete numerical data, tables and line graphs for time–series data and know their appropriate use

S4 interpret, analyse and compare the distributions of data sets from univariate empirical distributions through: …

* appropriate measures of central tendency (median, mean, mode and modal class) and spread (range, including consideration of outliers)

**PRIOR KNOWLEDGE**

Students should be able to calculate the midpoint of two numbers.

Students will have drawn the statistical diagrams in unit 3.

Students will have used inequality notation.

**KEYWORDS**

Mean, median, mode, range, average, discrete, continuous, qualitative, quantitative, data, sample, population, stem and leaf, frequency, table, sort, pie chart, estimate, primary, secondary, interval, midpoint, survey

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Specify the problem and:
* plan an investigation;
* decide what data to collect and what statistical analysis is needed;
* consider fairness;
* Recognise types of data: primary secondary, quantitative and qualitative;
* Identify which primary data they need to collect and in what format, including grouped data;
* Collect data from a variety of suitable primary and secondary sources;
* Understand how sources of data may be biased and explain why a sample may not be representative of a whole population;
* Understand sample and population.
* Calculate the mean, mode, median and range for discrete data;
* Interpret and find a range of averages as follows:
  + median, mean and range from a (discrete) frequency table;
  + range, modal class, interval containing the median, and estimate of the mean from a grouped data frequency table;
  + mode and range from a bar chart;
  + median, mode and range from stem and leaf diagrams;
  + mean from a bar chart;
* Understand that the expression 'estimate' will be used where appropriate, when finding the mean of grouped data using mid-interval values;
* Compare the mean, median, mode and range (as appropriate) of two distributions using bar charts, dual bar charts, pictograms and back-to-back stem and leaf;
* Recognise the advantages and disadvantages between measures of average.

**POSSIBLE SUCCESS CRITERIA**

Explain why a sample may not be representative of a whole population.

Carry out a statistical investigation of their own and justify how sources of bias have been eliminated.

Show me an example of a situation in which biased data would result.

State the median, mode, mean and range from a small data set.

Extract the averages from a stem and leaf diagram.

Estimate the mean from a table.

**OPPORTUNITIES FOR REASONING/PROBLEM SOLVING**

When using a sample of a population to solve contextual problem, students should be able to justify why the sample may not be representative of the whole population.

Students should be able to provide a correct solution as a counter-argument to statements involving the “averages”, e.g. Susan states that the median is 15, she is wrong. Explain why.

Given the mean, median and mode of five positive whole numbers, can you find the numbers?

**COMMON MISCONCEPTIONS**

The concept of an unbiased sample is difficult for some students to understand.

Often the ∑(*m* × *f*) is divided by the number of classes rather than ∑*f* when estimating the mean.

**NOTES**

Emphasise the difference between primary and secondary sources and remind students about the different between discrete and continuous data.

Discuss sample size and mention that a census is the whole population (the UK census takes place every 10 years in a year ending with a 1 – the next one is due in 2021).

Specify the problem and planning for data collection is not included in the programme of study but is a perquisite to understand the context of the topic.

Writing a questionnaire is not part of the new specification, but is a good topic to demonstrate bias and ways to reduce bias in terms of timing, location and question types that can introduce bias.

Encourage students to cross out the midpoints of each group once they have used these numbers to in *m* × *f*. This helps students to avoid summing *m* instead of *f*.

Remind students how to find the midpoint of two numbers.

Emphasise that continuous data is measured, i.e. length, weight, and discrete data can be counted, i.e. number of shoes.

When comparing the mean and range of two distributions support with ‘copy and complete’ sentences, or suggested wording.