UNIT 14: Statistics and sampling, cumulative frequency and histograms

SPECIFICATION REFERENCES

S1 infer properties of populations or distributions from a sample, whilst knowing the limitations of sampling apply statistics to describe a population

S3 **interpret and construct diagrams for grouped discrete data and continuous data, i.e. histograms with equal and unequal class intervals and cumulative frequency graphs, and know their appropriate use**

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

- Appropriate graphical representation involving discrete, continuous and grouped data, **including box plots**
- appropriate measures of central tendency (median, mean, mode and modal class) and spread (range, including consideration of outliers, **quartiles and inter-quartile range)**

S5 apply statistics to describe a population

PRIOR KNOWLEDGE

Students should understand the different types of data: discrete/continuous.

Students should have experience of inequality notation.

Students should be able to multiply a fraction by a number.

Students should understand the data handling cycle.

KEYWORDS

Tier 2

Sample, population, bias, random, frequency, mean, compare, spread, range

Tier 3

Fraction, decimal, percentage, stratified sample, cumulative frequency, box plot, histogram, frequency density, median, mode, lower quartile, upper quartile, interquartile range, outlier

14a. Collecting data (S1)	**Teaching time** 5–7 hours

OBJECTIVES

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

- Specify the problem and plan:
  - decide what data to collect and what analysis is needed;
  - understand primary and secondary data sources;
  - consider fairness;
- Understand what is meant by a sample and a population;
- Understand how different sample sizes may affect the reliability of conclusions drawn;
- Identify possible sources of bias and plan to minimise it;
- Write questions to eliminate bias, and understand how the timing and location of a survey can ensure a sample is representative (see note);

POSSIBLE SUCCESS CRITERIA/EXAM QUESTIONS

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

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

There are 1200 students at a school.

Kate is helping to organise a party.
She is going to order a pizza.

Kate takes a sample of 60 of the students at the school.
She asks each student to tell her **one** type of pizza they want.

The table shows information about her results.

Pizza	Number of students
ham	20
salami	15
vegetarian	8
margherita	17

Work out how much ham pizza Kate should order.
Write down any assumption you make **and** explain how this could affect your answer.

(Total 3 marks)

New SAMs Paper 1F qu.27 / 1H qu.7 (S1 – AO1/AO3)

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 the whole population.

NOTES

Emphasise the difference between primary and secondary sources and remind students about the difference 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).

Specifying the problem and planning for data collection is not included in the programme of study, but is a prerequisite to understanding the context of the topic.

Writing a questionnaire is also not included in the programme of study, but remains a good topic for demonstrating bias and ways to reduce bias in terms of timing, location and question types.

UNIT 3: Averages and range, collecting data, representing data

SPECIFICATION REFERENCES

G14 use standard units of measure and related concepts (length, area, volume/capacity, mass, time, money, etc.)

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

S3 **construct and interpret diagrams for grouped discrete data and continuous data i.e. histograms with equal and unequal class intervals …**

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

- appropriate graphical representation involving discrete, continuous and grouped data …
- appropriate measures of central tendency (median, mode and modal class) and spread (range, including consideration of outliers) …

S5 apply statistics to describe a population

S6 use and interpret scatter graphs of bivariate data; recognise correlation and know that it does not indicate causation; draw estimated lines of best fit; make predictions; interpolate and extrapolate apparent trends whilst knowing the dangers of so doing

PRIOR KNOWLEDGE

Students should be able to read scales on graphs, draw circles, measure angles and plot coordinates in the first quadrant.

Students should have experience of tally charts.

Students will have used inequality notation.

Students must be able to find midpoint of two numbers.

KEYWORDS

Tier 2

Mean, range, discrete, positive, negative, sample, population, frequency, estimate, table, sort

Tier 3

Median, mode, average, continuous, qualitative, quantitative, data, scatter graph, line of best fit, correlation, stem and leaf, pie chart

3a. Averages and range (G14, S2, S3, S4, S5)	**Teaching time** 6–8 hours

OBJECTIVES

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

- Design and use two-way tables for discrete and grouped data;
- Use information provided to complete a two-way table;
- Sort, classify and tabulate data and discrete or continuous quantitative data;
- Calculate mean and range, find median and mode from small data set;
- Use a spreadsheet to calculate mean and range, and find median and mode;
- Recognise the advantages and disadvantages between measures of average;
- Construct and interpret stem and leaf diagrams (including back-to-back diagrams):
  - find the mode, median, range, as well as the greatest and least values from stem and leaf diagrams, and compare two distributions from stem and leaf diagrams (mode, median, range);
- Calculate the mean, mode, median and range from a frequency table (discrete data);
- Construct and interpret grouped frequency tables for continuous data:
  - for grouped data, find the interval which contains the median and the modal class;
  - estimate the mean with grouped data;
  - understand that the expression 'estimate' will be used where appropriate, when finding the mean of grouped data using mid-interval values.

POSSIBLE SUCCESS CRITERIA/EXAM QUESTIONS

Be able to 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.

The table shows some information about the foot lengths of 40 adults.

Foot length (f cm)	Number of adults
$16 \leqslant f < 18$	3
$18 \leqslant f < 20$	6
$20 \leqslant f < 22$	10
$22 \leqslant f < 24$	12
$24 \leqslant f < 26$	9

(*a*) Write down the modal class interval.

(1)

(*b*) Calculate an estimate for the mean foot length.

(3)

(Total 4 marks)

New SAMs Paper 2F qu.27 / 2H qu.5 (S4 – AO1/AO2)

OPPORTUNITIES FOR REASONING/PROBLEM SOLVING

Students should be able to provide reasons for choosing to use a specific average to support a point of view.

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

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.

COMMON MISCONCEPTIONS

Students often forget the difference between continuous and discrete data.

Often the $\sum(m \times f)$ is divided by the number of classes rather than $\sum f$ when estimating the mean.

NOTES

Encourage students to cross out the midpoints of each group once they have used these numbers to in $m \times f$. 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.

Designing and using data collection is no longer in the specification, but may remain a useful topic as part of the overall data handling process.