MathsHUBS Resources

National Centre for Excellence in the Teaching of Mathematics

Mastery Professional Development

5 Statistics and probability

5.2 Statistical analysis

Guidance document | Key Stage 3

Making connections

The NCETM has identified a set of six 'mathematical themes' within Key Stage 3 mathematics that bring together a group of 'core concepts'.

The fifth of these themes is *Statistics and Probability*, which covers the following interconnected core concepts:

5.1 Statistical representations and measures

5.2 **Statistical analysis**

5.3 Probability

This guidance document breaks down core concept *5.2 Statistical analysis* into two statements of knowledge, skills and understanding:

5.2.1 Interpret reasonably statistical measures and representations

5.2.2 Choose appropriately statistical measures and representations

Then, for each of these statements of knowledge, skills and understanding we offer a set of key ideas to help guide teacher planning.

Please note that these materials are principally for professional development purposes. Unlike a textbook scheme they are not designed to be directly lifted and used as teaching materials. The materials can support teachers to develop their subject and pedagogical knowledge and so help to improve mathematics teaching in combination with other high-quality resources, such as textbooks.

Overview

When students worked with statistics at Key Stage 2, they chose the most appropriate representations for data and explained the reasons for their choice. They interpreted and constructed pie charts and line graphs to solve problems and may have also encountered and drawn graphs relating two variables. Students will have had experience of calculating the mean as an average and will know when it is appropriate to find the mean of a data set.

A key skill for students to develop at Key Stage 3 is the ability to make an informed choice of what statistical analysis and representation to use for discrete, continuous and grouped data. Being able to construct representations and calculate values that indicate a measure of central tendency or measure of spread is important in order to represent and summarise data accurately. Just as important is the need for students to be able to make an informed choice about what statistical tools to use, and understand the effect that these choices have on the interpretation, and misinterpretation, of data, including the potential impact of outliers.

The use of real-life examples will play an important part in developing students' understanding of the various ways that statistics can be used to represent data sets. Analysis of authentic statistics will also help students to appreciate the meaning and significance of different ways of presenting data.

Students should understand the importance of having both a measure of central tendency (mean, median and mode) and a measure of spread (range, including a consideration of outliers) in order to appreciate the distribution of a set of data. Students should be presented with summary data to interrogate, so they can appreciate the limitations of such information when the raw data is no longer available. Students can use spreadsheets and other technologies in their interrogation of statistics, and analyse large and more complex data sets. Dealing with inaccuracies, outliers and other contextual issues will give students a greater appreciation of the realistic nature of statistical analysis.

Additionally, students should have opportunities to describe simple mathematical relationships between two variables (bivariate data) in observational and experimental contexts, and to illustrate such relationships using scatter graphs. This will be developed further in Key Stage 4, alongside more sophisticated measures of central tendency (including modal class) and spread (including quartiles and inter-quartile range).

Prior learning

Before beginning to teach *Statistical analysis* at Key Stage 3, students should already have a secure understanding of the following from previous study:

Key stage	Learning outcome
Upper Key Stage 2	• Interpret and present discrete and continuous data using appropriate graphical methods, including bar charts and time graphs • Solve comparison, sum and difference problems using information presented in bar charts, pictograms, tables and other graphs • Solve comparison, sum and difference problems using information presented in a line graph • Complete, read and interpret information in tables, including timetables • Interpret and construct pie charts and line graphs and use these to solve problems • Calculate and interpret the mean as an average
Key Stage 3	• 5.1.1 Understand and calculate accurately measures of central tendency and spread • 5.1.2 Construct accurately statistical representations **Please note:** Numerical codes refer to statements of knowledge, skills and understanding in the NCETM breakdown of Key Stage 3 mathematics.

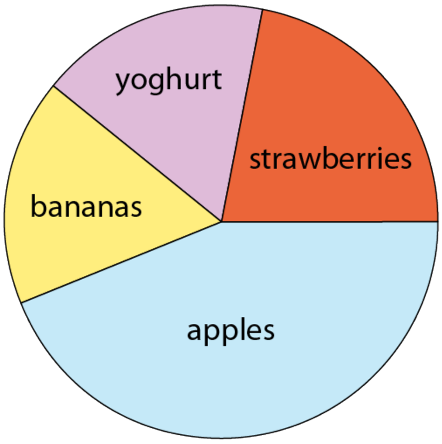
You may find it useful to speak to your partner schools to see how the above has been covered and the language used.

You can find further details regarding prior learning in the following segments of the NCETM primary mastery professional development materials[1]:

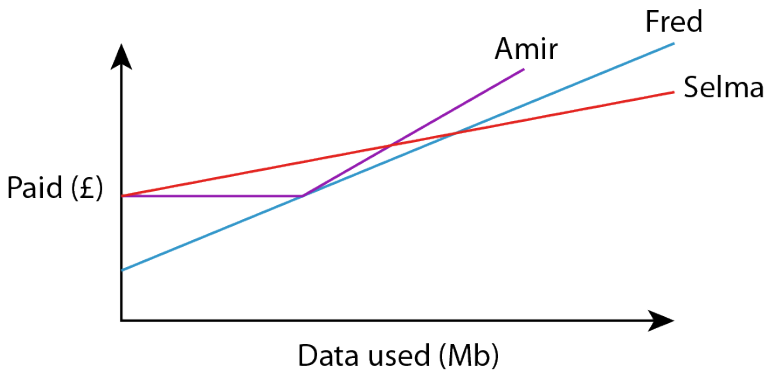
- Year 2: 1.12 Subtraction as difference
- Year 3: 1.17 Composition and calculation: 100 and bridging 100
- Year 5: 1.27 Negative numbers: counting, comparing and calculating
- Year 6: 2.26 Mean average and equal shares

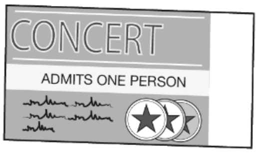
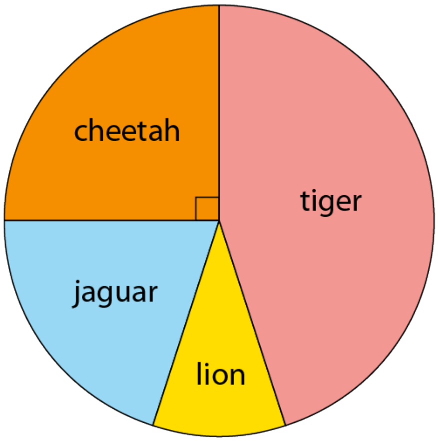
Checking prior learning

The following activities from the NCETM primary assessment materials[2] and the Standards & Testing Agency's past mathematics papers[3] offer useful ideas for assessment, which you can use in your classes to check whether prior learning is secure:

<table>
<tr><th>Reference</th><th>Activity</th></tr>
<tr><td>Year 6 pages 37 and 38</td><td>

The pie chart represents the proportions of the four ingredients in a smoothie drink.
The sector representing the amount of strawberries takes up 22% of the pie chart.
The sector representing the amount of apples is twice as big as the sector representing the amount of strawberries.
The sectors representing the amount of yoghurt and the amount of bananas are identical.
Calculate the percentage of bananas needed to make a smoothie drink.
What percentage of bananas would be needed to make two smoothie drinks?
Explain your reasoning. </td></tr>
<tr><td>Year 6 page 38</td><td> Ten pupils take part in some races on Sports Day, and the following times are recorded.
Time to run 100 m (seconds): 23, 21, 21, 20, 21, 22, 24, 23, 22, 20.
Time to run 100 m holding an egg and spoon (seconds): 45, 47, 49, 43, 44, 46, 78, 46, 44, 48.
Time to run 100 m in a three-legged race (seconds): 50, 83, 79, 48, 53, 52, 85, 81, 49, 84.
Calculate the mean average of the times recorded in each race.
For each race, do you think that the mean average of the times would give a useful summary of the ten individual times?
Explain your decision. </td></tr>
</table>

Year 6 page 39	*Three mobile phone companies each have different monthly pay-as-you-go contracts.* *Phil's Phones: £5 fee every month and 2p for each Mb of data you use.* *Manish's Mobiles: £7 fee every month and 1p for each Mb of data you use.* *Harry's Handsets: £7 fee every month and 200 Mb of free data, then 3p for each Mb of data after that.* *Amir, Selma and Fred have mobile phones and they have recorded for one month how much data they have used (in Mb) and how much they have paid (in £). They have represented their data on this graph.*  *With which company do you think Amir has his contract?* *With which company do you think Selma has her contract?* *With which company do you think Fred has his contract?* *Explain each of your choices.*

2018 Key Stage 2 Mathematics paper 2: reasoning Question 18	*Last year, Jacob went to four concerts.* *Three of his tickets cost £5 each.*  *The other ticket cost £7.*  *What was the* ***mean*** *cost of the tickets?* *Show your method.* Source: *Standards & Testing Agency* Public sector information licensed under the Open Government Licence v3.0
2018 Key Stage 2 Mathematics paper 3: reasoning Question 6	*This chart shows the number of different types of big cat in a zoo.* *There are* ***20*** *big cats in the zoo altogether.*  *Here are some statements about the chart.* *Tick the statements that are* ***true****.* *There are more cheetahs than jaguars.* ☐ *The total number of lions and tigers is 10.* ☐ *One-quarter of the big cats are cheetahs.* ☐ *There are more than 5 jaguars.* ☐ Source: *Standards & Testing Agency* Public sector information licensed under the Open Government Licence v3.0

Key vocabulary

Term	Definition
(arithmetic) mean	The sum of a set of numbers, or quantities, divided by the number of terms in the set. Example: The arithmetic mean of 5, 6, 14, 15 and 45 is (5 + 6 + 14 + 15 + 45) ÷ 5 i.e. 17.
bivariate	Involving two random variables; used in statistics as a bivariate distribution.
dispersion	Dispersion (also called 'variability', 'scatter' or 'spread') is the extent to which the data in a distribution is spread out. A simple measure of spread is the range. Other common measures are the variance, standard deviation and interquartile range.
measure of central tendency	In statistics, a measure of how the values of a particular variable are located in terms of the values collected for a particular sample, or for the relevant population as a whole. In school mathematics up to Key Stage 4, there are three important measures of central tendency: the arithmetic mean, the median and the mode. These are all statistical averages and often one is more useful than another, depending on the spread of the values under consideration.
median	The middle number or value when all values in a set of data are arranged in ascending order. Example: The median of 5, 6, 14, 15 and 45 is 14. When there is an even number of values, the arithmetic mean of the two middle values is calculated. Example: The median of 5, 6, 7, 8, 14 and 45 is (7 + 8) ÷ 2, i.e. 7.5 The median is one example of an average.
mode	The most commonly occurring value or class with the largest frequency. Example: The mode of this set of data: 2, 3, 3, 3, 4, 4, 5, 5, 6, 7, 8 is 3. Some sets of data may have more than one mode.
outlier	In statistical samples, an outlier is an exceptional trial result that lies beyond where most of the results are clustered. Example: Six people have the following salaries: £20 000, £25 600, £2 000, £19 000, £30 000, £160 000. The salary of £160 000 is clearly out of line with the others and is an outlier. At the other end, £2 000 is also well below the central cluster of values and so may also be considered as an outlier.
range	A measure of spread in statistics. The difference between the greatest value and the least value in a set of numerical data.

scatter graph	A graph on which paired observations are plotted and which may indicate a relationship between the variables. Example: The heights of a number of people could be plotted against their arm span measurements. If height is roughly related to arm span, the points that are plotted will tend to lie along a line.

Collaborative planning

Below we break down each of the two statements within *Statistical analysis* into a set of key ideas to support more detailed discussion and planning within your department. You may choose to break them down differently depending on the needs of your students and timetabling; however, we hope that our suggestions help you and your colleagues to focus your teaching on the key points and avoid conflating too many ideas.

Please note: We make no suggestion that each key idea represents a lesson. Rather, the 'fine-grained' distinctions we offer are intended to help you think about the learning journey irrespective of the number of lessons taught. Not all key ideas are equal in length and the amount of classroom time required for them to be mastered will vary, but each is a noteworthy contribution to the statement of knowledge, skill and understanding with which it is associated.

The following letters draw attention to particular features:

D Suggested opportunities for **deepening** students' understanding through encouraging mathematical thinking.

L Examples of shared use of **language** that can help students to understand the structure of the mathematics. For example, sentences that all students might say together and be encouraged to use individually in their talk and their thinking to support their understanding (for example, *'The smaller the denominator, the bigger the fraction.'*).

R Suggestions for use of **representations** that support students in developing conceptual understanding as well as procedural fluency.

V Examples of the use of **variation** to draw students' attention to the important points and help them to see the mathematical structures and relationships.

PD Suggestions of questions and prompts that you can use to support a **professional development** session.

For selected key ideas, marked with an asterisk (*), we exemplify the common difficulties and misconceptions that students might have and include elements of what teaching for mastery may look like. We provide examples of possible student tasks and teaching approaches, together with suggestions and prompts to support professional development and collaborative planning. You can find these at the end of the set of key ideas.

Key ideas

5.2.1 Interpret reasonably statistical measures and representations

The measures of central tendency and spread will each be most appropriate and meaningful in different situations. By experiencing data sets arising in varying contexts, and comparing and contrasting ways of analysing and representing them, students can be encouraged to explain, justify and be critical of their choice of measure. These skills continue to develop at Key Stage 3, and students should now be made aware of the potential impact of an outlier on measures of central tendency and spread.

Students should appreciate the differences between a frequency-based chart (such as a bar chart or pictogram) and a proportional chart (such as a pie chart) and how different aspects of the data can, and cannot, be inferred from each.

Students should also be able to 'describe simple mathematical relationships between two variables (bivariate data) in observational and experimental contexts and illustrate using scatter graphs' (Department for Education, 2013)[†]. In Key Stage 4, students will secure and deepen this understanding by exploring correlation and causation, using estimated lines of best fit to make predictions and interpolating and extrapolating apparent trends while being aware of the risks of doing so.

5.2.1.1 Understand that the different measures of central tendency offer a summary of a set of data

5.2.1.2 Understand how certain statistical measures may change as a result of changes in data

5.2.1.3 Understand range as a measure of spread, including a consideration of outliers

5.2.1.4 Understand that the different statistical representations offer different insights into a set of data

5.2.1.5* Use the different measures of central tendency and spread to compare two sets of data

5.2.1.6 Use the different statistical representations to compare two sets of data

5.2.1.7 Recognise relationships between bivariate data represented on a scatter graph

5.2.2 Choose appropriately statistical measures and representations

Situations that require statistical techniques to be employed will probably begin with an issue, a question or a problem. For example, *'Which was the wettest month this year?'* or *'What different flavours of crisps (and how many packets) should we order for the school tuck shop each month?'*. These situations require a number of decisions to be made:

- *Which data do I need to collect?*
- *How will I organise this data?*
- *How will I analyse this data in order to address the original question/problem?*
- *Which representation should I choose in order to address the original question/problem and communicate clearly my findings?*

Students should have the opportunity to consider all of these aspects at different stages of their work on statistics. The use of real-life contexts and issues are important to give statistics meaning.

Students should be given opportunities to solve problems that do not necessarily have a correct answer, but be required to justify decisions made and be prepared to be challenged. As statistical

[†] Department for Education, 2013, *National curriculum in England: mathematics programmes of study, Key Stage 3*

problems often involve prediction of trends and forecasting, probability could be linked with this element of problem-solving so students begin to gain an understanding of confidence and statistical significance, which is more fully developed in Key Stage 4 and beyond.

5.2.2.1 Given a statistical problem, choose what data needs to be analysed to explore that problem

5.2.2.2* Given a statistical problem, choose appropriate statistical measures to explore that problem

5.2.2.3 Given a statistical problem, choose appropriate representations to explore that problem

5.2.2.4 Given a statistical problem, choose appropriate measures and representations to effectively summarise and communicate conclusions

Exemplified key ideas

5.2.1.5 Use the different measures of central tendency and spread to compare two sets of data

Common difficulties and misconceptions

Students often learn to calculate the range alongside the different averages and do not always understand the distinction between the range as a measure of spread and the averages as a measure of central tendency.

Students should be encouraged to interpret summary values and draw conclusions from them rather than just being able to mechanically calculate them. Reversing the process and asking students to construct a data set to match a set of summary statistics gives students a much deeper understanding of the summary values.

R Asking students to match a data set, the summary data and an associated graphical representation can be a powerful activity to support students with making connections and going beyond mechanically calculating measures of central tendency. This is exemplified further in *Example 4* below.

What students need to understand

Accurately compare two data sets using one or more summary statistic.

Example 1:

The daily temperatures across March last year for two cities are summarised in this table.

City	Mean maximum daily temperature	Range of maximum daily temperature
A	22 °C	6 °C
B	22 °C	13 °C

Which city should you choose to visit if you want to enjoy high temperatures?

Justify your answer.

Guidance, discussion points and prompts

V *Example 1* is designed to draw students' attention to the fact that range is a basic measure of spread and not central tendency. Students should realise that the smaller the range, the more consistent the data.

PD What examples of data sets with summary statistics do students currently experience? How might you ensure that students experience a variety of measures of central tendency, as well as data sets including outliers?

Example 2:

The daily temperatures across March last year for two cities are summarised in this table.

City	Mean maximum daily temperature	Range of maximum daily temperature
C	12 °C	8 °C
D	21 °C	8 °C

Which city should you choose to visit if you want to enjoy high temperatures?

Justify your answer.

V *Example 2* is designed to draw students' attention to the fact that mean is a measure of central tendency and not the spread of the data. Students should realise that the smaller the mean, the colder the average daily temperature.

Example 3:

The daily temperatures last week for two other cities are summarised in this table.

City	Mean maximum daily temperature	Modal maximum daily temperature
E	24 °C	18 °C
F	20 °C	23 °C

How would you argue that city E enjoyed warmer days last week? How could you argue that city F did?

Example 3 draws students' attention to the subtle differences in averages. City E has a higher mean maximum daily temperature than city F, but that could be due to one extremely warm day (an outlier). A comparison of the modes might suggest that city F enjoyed warmer days than city E.

D A challenging, but worthwhile, activity to develop understanding, is to ask students to construct some possible data sets for each city to help them discern which city might have been the warmer. In this case, the context of the question would prevent *very* large values for an outlier.

For example:

City	Maximum daily temperature in °C						
E	18	18	18	24	29	30	31
F	16	18	19	20	21	23	23

Or

City	Maximum daily temperature in °C						
E	18	18	20	23	24	25	40
F	16	18	19	20	21	23	23

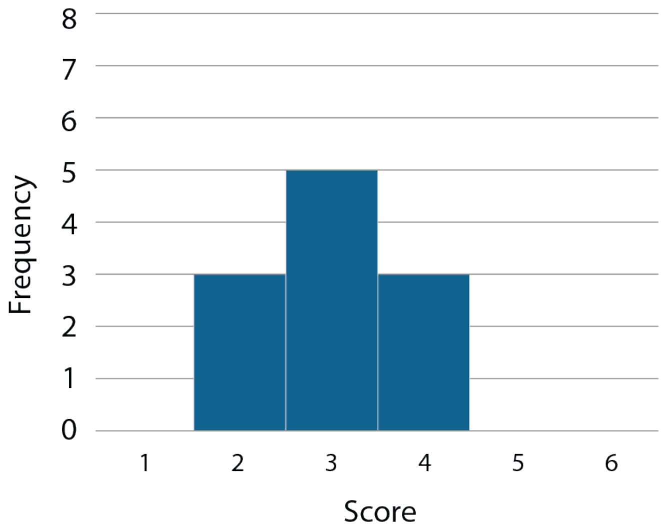
Understand that data sets can have the same measures of central tendency but a different measure of spread.

Example 4:

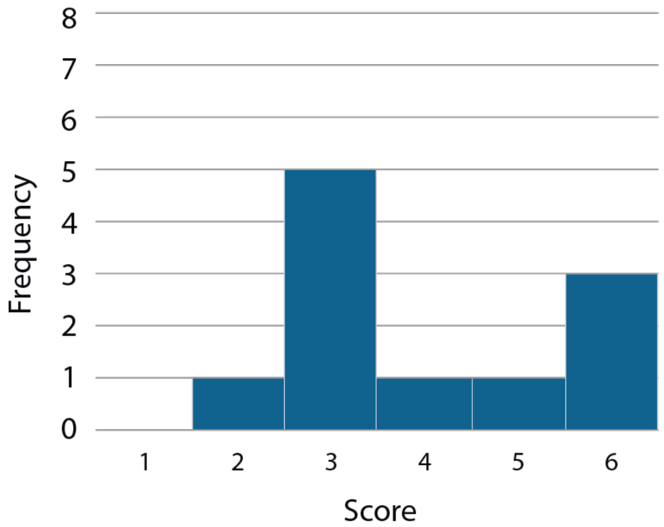
a) *Match the charts with the correct measures of central tendency and spread.*

b) *Find the missing values.*

Bar chart A



Bar chart B

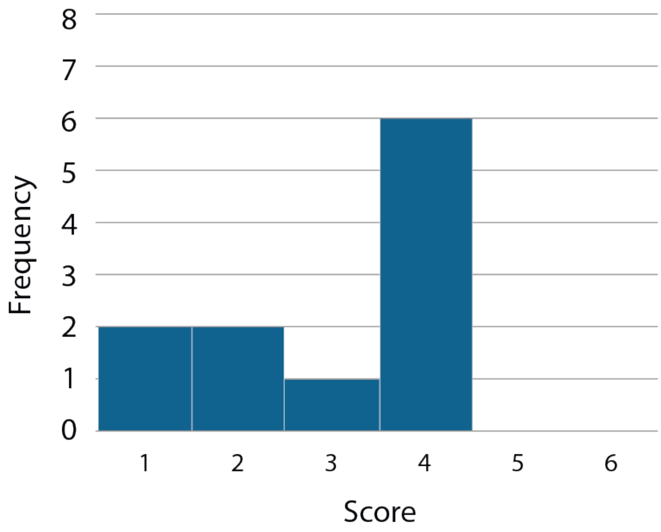


V The choice of what to vary and what not to vary can draw students' attention to the key ideas. With *Example 4*, you could ask students questions, such as: *'Looking at the set of charts and tables, what do you notice? What's the same? What's different?'*

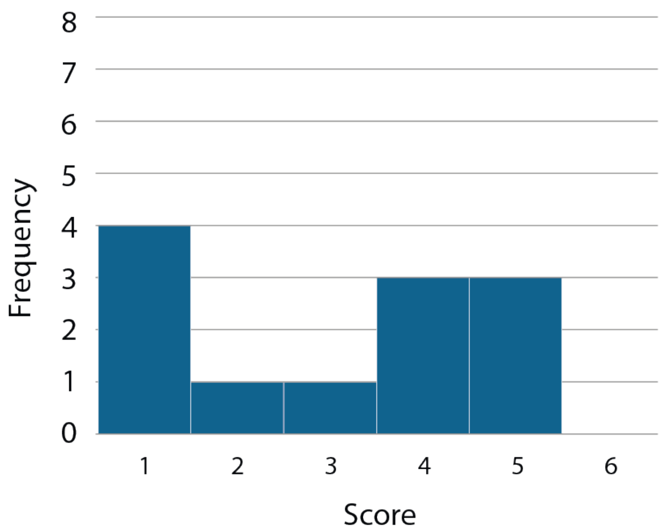
D Providing the statistical measures and asking students to match the tables with the bar charts actively encourages them to focus on the meaning of these measures and to use reasoning and problem-solving skills, thus deepening their understanding.

PD What strategies would you expect students to use when solving *Example 4*? You could consider asking students questions, such as: *'Is there just one solution to the matching activities?'*

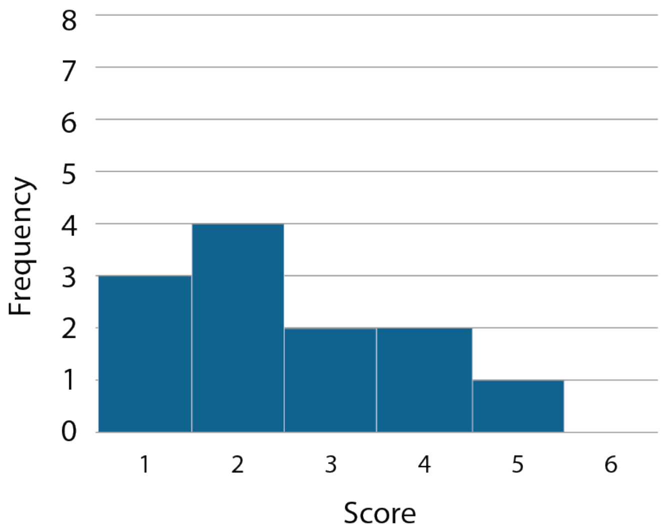
Bar chart C



Bar chart D



Bar chart E



Stats A

Mean	3
Median	4
Mode	4
Range	3

Stats B

Mean	3
Median	3
Mode	3
Range	

Stats C

Mean	
Median	3
Mode	3
Range	4

Stats D

Mean	
Median	2
Mode	2
Range	4

Stats E

Mean	3
Median	
Mode	1
Range	4

Source: Adapted from *Improving Learning in Mathematics (The Standards Unit box) – S4 Understanding mean, median, mode and range*
DfES Standards Unit

Understand range is a measure of spread and not a measure of central tendency.

Example 5:

The test scores for two classes are summarised in this table.

	Class A	Class B
Mean	70	70
Mode	70	70
Median	70	70
Range	15	10

Archie thinks some students in class A have scored higher than class B because it has a larger range.

Do you agree? Explain your answer.

V *Example 5* is designed to encourage students to notice that the range is a measure of spread and not a measure of central tendency. For example, class B could have a maximum value of 75 and minimum value of 65, whereas class A could have a maximum value of 73 and minimum value of 58.

PD How would you support students who were struggling with this question?

Solve familiar and unfamiliar problems, including real-life applications.

Example 6:

Three students each run five speed trials. These are the times (in seconds) for their five runs.

Abdul	Beattie	Chris
92	99	84
86	79	92
89	80	90
87	96	92
91	96	97

Who is the fastest runner?

Can you provide reasons (with evidence) for why each student might be the fastest?

D *Example 6* provides students with scope to make their own choices and to discuss them. In addition to using averages and range, the maximum and minimum data values might also be justifiable measures in this context.

PD Devise other problems, in contexts suitable for your classes, where the need to use the different measures of central tendency and spread to compare two sets of data is relevant.

Solve problems where there is more than one answer and where there are elements of experimentation, investigation, checking, reasoning, proof, etc. *Example 7:* *Is this statement always, sometimes or never true?* *'The measure of spread of a data set is greater than any of the measures of central tendency of the data set.'* *Explain your answer.*	Students tackling this question will need to explore a range of possibilities. Constructing some possible data sets to illustrate when it is true and when it is not true can be helpful.
Example 8: *Is this statement always, sometimes or never true?* *'Two data sets with the same measure of spread and measures of central tendency have the same maximum and minimum values.'* *Explain your answer.*	**D** For problems such as *Examples 7* and *8*, students could be encouraged to present a convincing argument or proof.

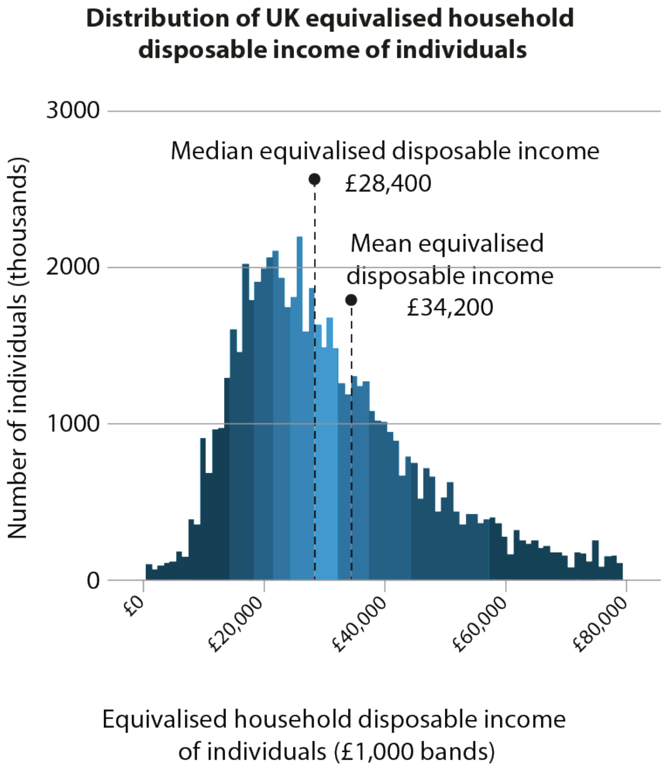
5.2.2.2 Given a statistical problem, choose appropriate statistical measures to explore that problem

Common difficulties and misconceptions
Students may think that all statistical measures are equally valid and do not appreciate the subtle differences between them. For example, students may see the mode as too simplistic and not understand that it can be used for qualitative data, whereas mean and median cannot. Students need to understand that the mean is a way of 'evening out' the data (keeping the total the same but distributing everything evenly), but does not give a good measure of central tendency when there are extreme outliers or the data set is very skewed. The median is useful if the data is evenly spaced, but may give a distorted view of the data if not. Students often confuse the range with measures of central tendency. Representing the data as points on a number line can help students to have a visual sense of the range and so more easily distinguish it from a measure of central tendency. When considering data sets with extreme outliers, it is helpful to ask students to consider what might be a more appropriate way of measuring how dispersed the data set is by ignoring the extremes and trying to measure the spread of the middle portion. This can provide a foundation for later study involving inter-quartile range and box plots in Key Stage 4.

What students need to understand	Guidance, discussion points and prompts
Explain why a certain average is the best choice of measure of central tendency. *Example 1:* *Consider the following data sets on property for a given geographical area:* • *Average type of property* • *Average number of bedrooms* • *Average property price* *For which data set would the mode be an appropriate average to use?* *For which data set would the mean be an appropriate average to use?*	**V** *Example 1* is designed so that two of the examples are quantitative and one qualitative to encourage students to notice that the mode is the only average that may be possible with qualitative data and that the mean is usually the best measure of central tendency to use when the data distribution is continuous and symmetrical. **PD** Students are often presented with questions about the mode, the mean and the median simultaneously. To what extent does this contribute to students thinking that they should always calculate all three? Consider strategies to avoid this misapprehension. What are the advantages and disadvantages of working on the three measures separately and simultaneously?

Example 2:

This graph is produced by the Office for National Statistics. It shows the distribution of UK household disposable income for the financial year ending 2018.



Source*: Office for National Statistics*
Public sector information licensed under the
Open Government Licence v3.0

When comparing household disposable incomes, why is the median often preferred over the mean to represent the 'typical' income?

V *Example 2* is designed to encourage students to notice that the median is an appropriate measure when the data is skewed or distorted by outliers.

(Note that the mode falls in the £26 000 to £27 000 bracket.)

PD Real contexts are often used in statistics, some of which are more accessible and relevant for students than others. What data sets would be relevant for your class, for which the median would be an appropriate average to use?

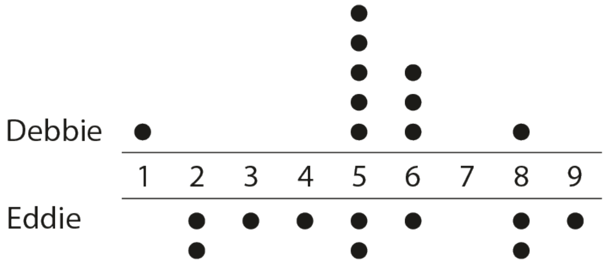
Example 3:

What would be the best average to choose to represent the height of a 'typical' girl?

D *Example 3* encourages students to think more deeply about the meaning of the three averages (mean, median and mode). Discussion can be prompted by questions such as:

- *'Where do you think the modal height might be on this graph? Why?'*
- *'Where do you think the median is? Why?'*
- *'Can you say anything about the mean in this sort of distribution? Explain your reasoning.'*

Note, in a normal distribution like this, the mode, median and mean all have the same value.

Explain why certain statistical measures give a better measure of dispersion. *Example 4:* *Two friends play a darts game and record their scores.* *Debbie: 5, 1, 5, 6, 6, 5, 6, 5, 8, 5* *Eddie: 4, 8, 3, 5, 6, 8, 9, 2, 2, 5* *Fatima calculates the range for each person in order to compare who is the most consistent player. She argues that, in this case, the range alone is not sufficient in considering who is the most consistent player.* *Do you think that she is right? Explain your answer.*	In *Example 4* both data sets have a range of seven. They also have the same mean of 5.2. **R** A dot plot representation provides more information on what Fatima has perhaps noticed:  Debbie seems more consistent because only her score of one was an extreme outlier and not representative of her set of scores.

Weblinks

1 NCETM primary mastery professional development materials
https://www.ncetm.org.uk/resources/50639

2 NCETM primary assessment materials
https://www.ncetm.org.uk/resources/46689

3 Standards & Testing Agency past mathematics papers
https://www.gov.uk/government/collections/national-curriculum-assessments-practice-materials#key-stage-2-past-papers