# Teacher Delivery Guide Statistics: 2.02 Data Presentation and Interpretation

| **OCR****Ref.** | **Subject Content** | **Stage 1 learners should…** | **Stage 2 learners additionally should…** | **DfE Ref.** |
| --- | --- | --- | --- | --- |
| **2.02 Data Presentation and Interpretation** |
| 2.02a2.02b | Single-variable data | a) Be able to interpret tables and diagrams for single variable data.*e.g. vertical line charts, dot plots, bar charts, stem-and-leaf diagrams, box-and-whisker plots, cumulative frequency diagrams and histograms (with either equal or unequal class intervals).**Includes non-standard representations.*b) Understand that area in a histogram represents frequency.*Includes* *the link between histograms and probability distributions.* *Includes understanding, in context, the advantages and disadvantages of different statistical diagrams.* |  | ML1 |
| 2.02c2.02d2.02e | Bivariate data | c) Be able to interpret scatter diagrams and regression lines for bivariate data, including recognition of scatter diagrams which include distinct sections of the population.*Learners may be asked to add to diagrams in order to interpret data, but not to draw complete scatter diagrams.*[*Calculation of equations of regression lines is excluded.*]d) Be able to understand informal interpretation of correlation.e) Be able to understand that correlation does not imply causation. |  | ML2 |
| 2.02f | Measures of average and spread | f) Be able to calculate and interpret measures of central tendency and variation, including mean, median, mode, percentile, quartile, interquartile range, standard deviation and variance.*Includes understanding that standard deviation is the root mean square deviation from the mean.**Includes using the mean and standard deviation to compare distributions.* |  | ML3 |
| 2.02g | Calculations of mean and standard deviation | g) Be able to calculate mean and standard deviation from a list of data, from summary statistics or from a frequency distribution, using calculator statistical functions.*Includes understanding that, in the case of a grouped frequency distribution, the calculated mean and standard deviation are estimates.**Learners should understand and be able to use the following formulae for standard deviation:**,* [*Formal estimation of population variance from a sample is excluded. Learners should be aware that there are different naming and symbol conventions for these measures and what the symbols on their calculator represent.*] |  | ML3 |
| 2.02h2.02i2.02j | Outliers and cleaning data | h) Recognise and be able to interpret possible outliers in data sets and statistical diagrams.i) Be able to select or critique data presentation techniques in the context of a statistical problem.j) Be able to clean data, including dealing with missing data, errors and outliers.*Learners should be familiar with definitions of outliers:**1. more than 1.5 × (interquartile range) from the nearer quartile**2. more than 2 × (standard deviation) away from the mean.* |  | ML4 |
| 2.04h | Selecting an appropriate distribution |  | h) Be able to select an appropriate probability distribution for a context, with appropriate reasoning, including recognising when the binomial or normal model may not be appropriate.*Includes understanding that a given binomial distribution with largecan be approximated by a normal distribution.*[*Question*s *explicitly requiring calculations using the normal approximation to the binomial distribution are excluded.*] | MN2MN3 |

# Thinking Conceptually

General approaches

Prior to working on data presentation and interpretation, it would be beneficial if learners had a firm understanding of the statistical problem solving cycle. This should be a core component of the initial approach. Data presentation and interpretation allows learners to use purposeful enquiry, using situations that are of interest to the learners.

Learners should have the opportunity to work with a range of different data sets, including but not restricted to the pre-release Large Data Set, both collected themselves and from published sources. They should develop confidence using technology to draw charts and calculate summary statistics, as well as being able to do these techniques manually.

Learners should also spend time critiquing different representations of data.

This is an area of mathematics with an exciting range of skills at its core. This includes investigation and gathering, presenting and examining the information collected. Learners understanding should be deepened by a hands-on approach.

The context of any data needs to be understood by learners in order for them to interpret their calculations correctly in the context. Teachers should make an emphasis on making inferences from the data.

Learners should understand the need to organise the data in ways that make it easy to see the main features. They should be encouraged to organise data to use suitable data displays and summary measures and should use the statistics to draw inferences about the population.

There is a lot of scope to explore large data sets and to use different ICT to perform calculations and produce different diagrams.

This part of the course needs to include working with the examination pre-release Large Data Set and learners should spend time understanding the context of the Large Data Set prior to any interpretations and data presentation.

A good starting point for introducing the Large Data Set is to look at the data for your own local authority, choosing a range of presentation types and interpreting the data in the context of students’ experience and local knowledge. You can then extend to nearby local authorities, and to some contrasting local authorities with very different data, before trying to work with the whole data set. It is not necessary for students always to create graphs and presentations for themselves, particularly early on, and much can be gained from an exploratory approach based on graphs and summary data produced by the teacher.

The majority of this unit of work can be covered using the pre-release Large Data Set; this can be used to draw out the important statistical concepts. Final assessment will be based on the assumption that learners are familiar with the underlying context of the data and the terminology used.

### Common misconceptions or difficulties learners may have

There are a number of misconceptions that learners may hold, or develop regarding data presentation and interpretation and care should be taken to avoid these becoming ingrained in learners.

One source of confusion for learners may be in differentiation between visually similar graph types (e.g. frequency bar charts, histograms) and the different methods used to evaluate the different representations. This should be explored to ensure the difference is clear.

Learners need to be clear about the language associated with data presentation and interpretation. The difference between the common terms such as population, sample, correlation, causation needs to be clarified and understood by all learners to avoid confusion with this unit of work.

Teachers should ensure that time is spent discussing the concept of variability of data in general, and should not limit the focus to calculating variability (i.e. range, standard deviation). Learners should have a sense of what is meant by variability of data, and developing the concept of variability and make comparisons of variability within the *context* of the data given.

Learners should be encouraged to pay attention to the context of the data; learners often confuse their calculated answer as a final result in itself, rather than referring back to the initial context given to give their calculated answer meaning. Teachers should encourage learners to check the reasonableness of their results within the context of the data, and the context should remain at the centre of any learning.

It is often the case that learners present conclusions from data that are too assertive, for example “This data shows that….” instead of “This data suggests that …” or “There is some evidence that….” It is important that they understand that deductive statements *can* be validly made, “In this sample it is the case that …” but that conclusions about a population from a sample are almost always inferential.

Conclusions from scatter diagrams (and bivariate data in general) need to be treated with caution. The old adage that “correlation does not imply causation” is a key one for learners to understand. It is often over-interpreted as meaning that when there *is* correlation there is definitely no causation, which is not the case, so the direction of implication is important. It can be very instructive to think about what relationship there might be between two variables, other than direct causation. If you want to demonstrate causation, then these other relationships need to be ruled out. Another common misconception is that strong correlation implies *linear* correlation; using scatter diagrams to check the type of association is an important tool.

Outliers are often a source of confusion; identifying an outlier (2.02h and j) should not automatically lead to removing that data point. Outliers can simply be caused by sampling variation, by unusual examples or subpopulations. Not only should outliers not automatically be removed, they are sometimes the starting point of an interesting investigation around why it is that the particular example is so different from the rest of the data. In a sample, or data set, with apparently many outliers it might be more instructive to attempt to identify a subpopulation that needs separate consideration and/or to use robust measures such as median and interquartile range in preference to mean and variance, in order to control for the skewing effects of those outliers.

The specification specifically excludes using the normal distribution as an approximation to the binomial distribution, but *does* include (at A Level only - 2.04h) understanding the connection between them in general terms. A common misconception is that binomial distributions tend towards normal as *n* tends to infinity for *any* value of *p* rather than just those close to 0.5. This can be explored using technology, but it is important to explore the binomial for large *n* and *p* not close to 0.5 as a non-example.

When working with the Large Data Set it is easy to forget that it is in effect population data, not a sample. Conclusions *can* be more assertive, though care still needs to be taken, particularly with the Age Structure data where the grouping has lost some of the original data. The idea that calculations made from grouped data (for example of mean and standard deviation) are estimates of the population parameters, despite the data not being from a sample, is a key point.

In the Age Structure part of the data set it is important to make sure that learners appreciate the effects of unequal grouping. If you simply produce a bar chart using the classes as categories, then it gives the impression of a very heavily middle-aged population with very few young people.

# Thinking Contextually

Learners need to see the relevance of their learning to real life events; they often struggle to understand the concepts in mathematics unless they can see the relevance.

The very nature of data presentation and interpretation is contextual and many different areas can be used to enhance learners understanding; these can be as basic as collection of data within the classroom about themselves that they can then interpret, through to more complex examples such as data collection for world population data or CO2 emissions to interpret the world around us.

Learners will be more successful if they can see how the concepts can be used outside of the classroom. If scenarios are chosen that are meaningful to the learners this will help to maintain their interest and motivation. With this in mind it is useful to have learners investigate a great range of contexts and data sets to highlight the scope of applications that these skills can be put to use. This will also help learners to focus on the mathematics and lead to independent thinking and greater retention of the skills.

# Past paper examples

[2018 H230/01](https://www.ocr.org.uk/Images/535662-question-paper-pure-mathematics-and-statistics.pdf) Q 13: Data, from the Large Data Set (LDS), presented in an unusual (radar diagram) manner. Some parts focus on interpreting the data provided, whilst the final part requires candidates to make use of their knowledge of the LDS.

[2018 H240/02](https://www.ocr.org.uk/Images/535611-question-paper-pure-mathematics-and-statistics.pdf) Q 11: This question requires candidates to use their knowledge of bivariate data and PPMC and apply this knowledge to interpret results from comparisons of columns from the LDS.

# Resources

| **Title** | **Organisation** | **Description** | **Ref** |
| --- | --- | --- | --- |
| [OCR Large Data Set](http://www.ocr.org.uk/Images/308727-units-h230-and-h240-large-data-set-lds-sample-assessment-material.xlsx) | OCR | This is the excel spreadsheet pre-release Large Data Set that will be used for some of the questions in the H230/H240 statistical papers. | 2.02a, 2.02b, 2.02f, 2.02g and 2.02h |
| [2.02 Data presentation and interpretation Check In](https://www.ocr.org.uk/Images/429923-section-check-in-2.02-data-presentation-and-interpretation.docx) | OCR | Set of 10 Questions, with worked solutions, on section 2.02 | 2.02 |
| [Large Data set: starter activities 1](https://www.ocr.org.uk/Images/422843-large-data-set-starter-activities.doc) | OCR | This resource consists of five short starter activities on data presentation and interpretation using data from the large data set.  | 2.02 |
| [Large Data set: starter activities 2](https://www.ocr.org.uk/Images/429548-large-data-set-starter-activities-set-2.doc) | OCR | This second resource on using the Large Data Set consists of four more short starter activities. | 2.02 |
| [Large Data set: statistical calculations](https://www.ocr.org.uk/Images/361199-lds-investigating-the-usage-of-umlrt-statistical-calculations-lesson-element.doc) | OCR | A resource using the large data set (LDS) to investigate whether usage of underground, metro, light rail and tram (UMLRT) has increased between 2001 and 2011.  | 2.02 |
| [Data Sets](http://mei.org.uk/data-sets) | MEI | This is a selection of large data sets. These data sets are provided for teachers of statistics to use with their students. Information is given about the data and an indication is given of statistical techniques that may be useful when working with the data set. | 2.02a, 2.02b, 2.02f, 2.02g and 2.02h. |
| [LDS Resource Videos](https://www.youtube.com/playlist?list=PL9HEotZZw6R8FA0CzwDLSM9kAcHLDjuN7) | MEI | Set of videos looking at the use of technology for investigating the Large Data Set and general support for using computer technology in teaching statistics. | 2.02 |
| [Astro Pi Flight Data Analysis](https://www.raspberrypi.org/learning/astro-pi-flight-data-analysis/worksheet/) | Raspberry Pi Foundation | This is a Large Data Set taken from the International Space Station. The data can be analysed by students, and there are examples of possible analyses to use.  | 2.02a, 2.02b, 2.02f, 2.02g and 2.02h |
| [Single variable basic tools](https://www.geogebra.org/m/SVvsgGem) | Geogebra | A basic introduction to using the spreadsheet and statistical chart functionality of geogebra. | 2.02a |
| [Measures of spread](https://www.geogebra.org/m/DjkkHmtd) | Geogebra | An interactive box-and-whisker plot that can be manipulated to correspond to a random list of integers. | 2.02a and 2.02f |
| [LDS – Investigating bicycle use in England and Wales (repeated sampling)](http://www.ocr.org.uk/Images/361199-lds-investigating-the-usage-of-umlrt-statistical-calculations-lesson-element.doc) | OCR | Learners will investigate whether usage of ‘underground, metro, light rail and tram’ (henceforth UMLRT) has increased from 2001 to 2011. The activity highlights the need to interpret percentage change with caution. Learners will need to manage missing data and to use a spreadsheet to calculate summary statistics. This activity covers aspects of 2.02a (interpret tables for single-variable data), 2.02f (measures of average and spread) and 2.02j (cleaning data). | 2.02a, 2.02f and 2.02j |
| [Analysing Networks](http://nrich.maths.org/12124) | Nrich | This activity allows students to understand and use different ways of interpreting the data about social networks, and what this tells us about how individuals interact with each other. This is then used to suggest models for social behaviour in the outbreak of a disease. | 2.02a and 2.02b |
| [Data Analysis](http://www.nuffieldfoundation.org/sites/default/files/files/FSMQ%20A-level%20results%20%28Feb%2011%29.pdf) | Nuffield Foundation | A lesson outline with exam results data for students to analyse with questions for them to use the analysis to answer. Students could also then use their school exam results with the data in question for comparison.Data for different years is available via the JCQ website: <https://www.jcq.org.uk/examination-results/a-levels>  | 2.02a, 2.02f and 2.02i |
| [The Lives of Presidents](https://nrich.maths.org/11007) | Nrich | A set of data on US Presidents with questions for students to try to address by analysing the spreadsheet of data.  | 2.02a, 2.02b, 2.02f and 2.02g |
| [Tinkerplots](https://www.tinkerplots.com/) | Tinkerplots | This is a data visualisation tool which is intuitive and useful for developing interpretation of non-standard presentations. | 2.02a, 2.02i |
| [Trendalyzer](https://www.gapminder.org/tools/#_chart-type=bubbles) | Gapminder | Trendalyzer is a great tool for creating scatter diagrams for bivariate data, but also for exploring multivariate and dynamic visualisations. | 2.02a, 2.02c, 2.02i |
| [Level 3 Data Analysis](http://www.nuffieldfoundation.org/fsmqs/level-3-data-analysis) - Stature | Nuffield Foundation | This website has many different statistical tasks; the one discussed here is Stature. This is an activity for students to use data from different countries to draw histograms and draw conclusions. Extends to using the normal distribution. | 2.02b |
| [Scatter Plots](http://www.mathsisfun.com/data/scatter-xy-plots.html)  | Maths is fun | An activity covering examples of scatter diagrams and correlation. There are some questions to check understanding.  | 2.02c, 2.02d and 2.02e |
| [Devising a Measure](http://map.mathshell.org/lessons.php?unit=9410&collection=8) | MARS | This is a series of lessons intended to assess how well students understand positive correlation. This allows students to work with other responses to identify misconceptions.  | 2.02c, 2.02d and 2.02e |
| [Understanding Standard Deviation](https://www.geogebra.org/m/s2BaqWpm) | Geogebra | A nice visual demonstration of standard deviation using 5 variables on a number line. | 2.02f |
| [Standard Deviation Formulae](https://www.geogebra.org/m/DS6PUaXy) | Geogebra | Change the numbers and then go through the steps of the calculation. | 2.02f |
| [Standard Deviation](https://www.examsolutions.net/tutorials/standard-deviation/?board=OCR&module=s1&topic=1820) | Exam Solutions | Two videos that cover this topic area. The first is examples on how to calculate standard deviation using the formula and the second is the same but for frequency tables.  | 2.02f |
| [Descriptive statistics](https://www.thinkib.net/mathhlsl/page/16476/descriptive-statistics) | In Thinking | Notes and set of questions on charts and calculations. | 2.02f |
| [Casio fx-991EX Classwiz calculator. Finding mean, variance, standard deviation etc](https://www.youtube.com/watch?v=Y1gsxlSbojs) | Youtube | A run through of the statistical functions on the Casio fx 991 EX Classwiz. | 2.02g |
| [TI 36x Pro Basic Statistics: Standard Deviation and Mean Tutorial](https://www.youtube.com/watch?v=KQHSysxSyXE) | Youtube | Although a slightly different Texas Instruments model than that available in UK, the instructions match those for the TI 30X Pro. | 2.02g |
| [Outliers](http://www.mathsisfun.com/data/outliers.html) | Maths is fun | An activity covering examples of outliers and their effect on the mean, median and mode. There are some questions to check understanding.  | 2.02h |
| [Outlier Test](http://www.s253053503.websitehome.co.uk/msv/msv-4/msv-4.xls) | MSV | This is a really simple but fun spreadsheet that gives students a chance to explore the definition of an outlier that uses the IQR. Used in the whole class setting, students' intuitions about what is likely to be an outlier can be offered up through a range of different problems. | 2.02h and 2.02j |
| [Impact on Median and Mean: Increasing an Outlier](https://www.khanacademy.org/math/probability/data-distributions-a1/summarizing-center-distributions/v/impact-on-median-and-mean-when-increasing-highest-value) | Khan Academy | A video covering this topic area. Shows a worked example and goes on to further practice. | 2.02h, and 2.02j |
| [Box plots and outliers](https://www.geogebra.org/m/brwwN1pM) | Geogebra | This is designed to investigate outliers. The points on the line represent the scores of 50 students in an examination. The points can be dragged along the line to investigate how the shape of the boxplot changes. | 2.02j |
| [Data cleansing](https://en.wikipedia.org/wiki/Data_cleansing) | Wikipedia | Some general notes on the reasons why data may need to be cleaned. | 2.02j |
| [Myth busting with calculators including statistical tables - Steven Kean, Science Studio](https://www.youtube.com/watch?v=I8xa_s0WS-o&list=PLtzR6sheDAMHP9E6GnRe9kIfwNGX7bEKi&index=10) | OCR | A short presentation on using scientific and graphical calculators, including the use of calculators for statistical calculations. | 2.02 |

**OCR Resources**: *the small print*OCR’s resources are provided to support the delivery of OCR qualifications, but in no way constitute an endorsed teaching method that is required by the Board, and the decision to use them lies with the individual teacher. Whilst every effort is made to ensure the accuracy of the content, OCR cannot be held responsible for any errors or omissions within these resources.

© OCR 2020 - This resource may be freely copied and distributed, as long as the OCR logo and this message remain intact and OCR is acknowledged as the originator of this work.

OCR acknowledges the use of the following content: n/a

Please get in touch if you want to discuss the accessibility of resources we offer to support delivery of our qualifications: resources.feedback@ocr.org.uk

We’d like to know your view on the resources we produce. By clicking on ‘Like’ or ‘Dislike’ you can help us to ensure that our resources work for you. When the email template pops up please add additional comments if you wish and then just click ‘Send’. Thank you.

Whether you already offer OCR qualifications, are new to OCR, or are considering switching from your current provider/awarding organisation, you can request more information by completing the Expression of Interest form which can be found here: [www.ocr.org.uk/expression-of-interest](http://www.ocr.org.uk/expression-of-interest)

Looking for a resource? There is now a quick and easy search tool to help find free resources for your qualification:
[www.ocr.org.uk/i-want-to/find-resources/](http://www.ocr.org.uk/i-want-to/find-resources/)