# OCR - Oxford Cambridge and RSA Teacher Delivery Guide Statistics: Data Presentation and Interpretation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Specification** | **Ref.** | **Learning outcomes**  | **Notes** | **Notation** | **Exclusions** |
| **STATISTICS: DATA PRESENTATION AND INTERPRETATION (1)** |
| Data presentation for single variable | MD1 | Be able to recognise and work with categorical, discrete, continuous and ranked data.  Be able to interpret standard diagrams for grouped and ungrouped single-variable data. | Includes knowing this vocabulary and deciding what data presentation methods are appropriate: bar chart, dot plot, histogram, vertical line chart, pie chart, stem-and-leaf diagram, box-and-whisker diagram (box plot), frequency chart.Learners may be asked to add to diagrams in examinations in order to interpret data. | A frequency chart resembles a histogram with equal width bars but its vertical axis is frequency. A dot plot is similar to a bar chart but with stacks of dots in lines to represent frequency. | Comparative pie charts with area proportional to frequency. |
| D2 | Understand that the area of each bar in a histogram is proportional to frequency. Be able to calculate proportions from a histogram and understand them in terms of estimated probabilities. | Includes use of area scale and calculation of frequency from frequency density. |  |  |
| D3 | Be able to interpret a cumulative frequency diagram. |  |  |  |
| D4 | Be able to describe frequency distributions. | Symmetrical, unimodal, bimodal, skewed (positively and negatively). |  | Measures of skewness. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Specification** | **Ref.** | **Learning outcomes** | **Notes** | **Notation** | **Exclusions** |
| **STATISTICS: DATA PRESENTATION AND INTERPRETATION (1)** |
| Data presentation | MD5 | Understand that diagrams representing unbiased samples become more representative of theoretical probability distributions with increasing sample size. | e.g. A bar chart representing the proportion of heads and tails when a fair coin is tossed tends to have the proportion of heads increasingly close to 50% as the sample size increases. |  |  |
| D6 | Be able to interpret a scatter diagram for bivariate data, interpret a regression line or other best fit model, including interpolation and extrapolation, understanding that extrapolation might not be justified. | Including the terms association, correlation, regression line.Leaners should be able to interpret other best fit models produced by software (e.g. a curve).Learners may be asked to add to diagrams in examinations in order to interpret data. |  | Calculation of equation of regression line from data or summary statistics. |
| D7 | Be able to recognise when a scatter diagram appears to show distinct sections in the population. Be able to recognise and comment on outliers in a scatter diagram. | An outlier is an item which is inconsistent with the rest of the data.Outliers in scatter diagrams should be judged by eye. |  |  |
|  | D8 | Be able to recognise and describe correlation in a scatter diagram and understand that correlation does not imply causation. | Positive correlation, negative correlation, no correlation, weak/strong correlation. |  |  |
|  | D9 | Be able to select or critique data presentation techniques in the context of a statistical problem. | Including graphs for time series. |  |  |
| **Bivariate data, association and correlation** |
| **Bivariate data** consists of two variables for each member of the population or sample. An **association** between the two variables is some kind of relationship between them. **Correlation** measures linear relationships. At A level, learners are expected to judge relationships from scatter diagrams by eye and may be asked to interpret given correlation coefficients – see MAH10. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Specification** | **Ref.** | **Learning outcomes** | **Notes** | **Notation** | **Exclusions** |
| **STATISTICS: DATA PRESENTATION AND INTERPRETATION (1)** |
| Summary measures | MD10 | Know the standard measures of central tendency and be able to calculate and interpret them and to decide when it is most appropriate to use one of them. | Median, mode, (arithmetic) mean, midrange. The main focus of questions will be on interpretation rather than calculation.Includes understanding when it is appropriate to use a weighted mean e.g. when using populations as weights. | Mean |  |
|  | D11 | Know simple measures of spread and be able to use and interpret them appropriately. | Range, percentiles, quartiles, interquartile range. |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Specification** | **Ref.** | **Learning outcomes** | **Notes** | **Notation** | **Exclusions** |
| **STATISTICS: DATA PRESENTATION AND INTERPRETATION (1)** |
| Summary measures | MD12 | Know how to calculate and interpret variance and standard deviation for raw data, frequency distributions, grouped frequency distributions.Be able to use the statistical functions of a calculator to find mean and standard deviation. | sample variance: (†)where  sample standard deviation: (‡) |  | Corrections for class interval in these calculations. |
| D13 | Understand the term outlier and be able to identify outliers. Know that the term outlier can be applied to an item of data which is:* at least 2 standard deviations from the mean;

OR* at least 1.5 × IQR beyond the nearer quartile.
 | An outlier is an item which is inconsistent with the rest of the data. |  |  |
| D14 | Be able to clean data including dealing with missing data, errors and outliers. |  |  |  |
| **Notation for sample variance and sample standard deviation** |
| The notations *s*2 and *s* for sample variance and sample standard deviation, respectively, are written into both British Standards (BS3534-1, 2006) and International Standards (ISO 3534). The definitions are those given above in equations (†) and (‡). The calculations are carried out using divisor .In this specification, the usage will be consistent with these definitions. Thus the meanings of ‘sample variance’, denoted by *s*2,and ‘sample standard deviation’, denoted by *s*, are defined to be calculated with divisor .In early work in statistics it is common practice to introduce these concepts with divisor *n* rather than . However there is no recognised notation to denote the quantities so derived.Students should be aware of the variations in notation used by manufacturers on calculators and know what the symbols on their particular models represent. |

# 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 undertake purposeful enquiry in situations that are of interest to the learners.

Learners should have the opportunity to work with a range of different data sets, 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 where appropriate.

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.

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 as part of any interpretations and data presentation.

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 emphasise 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; learners should use the statistics to make 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.

The majority of this unit of work can be covered using the pre-release large data set. Final assessment will be based on the assumption that learners are familiar with the underlying context of the data and are able to manipulate the data.

The pre-release data set is large and is provided on a spreadsheet, it will not be practical to carry out analysis on the whole data set by hand.

### Common misconceptions

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.

Time needs to be given to developing learners’ understanding of the meaning of variability of the data.

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 measures of 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 with the final result, rather than referring back to the initial context given to give the result 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 useful to present incorrectly drawn presentations of the data to learners to encourage them to highlight and discuss the errors.

# 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 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.

# Resources

| **Title** | **Organisation** | **Description** | **Ref** |
| --- | --- | --- | --- |
| [Data processing, presentation and interpretation (AS)](http://www.mei.org.uk/files/sow/14-data-processing-presentation-interpretation.pdf) | MEI | A commentary of the underlying mathematics, a sample resource, a use of technology, links with other topics, common errors, opportunities for proof and questions to promote mathematical thinking. | D1- D14 |
| [Large Data Set (LDS\_1)](http://www.ocr.org.uk/Images/400305-large-data-set-lds_1-h630-2018-h640-2018-2019-pre-release-data-mei.xlsx) | OCR | This is the excel spreadsheet pre-release Large Data Set that will be used for some of the questions in the H630/H640 statistical papers:2018 H630 and H640 2019 H640 | D1, D2, D10, D11, D12 and D13 |
| [Large Data Set (LDS\_2)](http://www.ocr.org.uk/Images/400306-large-data-set-lds_2-h630-2019-h640-2020-pre-release-data-mei.xlsx) | OCR | This is the excel spreadsheet pre-release Large Data Set that will be used for some of the questions in the H630/H640 statistical papers:2019 H630 2020 H640 | D1, D2, D10, D11, D12 and D13 |
| [Large Data Set (LDS\_3)](http://www.ocr.org.uk/Images/400307-large-data-set-lds_3-h630-2020-h640-2021-pre-release-data-mei.xlsx) | OCR | This is the excel spreadsheet pre-release Large Data Set that will be used for some of the questions in the H630/H640 statistical papers:2020 H630 2021 H640 | D1, D2, D10, D11, D12 and D13 |
| [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. | D1, D2, D10, D11, D12 and D13 |
| [Single variable basic tools](https://www.geogebra.org/m/SVvsgGem) | Geogebra | A basic introduction to using the spreadsheet and statistical chart functionality of geogebra. | D1 |
| [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. | D1 and D2 |
| [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.  | D1, D2, D10, D11, D12 and D13 |
| [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. | D1, D10 and D11 |
| [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. | D1, D10, D11 and D14 |
| [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>  | D1, D10 and D11 |
| [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.  | D1, D2, D10, D11 and D12 |
| [Level 3 Data Analysis - Stature](http://www.nuffieldfoundation.org/fsmqs/level-3-data-analysis) | 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. | D2 |
| [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.  | D6, D7 and D8  |
| [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.  | D6, D7 and D8  |
| [Understanding Standard Deviation](https://www.geogebra.org/m/s2BaqWpm) | Geogebra | A nice visual demonstration of standard deviation using 5 variables on a number line. | D10, D11 |
| [Standard Deviation Formulae](https://www.geogebra.org/m/DS6PUaXy) | Geogebra | Change the numbers and then go through the steps of the calculation. | D10, D11 |
| [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.  | D10, D11 |
| [Descriptive statistics](https://www.thinkib.net/mathhlsl/page/16476/descriptive-statistics) | In Thinking | Notes and set of questions on charts and calculations. | D10, D11 |
| [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. | D12 |
| [TI 36x Pro Basic Statistics: Standard Deviation and Mean Tutorial](https://www.youtube.com/watch?v=KQHSysxSyXE) | Youtube | Although slightly different Texas Instruments model than that available in UK, the instructions match those for the TO30X Pro. | D12 |
| [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. | D12 |
| [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.  | D13 |
| [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. | D13 and D14 |
| [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. | D13, and D14 |
| [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. | D14 |
| [Data cleansing](https://en.wikipedia.org/wiki/Data_cleansing) | Wikipedia | Some general notes on the reasons why data may need to be cleaned. | D14 |
| [Graph it!](http://www.censusatschool.com/resources.html) | Census at School | Matching graphs to their names and purposes. There are many other useful resources in the Census at School resource archive. | D1, D2, D6 |

**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 2017 - 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/)