**Y12 AS Mathematics**

**Units 14 & 15: Data collection, processing, presentation and interpretation 2 weeks**

## Teaching objectives

**a** Understand and use the terms ‘population’ and ‘sample’

**b** Use samples to make informal inferences about the population

**c** Understand and use sampling techniques, including simple random sampling, stratified sampling, systematic sampling, quota sampling and opportunity (or convenience) sampling

**d** Select or critique sampling techniques in the context of solving a statistical problem, including understanding that different samples can lead to different conclusions about the population

**e** Interpret diagrams for single-variable data

**f** Understand that area in a histogram represents frequency

**g** Connect to probability distributions

**h** Interpret scatter diagrams and regression lines for bivariate data, including recognition of scatter diagrams which include distinct sections of the population (calculations involving regression lines are excluded)

**i** Understand informal interpretation of correlation

**j** Understand that correlation does not imply causation

**k** Interpret measures of central tendency and variation, extending to standard deviation

**l** Be able to calculate standard deviation, including from summary statistics

**m** Recognise and interpret possible outliers in data sets and statistical diagrams

**n** Select or critique data presentation techniques in the context of a statistical problem

**o** Be able to clean data, including dealing with missing data, errors and outliers

**Resources for advance preparation:**

[**TES: Sampling Techniques**](https://www.tes.com/teaching-resource/maths-gsce-statistics-sampling-techniques-6082332)– 1 set of cards for each small group

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Starter** | **Main teaching**  Including key questions, key teaching points, models and resources | **Notes**  Including Support and Extension | **Consolidation/Plenary**  Including key questions and homework |
| 1 | Look at one of the LDS. What questions could I ask about this data???  I wonder why……..?  Is there a connection between……?  Does that match with my experience?  Have things changed over time?  How can I get a sensible amount of data from this huge data set?  How can I organise the data?  Maybe if I extracted certain parts of the data?  What sampling methods could I use?  What diagrams might help?  What calculations might help? | [**TES: Sampling Techniques**](https://www.tes.com/teaching-resource/maths-gsce-statistics-sampling-techniques-6082332)  Take a question from the starter and select a simple random sample. Create a Box Plot. Repeat for different random samples – the same size and different sizes.  What do they notice when the sample size is the same?  What do they notice about different sample sizes? | Record questions and techniques required for future reference  Ensure that students are clear about simple random sampling: *all possible samples of the given size need to be equally likely*. This is not the same as all elements being equally likely.  It is easy to do repeated random samples and Box Plots in Gnumeric. | Homework:  [**Integral: Exercise Section 1 level 1**](https://2017.integralmaths.org/pluginfile.php/10495/mod_resource/content/0/assd1ax_level1.pdf)  [**Integral: Exercise Section 1 level 2**](https://2017.integralmaths.org/pluginfile.php/10496/mod_resource/content/0/assd1ax_level2.pdf) |
|  | **Starter** | **Main teaching**  Including key questions, key teaching points, models and resources | **Notes**  Including Support and Extension | **Consolidation/Plenary**  Including key questions and homework |
| 2 | As for lesson 1, but using a different LDS, so that different questions are raised requiring different techniques.  Note that each LDS has different strengths and weaknesses as teaching resources, so you will need to use more than one in order to cover all the necessary content.  In any case, using a variety of LDS makes the lessons more interesting! | Use ‘An AS lesson using a LDS’ to explore further questions involving averages and measures of spread.  (Editable powerpoint attached.) | Using a spreadsheet to calculate, for example, standard deviation, helps students to understand what these calculations represent, while being much quicker, and allowing larger samples to be used, than calculations by hand.  Use of calculators to simply generate the numbers follows once understanding is secure. | As for lesson 1. |
| 3 | Look at Edexcel LDS.  Which factors affect visibility? Why this could be important?  Is there a connection between mean visibility and daily maximum relative humidity?  Have things changed between 1987 and 2015?  How can I get a sensible amount of data from a huge data set?  What sampling methods can I use?  What diagrams might help?  What calculations might be useful? | Take random samples of data from different locations in pairs from 1987 and 2015. Sample sizes of 20 or larger.  Use Casio CG-20 and Task 4 description from MEI to enter and analyse data, using scatter graph  [**http://mei.org.uk/files/ict/casio-statistics-tasks.pdf - page=1**](http://mei.org.uk/files/ict/casio-statistics-tasks.pdf#page=1)  What do you notice from different sample sizes and locations?  Choose your own variables from the LDC to investigate your hypotheses on correlation. | Talk about correlation and causation and the difference between them.  Ensure students are clear about simple random sampling.  Teach generating random numbers in Excel, using RAND() function. | Homework:  [**Watch a video on calculating PMCC on Integral**](https://2017.integralmaths.org/mod/resource/view.php?id=5961)  [**Integral: Exercise Section 3 Level 2**](https://2017.integralmaths.org/pluginfile.php/19224/mod_resource/content/1/edexcelassd3ax_level2.pdf) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Starter** | **Main teaching**  Including key questions, key teaching points, models and resources | **Notes**  Including Support and Extension | **Consolidation/Plenary**  Including key questions and homework |
| 4 | [**Integral: Histogram reconstruction**](http://mei.org.uk/files/sow/14-data-processing-presentation-interpretation-res.pdf)  Students should be familiar histograms from GCSE, so this is for revision. | Measures of spread. Recap Interquartile Range (see previous lessons Box Plots) as a measure of spread. Then use data from an LDS to introduce standard deviation. Set up the calculation in a spreadsheet so students can see how it works. Investigate standard deviation for repeated sampling.  Note that MEI require students to work with sample standard deviation [divide by (n-1)] not population standard deviation. Some discussion will be needed about the difference. | Note that simply subtracting the mean from each data item would give a total of zero. You could take the modulus, but it’s easier to calculate with squaring instead. Also, this exaggerates the effect of data which is further from the mean. Why might this be useful?  Note that students need to understand what the calculation is, which is easier to do from a spreadsheet.  Remember: no method marks – students are expected to use advanced calculators. | Look at how to calculate standard deviation from summary statistics, and by entering data into a graphical calculator or classwizz or equivalent.  Homework:  [**Integral: Exercise Section 2 Level 1**](https://2017.integralmaths.org/pluginfile.php/915/mod_resource/content/0/assd2ax_level1.pdf)  NB the point of much of this exercise is to practise using the statistical functions on their calculators! |
| 5 |  | Investigate missing data and outliers in one or more of the LDS. Discuss different options for dealing with them. |  |  |
| 6&7 | Use teaching activities from [**Integral**](https://2017.integralmaths.org/course/view.php?id=5&sectionid=84) to address any gaps. | Students to investigate some of their questions about one or more of the Large Data Sets, using calculations and diagrams as appropriate. |  | Homework:  [**Integral: Exercise Section 2 Level 2**](https://2017.integralmaths.org/pluginfile.php/916/mod_resource/content/0/assd2ax_level2.pdf) |
| 9 |  | Bivariate data. Ask questions about the LDS. What problems arise? Students need to note when there is more than one distinct population. | ‘Correlation does not imply causation’  Note that informal interpretations only are required. *Not* correlation coefficients.  *Not* regression lines |  |
| 9 |  | Further activities with the LDS |  |  |

**Additional resources now available on Integral [FREE ACCESS WITH FMSP REGISTRATION], especially for the LDS.**