OCR
Oxford Cambridge and RSA

LEVEL 3 CERTIFICATE

Teacher guide

CORE MATHS B (MEI)

H869
For first teaching in 2015

Theme: Statistical Problem Solving

Version 1

MEI **Mathematics Education Innovation**

CONTENTS

- Most post-16 subjects involve some quantitative work.

- Quantitative Problem Solving qualification is designed to help those students of these subjects who do not take AS or A level Maths.

- The most common requirement is to be able to work competently and confidently with data.

- *Statistical Problem Solving* is designed to meet that need.

- Students who take this course can expect to do better in their other subjects.

- They will also be better prepared for higher education.

- This guide provides information and advice for teachers.

Would you prefer a Word version?

Did you know that you can save this pdf as a Word file using Acrobat Professional?

Simply click on **File > Save As Other . . .** and select **Microsoft Word**

We value your feedback

We'd like to know your view on the resources we produce. By clicking on the icon above you will help us to ensure that our resources work for you.

INTRODUCTION

Quantitative Problem Solving qualification is designed to meet the needs of the large number of post-16 students who are not taking AS or A level Mathematics. Many of them have chosen subjects where statistics is really important but are not comfortable with it.

They find the statistics is an impediment to their progress, holding them back in their chosen subjects. This course is designed to meet the needs of such students, giving them the background skills and experience that will allow them to use the statistics, and so have access to the learning across the curriculum that depends on it.

Students' problems do not end when they leave school or college. At the next stage in their lives, be it higher education or employment, they find that they need a background in statistics that allows them to work with data competently and confidently. At the moment many do not have this, and so once again are held back in their aspirations.

These two situations are obviously closely linked. The aim of the *Statistical Problem Solving* course is to address both of them.

Considerable thought has gone into the design of this course. How is it possible to turn this vision into reality? Two items proved critically important.

- A list of important skills for those entering higher education to study subjects that use statistics. This list was compiled after lengthy discussions with ten subjects (not including mathematics) and published in *A world full of data* (2013). These are clearly much the same skills that students will need in their other subjects.

- The Statistics Cycle. This encapsulates the framework in which statistics is used for problem solving. It shows how in real life applications statistics is an integrated whole rather than a set of isolated techniques. So the Statistics Cycle provides a model for how the subject should be presented at this level.

This guide explains how these ideas are built into the design of *Statistical Problem Solving*. Important features are the use of pre-release data and the way the content is organised and expressed. The course represents a departure from the way that statistics is often presented in school and college mathematics, at GCSE and subsequently in AS and A level.

The questions on the specimen examination paper are placed where appropriate through this guide. It will be seen that they are not about carrying out routine calculations, or drawing displays, for their own sake. Instead the emphasis is on using data to solve problems and to provide insights.

Our advice to teachers is that understanding the rationale for the design is the first step in helping your students get the most out of the course. So please read on…

WORKING WITH DATA

A key conclusion from the research underpinning *A world full of data* was that, across all user subjects, the need is not so much for students to learn particularly advanced techniques as to be very comfortable working with data.

Across all the subjects, there was strong agreement from those in higher education that new undergraduates would benefit from being sufficiently familiar with data that they can engage competently with new data sets as and when they meet them. This would mean that before arriving at university students should have done work that gave them a set of basic skills.

Reporting data	Using suitable diagrams to illustrate data. Using suitable measures to summarise data. Interpreting the main features of a data set, including any patterns, in terms of the situation that gave rise to it. Translating information from one form to another: tabular, graphical and narrative. Describing the information contained in a data set orally and in writing, in terms of the situation being investigated.
Data types	Knowing the difference between a sample and a population. Being aware of the source of a data set and the method of collection. Understanding the different types of variable and related measurement techniques. Understanding about natural variability and experimental error and the differences between them.
Analysing data	Framing suitable questions to ask when using data to investigate a situation. Being aware of the need for error analysis. Being prepared to criticise data and their presentation.

Above information taken from Porkess R. (2013) *A world full of data*, The Royal Statistical Society.

Statistical Problem Solving is designed to ensure that students gain the hands-on experience of working with real data that will allow them to acquire these skills. For that reason a large data set is provided as pre-release material at the start of the course.

The focus is thus rather different from AS and A level Mathematics. Those specifications map out a route leading towards mathematical statistics in higher education. So Mathematics students spend a considerable amount of their time developing a theoretical framework and relatively little of it on more practical work.

PRE-RELEASE DATA

To acquire the skills identified by higher education, students need to become so familiar with data that working with them is internalised. They need to reach a state where their reaction to a data set is along the lines of '*That looks interesting. Let me see what it is telling me.*'.

This will not be achieved through working with a few small data sets, but this is just the experience of many students currently being taught statistics. Instead, they need to learn for themselves what data feel like and what questions to ask.

The teaching of this course is expected to be data driven and with that in mind centres will be provided with a large pre-release data at the start of the course. This is described in the specification.

> In Statistical Problem Solving, the pre-release material will be a large data set that can be used as teaching material throughout the course. It is comparable to a set text for a literature course. It will be published, in electronic form, at the start of the course in September. In the examination it will be assumed that learners are familiar with the contexts covered by this data set. Half of the marks in the examination will be on questions that are based on the pre-release data set.
>
> *Quantitative Problem Solving (MEI) Specification*, 2014, OCR, pre-release material

There are a number of key points about the pre-release material.

- It should be seen primarily as teaching material providing data that can be used to draw out and illustrate important statistical concepts.

- The data set will be drawn from real life and will be rich and interesting.

- The assessment will be based on the assumption that students are familiar with the underlying context(s) and are reasonably able to find their way around the data.

- The data set will be so large that there will be no point in teachers trying to spot particular questions for their students. Some of the examination questions may be based on a particular sample from the set.

The pre-release data for *Statistical Problem Solving* Specimen Paper is drawn from the CIA Factbook. They give the location, population, birth rate, life expectancy and per capita GDP for every country in the world; the rankings of most of these fields are also given. In total there are 10 fields with up to 240 entries in each. The specimen examination questions on it are shown on . They show examples of some of the possible styles of question.

The data will be provided on a spreadsheet. The size of the data set means that it will not be practicable to carry out analysis by hand; using some form of software will be essential.

Working with spreadsheets should present no problems to students since it is covered in the specification requirements (IQR 2 to 7) of *Introduction to Quantitative Reasoning*; indeed the pre-release data provide a good opportunity for them to develop these skills. Students will not be expected to have access to a spreadsheet in the examination but questions may test whether they know how to use one. Question 5 on the *Introduction to Quantitative Reasoning* Specimen Paper does just that.

Some teachers may prefer to transfer the pre-release data into a statistics package and this is entirely acceptable. However, there is no requirement to do and such software will not be available in the examination.

Inevitably the data in the pre-release material will be secondary. However, many students will collect their own primary data as part of investigational work in at least one of their other subjects. These different experiences will complement each other. The experience of working with the pre-release material will give students the confidence, knowledge and skills to make sense of their own data.

Advice and help will be available for teachers on how to use the pre-release material in teaching particular topics. There may, however, be some items where its use would be artificial and teachers should not feel compelled to use it to cover the whole specification. Half of the marks on the examination paper will be on questions based on the pre-release material and the other half not.

Working with the pre-release material will help students to acquire many of the skills valued by higher education identified in the list on **page 4**. These will be enhanced by the content of the specification which is based around the Statistical Problem Solving Cycle. However, a word of caution is appropriate. These skills will not be acquired at a single exposure; it is not a tick-box exercise. The more opportunities students have to work with data, both within this course and in their other subjects, the better equipped they will be.

Teaching with the pre-release material will involve new classroom approaches and so we would like to recommend that teachers make use of the CPD opportunities associated with this course.

QUESTIONS FROM THE SPECIMEN PAPER BASED ON THE PRE-RELEASE MATERIAL

There are very many questions that could be set based on the pre-release material. Those on the specimen paper are just a sample of what could be set and so neither the questions nor their categorisation in any sense belong to a definitive list.

Question requiring a general familiarity with the data and the contexts

4. (a) The figures in the column headed "Population" have been summed to give the total of 7 174 654 290.
        Give two reasons why this is not the present world population.   **[2]**

    (b) Is it true or false that over a quarter of the world's population lives in just two countries? Justify your answer.   **[3]**

    (c) Use the data provided to compare the number of births per year in Ghana and the UK.   **[4]**

Question requiring understanding of the use of ICT to explore and analyse the data

5. In the accompanying data, GDP per capita is given by country.

    Explain how you would use the spreadsheet containing the pre-release data to calculate an estimate of the mean GDP per capita for the whole world.
    Explain what this figure represents.

    Demonstrate your method using data from Japan, Nigeria and Sri Lanka.   **[8]**

Long question based on a sample from the pre-release data

6. A research student thinks that valuable insights can be obtained from the data in the tables.
    In order to carry out a pilot investigation the student selects the follow sample of countries.

Algeria	Brazil	Costa Rica	Poland	Tonga
Bangladesh	Cambodia	France	Syria	Zambia
Barbados	Canada	New Zealand	Tajikistan	

    (i)   Why might the student want to use a sample for a pilot investigation when the full data set is available?   **[1]**

    (ii) Comment briefly on how it appears this sample was selected.   **[1]**

    The student uses this sample to investigate whether there is any association between the birth rate of a country and the life expectancy of its citizens.

    (iii) Write down null and alternative hypotheses for a 2-tail test.

          Carry out a hypothesis test, using Spearman's rank correlation coefficient. Use a 5% significance level.   **[9]**

    (iv) The student uses software to work out the product moment correlation coefficient for birth rate and life expectancy for all countries. The result is –0.84. Interpret this figure.   **[2]**

THE STATISTICAL PROBLEM SOLVING CYCLE

The importance of investigative work is widely recognised and, in one form or another, it is a part of nearly all subjects. It almost always starts with a problem; students then collect information, process and analyse it and draw conclusions. Finally they have to decide whether their conclusions provide a sensible solution to the problem. If not, they have to reconsider their approach. So a cyclical process is involved.

The information that students collect is often statistical data and so the procedures can be summarised by the Statistical Problem Solving Cycle. This is given in the specification for *Introduction to Quantitative Reasoning* and so is included in that for *Statistical Problem Solving*.

Quantitative Problem Solving (MEI) Specification, 2014, OCR, The Statistical Problem Solving Cycle
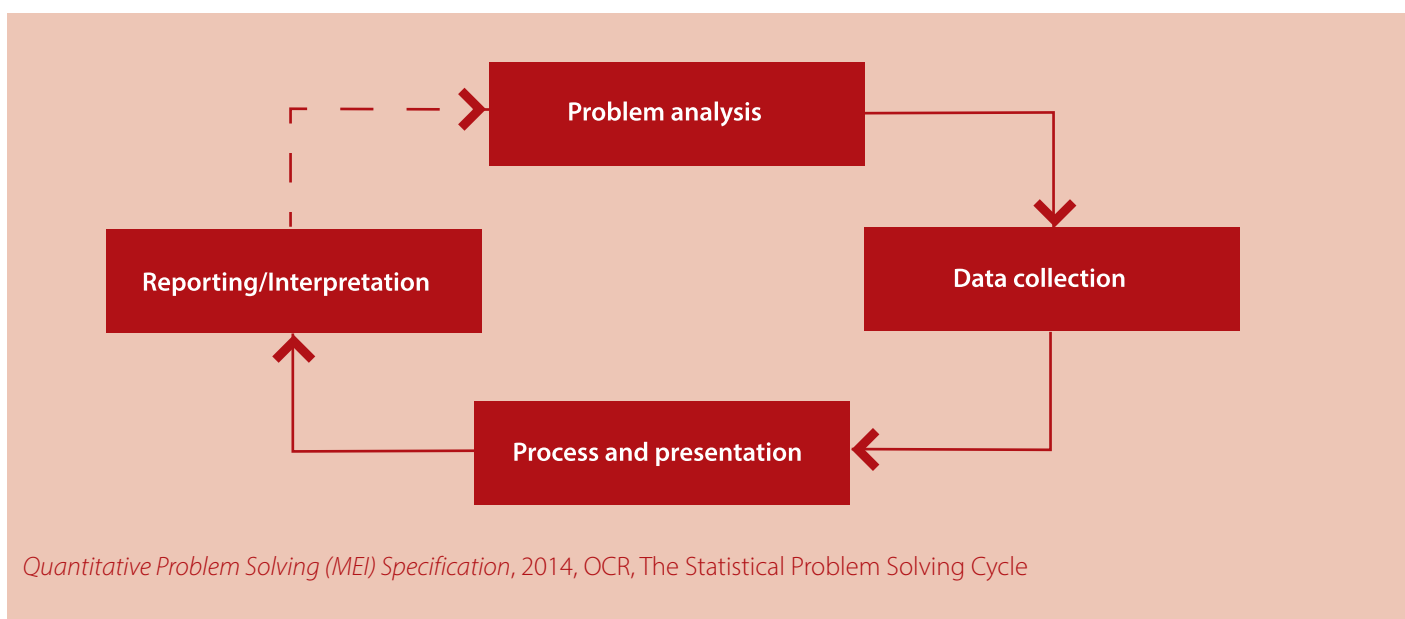
The specification for *Statistical Problem Solving* has four sections. Their headings are the same as the words in the four boxes of the Statistical Problem Solving Cycle. Thus every topic in the specification is placed within the overall context of statistical investigation. This has quite considerable implications.

- The importance of individual topics is seen less as deriving from their own right and more in terms of the contribution they make to using statistics to solve problems.

- Carrying out particular calculations (like standard deviation) and drawing data displays (like scatter diagrams) is less important than using them to illustrate and interpret data that are relevant to the overall problem.

- Consequently, the emphasis in the assessment will be on selecting and interpreting relevant techniques, rather on doing things that can easily be done using suitable technology.

- Many of the examination questions will require work that is located in more than one of the boxes in the Statistical Problem Solving Cycle.

The rest of this document looks at the content of the four sections in more depth. The headings and sub-headings are taken from the specification document.

PROBLEM ANALYSIS

Many situations in real life can be clarified by the use of suitable data. The process of problem analysis is essentially one of deciding what "suitable" means in the context and then planning a course of action. In the specification it is described in four stages.

Strategies for problem solving

Most real problems are complicated and need to be simplified to be tractable and allow statistical techniques to be applied. However their main features must not be lost.

Some subjects, such as psychology, require students to decide on their own investigations and how they are going to set about them. Such students should be seen as a rich classroom resource.

Selecting a model

A model is then set up. This may be a standard statistical model or it may be purpose made. Setting up models is not always easy and the practical activities in many subjects are based on given models. The Normal distribution and the uniform distribution are specifically mentioned in this specification. If some other model is required in the examination, sufficient guidance will be given.

Inputs

The model will require inputs in the form of data. At this stage it must be clear what data will be required and collected. This requirement should give rise to rich classroom discussions covering items raised under "Data types" in the list of higher education requirements on **page 5**: different types of data, their sources and their likely accuracy. These are very important concepts in statistics but they are often largely overlooked in statistics classrooms. Examination questions may ask for a critical appraisal of the choice of data.

Inference

The data will contain information but it has to be unlocked. There must be a clear plan as to how this will be done, whether through suitable summary measures and diagrams or deeper analysis. At this planning stage, students have to decide what they are going to do with the data once they have been collected. An all too common situation is for students to ask "*I have collected my data, what do I do next?*". They need to understand the various techniques available to them. These include hypothesis testing. Two tests are required, the χ^2 test using a contingency table and that based on Spearman's rank correlation coefficient; both these tests are commonly used in other subjects.

Finally it must be clear that it will be possible to interpret the information to shed light on the original problem. It must also be possible to make a judgement as to whether the problem has been adequately solved.

EXAMPLE FROM THE SPECIMEN PAPER

Modelling

1. A biologist is concerned about the possible decline in numbers of a type of bird in a wood. The bird eats insects and the biologist thinks that the use of insecticides on nearby fields may be one of the causes of the possible decline.

    The biologist observes 12 female birds during the breeding season and counts the number of fledglings (young that leave their nests). The results are shown on this chart.
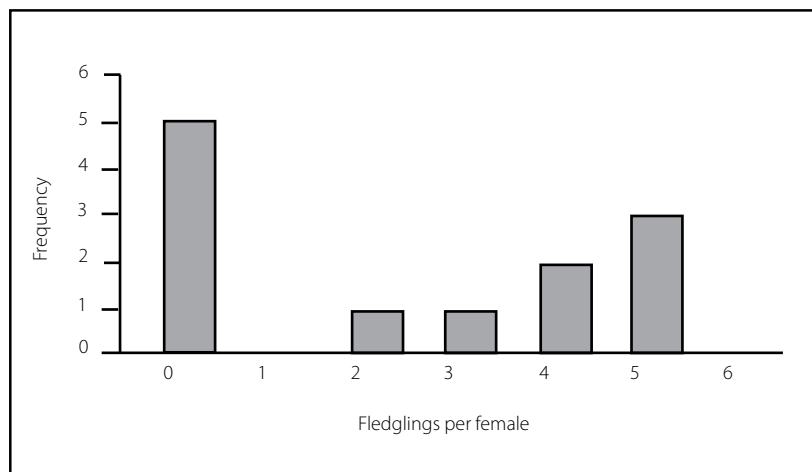
    She applies the following modelling assumptions to this situation.

    *   Half of the fledglings are male and half are female.
    *   75% of the fledglings will be taken by predators before they are old enough to breed.
    *   A female breeds three times in her lifetime.
    *   A female breeds once a year.

    (i)   Show that these observations and modelling assumptions lead to the conclusion that the birds will reduce in numbers.   **[4]**

    (ii)   Is there good evidence for the biologist to ask the farmer to stop using insecticides on the fields near the wood? Give three comments.   **[3]**

DATA COLLECTION

Sampling

The specification includes specific mention of a number of sampling techniques.

Sampling methods include:

- opportunity sampling

- simple random sampling

- stratified sampling

- cluster sampling

- self-selected sampling

Quantitative Problem Solving (MEI) Specification, 2014, OCR, Sampling methods

Students should become comfortable in discussing the merits and drawbacks of these, and any other methods that may come up in class discussion. For example biology students may talk about using quadrats and sociology students the use of snowball sampling, where subjects in hard-to-access groups (eg drug users) provide researchers with further contacts.

The assessment requirement will be for students to be able to comment sensibly on a method described in a question rather than show theoretical knowledge of any particular technique. They will, however, be expected to know the meanings of the six methods in the list above. A question may, for example, describe a sampling procedure and ask candidates to identify the sampling method that has been used, and to comment on its appropriateness.

Information sources

During the course students are expected to think about possible sources of data, and to use their own common sense and general knowledge.

The assessment of this course does not require students to have collected their own data. However, those who have done so will benefit greatly from the experience and teachers are advised to do all they can to encourage such work.

Some students will be fortunate enough to collect their own data in one of their other subjects. Teachers should look for opportunities for them to share their experience with their classmates.

The pre-release data provides the opportunity for students to carry out pilot studies, based on small samples taken from the main data set.

DATA PROCESSING AND PRESENTATION

In this part of the cycle, raw data are turned into forms which allow conclusions to be drawn that are relevant to the problem under investigation. Much of the work on a traditional statistics course can be found here. In any particular case, some or all of the following steps and techniques may be required.

Raw data

Real data often throw up problems such as:

• outliers

• missing data

• items that are not in the right format.

Dealing with these is often referred to as cleaning data.

In the assessment, students will be expected to deal thoughtfully with outliers, deciding whether or not they should be excluded in the circumstances.

Graphs, charts and summary measures

Once data have been cleaned, it is common for them to be displayed in ways that highlight their main features, for example their distribution. The facility to do this electronically is now widely used.

Consequently, the focus of the assessment will be on the selection of suitable data displays and their interpretation. Statistical diagrams will be given in many examination questions and students will be expected to draw conclusions from them.

Data are usually represented by summary measures such as mean and standard deviation (or median and interquartile range) and, for bivariate data, a correlation coefficient. The assessment will not require students to carry out lengthy calculations by hand; they will either be given or, in the case of small data sets, obtained electronically. The emphasis instead will be on what information these measures bring to solving the problem in hand. In most subjects particular index numbers are important. Students are not required to know all of these but they are expected to be able to interpret a given index number as a summary measure.

The choice of what data should be collected was made in the light of the model adopted for the problem. At this point, with data displays and summary measures available, it may be possible to decide whether the model has turned out to be appropriate. If, for example, a histogram shows the data to be bimodal, a Normal distribution model could be ruled out at this stage.

Parameters and inputs, calculations

The next stage in the process involves using the sample data to estimate the inputs required by a model or the parameters for a distribution. Such figures will often have been calculated already as summary measures. They are now put to use within the model or distribution.

The Normal distribution

Students are expected to be able to use the Normal distribution as a model and to know when it is appropriate to do so, and when not, for example for a skewed or bivariate distribution. Examination questions will test this.

The Normal distribution is also covered in *Introduction to Quantitative Reasoning* and the specification statements there (IQR s18-21) are subsumed and extended in this specification. Students will be expected to calculate z-values and to use them to estimate population proportions; this may involve the use of tables but it is hoped that many students will have calculators with the appropriate statistical capability. The Normal hypothesis test for a population mean is not required.

The χ^2 test

Sometimes the data displays and summary measures will give enough information to solve the problem. Other times a more formal analysis is needed, in the form of a statistical hypothesis test. Two such tests are included in the specification, the x^2 test for a contingency table and that using Spearman's rank correlation coefficient. Both these tests are commonly used in other subjects, as are several other tests. So it is important that students learn that the same methodology and language is used in all hypothesis testing, and not just how to conduct two particular tests. Some examination questions will involve hypothesis on samples that have been taken from the pre-release data for pilot investigations.

Bivariate data, Spearman's rank correlation and product moment correlation

Students will be familiar with the bivariate data, but possibly not by name, from GCSE and from *Introduction to Quantitative Reasoning* (IQR s14). They will know about scatter diagrams, correlation and lines of best fit (drawn by eye). These ideas are extended in this specification to include using Spearman's rank correlation coefficient as a test statistic.

Spearman's test is usually used with small data sets and so is suitable for many student investigations in a variety of subjects. Students who have met it in, say, geography will have interesting examples to share with the class.

This specification is also about large data sets. Typically these are stored electronically in a format that makes the product moment correlation coefficient readily available. Students are expected to interpret this measure.

Students are not expected to calculate or use least squares regression lines.

<div style="border:1px solid red; padding:10px;">

EXAMPLE FROM THE SPECIMEN PAPER

Data collection and the x^2 test

3. A psychology student has devised a test for how conformist people are. It places each person in one of three categories: C, D and N.

   C People who will always do what they are told without questioning it.

   D People who will only do what they are told without question if they respect the person asking them

   N People who will never do what they are told without first questioning it

   The student tried the test out on groups of rugby followers attending Six Nations matches. He wanted to know if there are differences between the nationalities involved. The test was carried out in bars near the matches and the student bought drinks for some of the participants to get them to take part.

   The results are summarised in this table.

Team supported	C	D	N
England	0	0	20
France	6	8	6
Ireland	5	7	8
Italy	6	8	6
Scotland	3	11	6
Wales	10	5	5

   (i) What sampling method did the student use?   **[1]**

   (ii) The information for England is an outlier. Explain what this means?   **[1]**

   (iii) Use the data, with the outlier excluded, to carry out a   test, using a 5% significance level. State clearly your null and alternative hypotheses.   **[9]**

   (iv) Make two comments on the conduct of the experiment.   **[2]**

</div>

REPORTING/INTERPRETATION

The last stage involves reporting the results of the investigation.

Students are likely to present write-ups or formal reports following investigative work in other subjects, but this is not an assessment requirement of this specification. However, questions will test the individual elements that would go into such a report.

Interpretation

Students will be expected to relate a proposed solution to the original problem.

This includes the outcomes from a hypothesis test when one is used; an answer of "Significant" or "$p = 0.02$" will not be sufficient. The question "So what?" must be addressed.

Students need to understand that statistics does not end with a number or a diagram. Interpretation is essential and this will be emphasised in the examination questions.

Checking

There are two sorts of checking that may be required in this specification.

- Students will be expected to apply common sense and general knowledge to detect an unreasonable proposed solution.

- They will also be expected to comment critically on the data that have been used with regard to their likely accuracy and to sources of bias. During this course, students should learn to question the accuracy of data rather than taking them at face value.

Evaluation

Students may be expected to question the validity of the approach that has been taken to solve a problem, including any modelling. If necessary they may be asked to suggest improvements or even a completely different approach. Alternatively the proposed solution may require additional data or it may be good enough for the purpose.

When a question is set on evaluation, there is unlikely to be a single correct answer. Students will be expected to reply thoughtfully. During the course students will benefit from meeting and discussing a number of examples.

Communication

Students may need to explain the results of an investigation or the solution of a problem to people who have not been involved at all. To do this they need to know not just the technical vocabulary covered in this specification but how to convey its meaning in everyday language. This is something that teachers themselves do on a daily basis.

EXAMPLE FROM THE SPECIMEN PAPER

Reporting, modelling and the Normal distribution

2. The residents of a small town find the cost of their house insurance has gone up. The insurance companies say that their risk of flooding is high. A river flows through the town. If the river level rises by more than 2 metres above a given datum level, the town will flood.
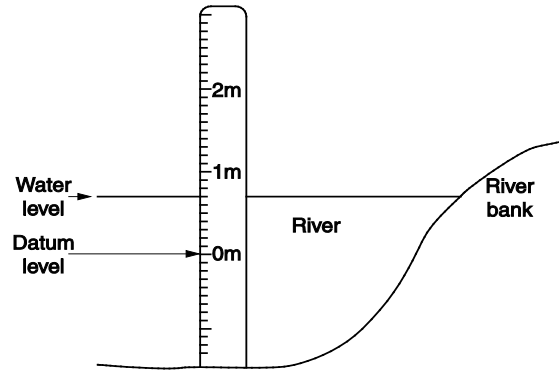

The residents decide to investigate the situation so that either they can refute the insurance companies' argument, or they can claim funding for flood defences.

The town's archives have records going back 120 years giving the greatest height of the water above the datum level each year. The mean of these heights is 0.61 m with standard deviation 0.48 m. One of the residents tries using the Normal distribution to model this situation.

(i) Show that using the Normal distribution as a model suggests that the flood risk in this town means that the flood can be described as a 'Once in 500 years event'. **[7]**

Another resident points out that she has experienced three floods and she is not very old. She draws this frequency diagram using the records of the greatest heights above the datum level from the town's archives.

Height in m above the datum level
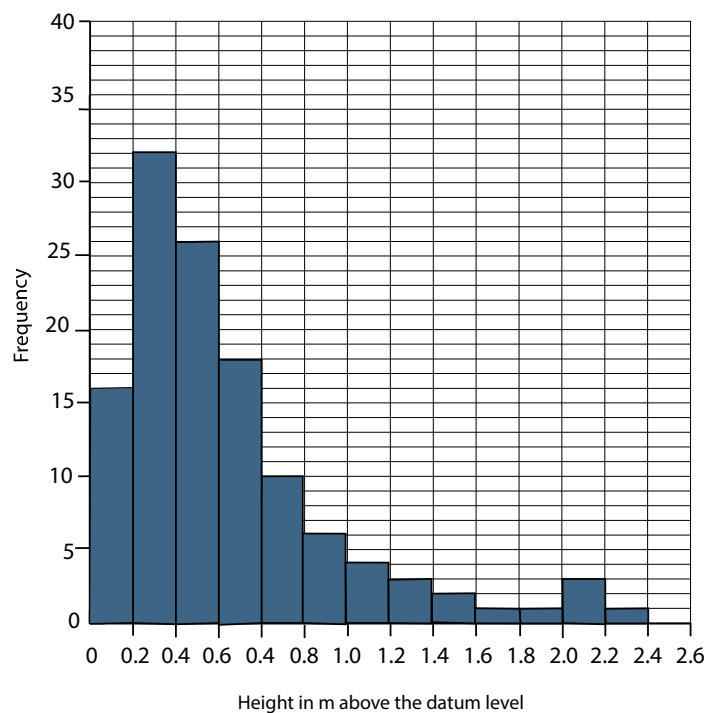
(ii) Taking this new information into account, comment on the work the residents have done so far and advise them how they should proceed. **[3]**

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/

www.ocr.org.uk

OCR Customer Support Centre

General qualifications
Telephone 01223 553998
Facsimile   01223 552627
Email general.qualifications@ocr.org.uk