

Teaching Statistics 1

Perceptions of S1

“Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.” (H G Wells). Statistics are widely used in many fields of study and employment and statistical literacy is a requirement for a full understanding of the modern world. However, despite the subject’s obvious usefulness and applicability, some students do not seem to see the point of statistics. “Why is A Level statistics so boring?” is a discussion thread on the NRICH website.

Students’ first experience of A Level statistics is in Statistics 1; their negative reactions may be because this is a surprisingly difficult module to teach. This is in spite of the perception, in some quarters, that it can be easier for students to do well in S1 than in M1. This paper is written to offer some practical advice; we hope it will be especially useful to the many teachers who find S1 challenging to teach. It is written for the MEI S1 module but the principles are also applicable to other S1 modules.

Transition Units: C1 and S1

It is inevitable that students beginning A Level Mathematics will have differing amounts of knowledge and understanding of the GCSE content. Even if everyone had taken GCSE Higher Tier (and they have not) the knowledge of some students will be patchy or insufficiently secure to build on.

For this reason, the specifications for both C1 and S1 contain some topics that some, but not all, students will have covered at GCSE (this contrasts with Mechanics and Decision Mathematics which are not covered at all at GCSE). This overlap is necessary for a smooth transition to A-level. The C1 and S1 textbooks support the specifications and so include material that will be familiar to some students.

It is vital that students understand this work well but teaching something that students have done before is not the same as teaching something which they have never met. Either they already understand it (or think they do) or they know they don’t understand and believe it to be difficult.

A brief overview of S1 content

The material which overlaps with GCSE is almost entirely in data presentation, statistical summary measures and probability. By contrast, hypothesis testing is completely new.

Data presentation

- Your students will have done some work on data presentation at GCSE; indeed, they are likely to have met bar charts, pie charts and stem and leaf diagrams at KS3.
- Students may not have made connections between the types of data and appropriate ways of presenting them and the idea of skewness is likely to be new.
- Although your students are likely to have met cumulative frequency graphs and boxplots at GCSE, identifying and dealing appropriately with outliers will be new to most of them. A significant number of students find it difficult to draw and interpret cumulative frequency graphs and boxplots correctly.
- Drawing and interpreting histograms is also a topic where students can have misconceptions that you need to tackle.

Statistical summary measures

- Your students will have met averages before but may not always appreciate when it is appropriate to use the different measures.
- Although students are used to using averages, they may be less familiar with using them in connection with measures of dispersion, which is how they are most usually used in S1, and beyond.
- Variance, standard deviation, mean square deviation and root mean square deviation are new to students and the examiners' reports for S1 show that confusion and careless mistakes are common when calculating them. Students who have taken GCSE Statistics will have come across standard deviation, but with divisor n , rather than $n-1$. They may find the change more acceptable if you explain that divisor $n-1$ gives a more natural way to estimate the population variance or standard deviation. The note on page 147 of the specification gives helpful background information; a Newsletter Extra from 2004 goes into greater detail and can be found on www.mei.org.uk.

Probability

- Your students will have some familiarity with concepts of probability from GCSE, and before.
- Although students are often successful at working with tree diagrams, Venn diagrams are not always used as successfully.
- Your students are likely to find conditional probability difficult and they may be unclear about whether or not events are independent.
- Sometimes there is more than one way to do a probability question but some methods are more complicated, and so more likely to result in errors, than others.
- The binomial distribution will be new to most students. Students who have taken GCSE Statistics are likely to have met simple cases of the binomial distribution but are unlikely to have studied it in the depth demanded by S1.

Hypothesis testing

- Hypothesis testing involves new and challenging concepts and is likely to need a considerable time allocation for your students to understand it fully. In particular, many students find it difficult to relate the conclusion to the original situation.
- Your students are likely to find critical regions particularly difficult but it is vital that they understand these if they are to be able to use hypothesis tests successfully.

Using statistics

Even students who are familiar with the techniques from GCSE can find it hard to make appropriate connections between the statistical techniques they are using and the real-world implications and interpretations. Understanding what the answer means is the key to statistical literacy and, consequently, an important aspect of statistics at all levels.

Implications for teaching S1

The fact that some students will already have some knowledge from GCSE means that it should not be necessary for every student to work through every exercise in the textbook. The time you allocate to teaching each section should not be proportional to the number of pages devoted to it in the S1 textbook. Some of the new topics in S1 are conceptually quite difficult, e.g. conditional probability and hypothesis testing; spending too long on the easier aspects can mean that you are left with insufficient time for these more difficult concepts. This is especially likely to happen if you are teaching S1 as the last of three AS modules. It is particularly important to allow sufficient time for students to understand hypothesis testing. The work done on this in S1 lays the foundations for future statistics modules; it is also useful for many students who end their formal studies of statistics with S1 but who meet hypothesis testing in their studies of science or social science.

From the start of the S1 course, it is important to gauge your students' understanding of concepts that they may have encountered before; this can be especially difficult with students you have not taught before. You may find that engaging in activities that promote discussion can help you to get a feel for what they already know and, hence, where to concentrate your teaching. The multiple choice tests in the online resources can also be useful for checking your students' understanding.

Possible orders for teaching S1

You will want to use your professional judgment to decide whether it is better for your class to start with concepts which are familiar and build on these or whether they will find it more motivating to start with something new. Three possible orders for teaching S1 are given below. The time allocations are only rough guidelines; different times will be needed for different classes. Additional time will be needed for students to practise and consolidate concepts, for internal assessments and for exam preparation.

Starting with the binomial distribution

This starts with a new topic and can work well with classes that already have a strong understanding of the statistical content of GCSE Mathematics Higher tier. However, it can also work well with students who have gaps in their knowledge from GCSE as it allows a fresh start with a new topic, and reviews tree diagrams in the first few lessons. This could, therefore, be a good way to start with a class who have a mixture of levels of previous understanding or where you do not know their level of understanding. It allows the students to start on something new and feel as though they are making progress before going back over previous learning.

Content	Approx time (hours)	S1 textbook chapter
Introduction to S1	2	Practical activity which can be modelled using the binomial distribution.
The binomial distribution	6	Chapter 6 (formal ideas of expectation have not been covered yet but most students are ready to accept $E(X) = np$ as intuitively fairly obvious)
Probability	6	Chapter 3
Exploring data	5	Chapter 1
Data presentation and related measures of centre and spread	4	Chapter 2
Discrete random variables	6	Chapter 4 (this can be conceptually difficult for some students because of the generality – refer back to the more specific binomial distribution).
Further probability	6	Chapter 5
Hypothesis testing using the binomial distribution	8	Chapter 7

Starting with data presentation

This follows the textbook order and begins with a review of some of the topics from GCSE Mathematics, emphasising the use of these techniques in making sense of data. It could be useful with students who are less confident with their understanding of the GCSE content, enabling them to grow in confidence gradually. This would also be a suitable approach with a new class, who may not know each other, starting on S1 as their first module.

Content	Approx time (hours)	S1 textbook chapter
Introduction to S1	2	Exploring a set of data, including gauging students' understanding of previous work.
Exploring data	5	Chapter 1
Data presentation and related measures of centre and spread	4	Chapter 2
Probability	6	Chapter 3
Discrete random variables	6	Chapter 4 (this can be conceptually difficult for some students because of the generality – refer to more specific distributions that they can work out the probabilities for).
Further probability	6	Chapter 5
The binomial distribution	6	Chapter 6
Hypothesis testing using the binomial distribution	8	Chapter 7

Starting with probability

The strongest students often find the introductory probability concepts from GCSE straightforward. However, conditional probability and questions which make use of permutations and combinations will extend their thinking. This makes this approach suitable for a class of very able students who are likely to enjoy the challenge of doing difficult questions early in the course.

Content	Approx time (hours)	S1 textbook chapter
Introduction to S1	2	Exploring a situation involving probability; this will help to gauge students' understanding.
Probability	6	Chapter 3
Further probability	6	Chapter 5
The binomial distribution	6	Chapter 6 (formal ideas of expectation have not been covered yet but most students are ready to accept $E(X) = np$ as intuitively fairly obvious)
Exploring data	4	Chapter 1
Data presentation and related measures of centre and spread	3	Chapter 2
Discrete random variables	6	Chapter 4 (this can be conceptually difficult for some students because of the generality – refer back to the more specific binomial distribution).
Hypothesis testing using the binomial distribution	8	Chapter 7

Engaging your students: how to start?

First impressions matter. What impression of statistics will students get from their first lesson on the S1 module? Will they think it is all about being able to do a set of techniques? Will they think it is just redoing things from GCSE? Will they recognise that statistics is about dealing with real problems and the associated data? Why not start with some data or a practical situation? Three suggestions which you could use as the first lesson are given in Appendix 1; each set of suggestions corresponds to one of the possible orders for teaching S1 given above.

Helping to gauge and develop understanding of topics from GCSE

Some further ideas to promote discussion and deeper understanding of material that has been studied at GCSE can be found in the DfES Standards Unit materials. Two of them are described below.

Chapter 1 – exploring data

Understanding mean, median, mode and range

Students will have covered these concepts at KS3 and GCSE. If you think some of your students have a weak understanding of what these measures are, there is an activity matching bar charts to summary statistics in the materials produced by the DfES Standards Unit (Improving learning in mathematics).

Chapter 2 – Data presentation and related measures of centre and spread

Interpreting frequency graphs, cumulative frequency graphs, box and whisker plots

These topics are typically covered at GCSE and students need to understand them thoroughly to succeed in S1. To check and develop their understanding you may find it helpful to use the materials produced by the DfES Standards Unit (Improving learning in mathematics) which include a matching activity.

Enrichment material

With pressures of time and targets to meet, it can be tempting to focus exclusively on teaching students the techniques they need to pass examinations. However, enrichment need not be seen as an optional luxury.

“Through art and science in their broadest senses it is possible to make a permanent contribution towards the improvement and enrichment of human life and it is these pursuits that we students are engaged in.” Frederick Sanger

As well as helping students to understand the subject matter better, looking at real-life examples, applications and misconceptions can help them to see that statistics is an interesting and enjoyable subject. Statistical graphs used in the media and in election campaigns can be a useful resource for promoting discussion, and so can the national lottery. There is a wealth of resources on the internet; you will find a few of these briefly described in Appendix 2.

Appendix 1

Ideas for the first lesson of S1

Introducing binomial probability: Can you guess what is in the teacher's mind?

Equipment/resources

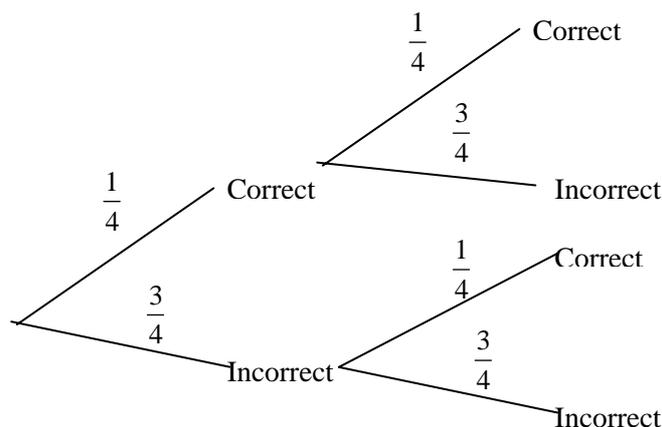
This requires a pack of ordinary playing cards.

Instructions for the initial activity

- Make sure the students know the names of the four suits.
- Shuffle the pack thoroughly.
- Tell them that you are going to pick the top card and look at it; you will not show it to them but they should guess which suit it is and write it down.
- When you have looked at the card, put it face down on the table.
- Repeat until you have looked at 5 cards.
- Tell the students that you will tell them what the suits were but that it will not be possible for them to know how good their guesses were unless they know how likely they were to get, say, 3 correct just by guessing randomly.

Follow up

- To work out the probabilities of getting 0, 1, 2, 3, 4 or 5 correct by guessing randomly, start by drawing a tree diagram, as shown below.



- The tree diagram becomes too time-consuming to draw and then it is time to look for patterns in the answers – this could lead on to the binomial distribution in the next few lessons.
- Checking answers by using the fact that the probabilities should add to 1 will connect with discrete random variables.
- When the probabilities are calculated, it is time to reveal what the cards were and to see how many each student got right. How do the probabilities of getting 0, 1, 2, 3, 4, 5 correct compare with what happened?
- You could revisit the data from this when doing hypothesis testing: how many cards would a student need to get correct before we concluded that he wasn't just guessing and getting them right by chance?

Points to be aware of

Some students may raise the issue of whether probabilities should change because you are not replacing the cards – this will introduce the idea of conditional probability for later use. In this case, because they do not know what cards have come up earlier, their probability of guessing correctly is still $\frac{1}{4}$.

Variations on this theme

1. If you think students will find it confusing to consider whether the probabilities should change or not, you can use the score on a die you throw instead – perhaps throwing it before the lesson and writing down each result on a small card. You can then look at each number, one at a time, while they try to guess it. Using a coin is not a good idea as the probability of being correct is equal to that of being incorrect and this obscures the pattern in the probabilities.
2. If you think that students will find it too strange to try and guess what is in your mind, <http://www.stat.auckland.ac.nz/~iase/publications/18/BOOK1/A4-6.pdf> gives a suggestion for a taste test using two types of cola; the tasting uses 3 cups with two the same and one different and the taster has to say which one is different (this means the probability of getting it right by guessing is $1/3$). This is how tasting panels work in the food industry.

Different ways of presenting data: UK Population

Equipment/resources

You need a set of suitable data; you could use any suitable set of data that allows a range of techniques to be used but download them yourself to avoid wasted time as students search for something suitable.

Estimates of population by age and sex for local authorities in the UK can be downloaded from <http://www.statistics.gov.uk/STATBASE/ssdataset.asp?vlnk=8611> . Download this data set and print it out; you may want to reduce the data set to make it more manageable, particularly with less confident classes.

Instructions for the initial activity

- The task is for students to produce a brief news story highlighting something from the data; it should contain at least one suitable diagram and may contain the results of statistical calculations.
- Students can work as individuals or as small groups of two or three; this can depend on the size and confidence of your class.
- You can start with a discussion of some possible stories from the data and what kinds of diagrams would be suitable.
- ICT can be used (or not) depending on availability.
- The initial activity could be set as homework.

Follow up

- Students present their “story” to the class and say why they chose the statistical techniques used.
- This should give you some information about what the class already knows to help you in planning the next few lessons appropriately.

Variations on this theme

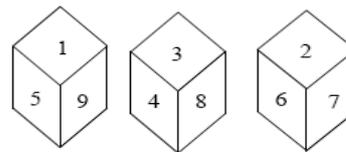
1. The National Statistics home page www.statistics.gov.uk gives some ideas of data that are making the news.
2. http://www.wsf.org.uk/docs/GHS_ng.doc gives you statistics on people’s participation in sports; you could do some comparison with sixth form students in your school/college.

Introducing probability: the strange dice

Equipment/resources

3 dice numbered as in the diagram (or several sets of such dice)

Opposite faces are numbered the same.



A B C

Instructions for the initial activity

- Two players roll different dice; the highest score wins.
- Which is the best die to have?
- Students could start by doing this practically then move onto a theoretical consideration using a sample space or they could start with a theoretical argument.

Follow up

A suitable way of finding the probability of die A beating die B is to use a sample space:

Winner		score on A					
		1	5	9	1	5	9
Score on B	3	B	A	A			
	4	B	etc				
	8						
	3						
	4						
	8						

- Students often find it surprising that A is more likely to win than B, when A plays B; B is more likely to win than C, when B plays C; but C is more likely to win than A, when C plays A.
- What would happen if A, B and C all played each other at once?
- Which is the best if you throw twice and add both scores?
- Is it possible to have 4 strange (intransitive) dice?

Variations on this theme

1. Two dice (standard or numbered differently, depending on your class). Students should write 5 numbers down to play dice bingo. The number they are allowed to cross off is the sum (or difference, or product) of the scores when you roll the dice. Which are good numbers to choose for a bingo card?
2. Write a list of events that students should be able to find the probability for, allowing space for the events to be cut out. Students should put the events in order of probability and justify their decisions.

Appendix 2

Enrichment material

Questions posed by historical gamblers

When the theory of probability was being developed, some wealthy gamblers wrote to mathematicians like Pascal and Galileo asking which outcomes would be more likely. These questions, originally answered by these famous mathematicians, are accessible to A Level students.

- Which is more likely: rolling at least one 6 in four throws of a single die or rolling at least one double 6 in 24 throws of a pair of dice? (The Chevalier de Mere to Pascal)
- When rolling three dice, there are 6 possible combinations that give a total of 9 and the same for a total of 10. Why does a total of 10 occur more often than 9? (The Grand Duke of Tuscany to Galileo)
- Which is more likely?
 - At least one 6 when rolling six dice
 - At least two 6s when rolling twelve dice
 - At least three 6s when rolling eighteen dice (Pepys to Newton)

Some websites you may find useful and what they say about themselves

<http://exploringdata.cqu.edu.au/> “This website contains activities, worksheets, overhead transparency masters, datasets and assessment to support data exploration. It also contains an extensive collection of articles designed to enhance the statistics knowledge of the teacher.”

<http://www.rsscse.org.uk/ts/> “Teaching Statistics seeks to help teachers of biology, business studies, economics, geography, mathematics, the sciences and the social sciences etc. by showing how statistical ideas can illuminate their work and how to make proper use of statistics in their teaching.”

http://chance.dartmouth.edu/chancewiki/index.php/Main_Page “Chance News reviews current issues in the news that use probability or statistical concepts. Its aim is to give the general public a better understanding of chance news as reported by the media and to allow teachers of probability and statistics courses to liven up their courses with current news.”

<http://www.mis.coventry.ac.uk/~nhunt/home/about.html> “The DISCUSS project is a modified and supplemented web-based implementation of the widely used DISCUS materials for teaching elementary Statistics. It aims to integrate the powerful interactive capabilities of Microsoft Excel with the convenience of web-based resource materials using Microsoft Internet Explorer.”

<http://www.stat.auckland.ac.nz/~iase/islp/mission> “The mission of the International Statistical Literacy Project (ISLP) is to support, create and participate in statistical literacy activities and promotion around the world. To facilitate communication among many countries and projects, we support this webpage, which is a forum where those interested in acquiring or providing statistical literacy can meet (in a virtual sense), exchange needs, information and resources, and learn to disseminate statistical literacy in their communities.”

<http://www.whfreeman.com/eeseen/eeseen.html> “One of our goals in developing the Encyclopedia of Statistical Examples and Exercises (EESSEE) was to provide a wide variety of timely, real examples with real data for use in Statistics classes.”

<http://www.swivel.com/> “Swivel is a website for curious people to explore data.”