

PART 1 A vision for statistics in schools

Chapter 1

What should all students understand and why is it important?

1.1 Who needs statistics anyway?

When planning a statistics curriculum for all, the result should be dependent on the end users. It is convenient, if reductive, to categorise users of statistics based on the level of integration of statistical concepts into their daily lives. Of course, the careful statistician will recognise that although this is a helpful process for building a model of essential skills, any individual member of the population is unlikely to fall neatly into a single descriptive group.

| | |
|--|--|
| Expert statisticians | People who work with data routinely and analyse the results directly. |
| Functional statisticians | People who receive data and statistics second-hand and use them for pre-defined, possibly standardised, tasks. |
| Occasional (and perhaps unwilling) statisticians | People who encounter analysis based on data and use this to make decisions. |

Table 1.1 Users of statistics

Expert statisticians include scientists, academics, engineers and people who might calculate a p -value without a second thought. These people are capable of being utterly fascinating or staggeringly boring at parties depending on the perspective of the observer. They have probably gained many of their statistical skills as part of a degree-level qualification. For this group, much of what they have learnt in school about statistics has been supplanted by sophisticated new techniques supported by powerful software packages. Despite this, examples exist of the use of data leading to unsafe conclusions: many examples are detailed by Dr Ben Goldacre at www.badsience.net and provide an excellent source of stimulating material for interested students.

The p -value problem

It is not just ‘bad’ science that produces potentially dangerous results of course. At the heart of the scientific method is the hypothesis test, a confirmatory test that essentially examines whether the data are unlikely to have happened by chance. If they are unlikely to have happened by chance, then this indicates that there is something interesting going on.

But how unlikely is ‘unlikely’? The standard for publication in most cases is about 5%, so in fact somewhere around 1 in 20 scientific studies are drawing their conclusions from data which just happened to come out looking a bit interesting by chance. This is unlikely for any particular individual study, but there are thousands and thousands of studies published every year.

This phenomenon is known as ‘the p -value problem’, and there is an ongoing debate in the scientific community about the level of confidence there ought to be in a conclusion before publishing it.

For future expert statisticians, a secondary-level course in statistics does not need to introduce them to the sophisticated techniques which will be learned later in the context of their area of study. It should, however, prepare them to be critical when engaging in analysis and experimental design, introducing them to the idea of selecting appropriate techniques rather than reaching for the first tool in their statistical toolbox. Students should gain a sense of the limitations of statistical analysis and the need to allow the wider context to inform any decisions and conclusions.

Functional statisticians are those who work in jobs where spreadsheets appear with sometimes alarming regularity in their email inbox. These people may have little experience beyond what they studied at secondary or college level, but will be expected to make judgements and predictions based on trends, averages and raw data. Many teachers fall into this category since student progress is increasingly judged by recorded data rather than professional insight. Often, teachers and senior leaders are not adequately supported to make the data-based predictions of student outcomes currently required. It is not uncommon in UK schools for teachers to be pressured to justify why an individual student has achieved below a target grade that was based on a third-party model of progress for ‘equivalent’ students. This sometimes happens even when a class of 30 students or a cohort of 200+ has achieved or exceeded the overall targets predicted by the model. This lack of understanding that a model may be used on a macro scale but is unsafe as a predictor of individual performance leads to considerable stress being inflicted on teachers and students due to a fundamentally incorrect assumption. It would be naïve to assume that this kind of misinterpretation is limited to the teaching

profession alone! For functional statisticians, a secondary course in statistics should provide them with knowledge of the limitations of data for inference and the ability to defend themselves against unsafe conclusions.

Occasional statisticians are those who work in jobs where they do not encounter data directly and have little need for it in their home life. These are the people most let down by the traditional approach to learning statistics. Every day they are bombarded with information and data that has been abused and wrangled into TV ads. They read web pages and newspaper articles that tell them their risk of cancer will go up if they drink too much red wine, and down if they eat vast quantities of kale. They vote in elections after watching politicians and journalists quote seemingly contradictory statistics demonstrating both that everybody is better off and that society is disintegrating around them. During their time in education they may have been unwilling to engage with a statistics curriculum which appeared irrelevant to, and disconnected from, their daily lives, so they do not have the skills needed to critically evaluate these statements and arguments. For the occasional statistician, secondary education must provide the skills to recognise when statistics are being used well, and when data are presented badly. It is often said that you can prove anything with statistics, but this is not the case. You can *imply* anything by misuse of statistics, and some would suggest it is a right of everybody completing compulsory education that they leave with the statistical literacy needed to understand and critically evaluate statistics and data-based arguments.

In his book *Thinking, fast and slow* Daniel Kahneman described how our brains work against us, convincing us that what we see and experience directly leads to mental models that we then over-generalise, often giving us an inadequate or skewed perception of the world around us. Statistical literacy is the ideal toolkit for exploring this concept in school so that individuals can critically evaluate their own world-view in the future and use data to challenge their perceptions of the world around them.

A secondary curriculum for statistical literacy should meet all the needs of the occasional statisticians, should be flexible enough to provide the functional statisticians with the skills they will need to make sound judgements, and should lay the conceptual foundations for the expert statisticians in anticipation of more technical study later.

1.2 Towards a new type of statistics curriculum

Many current curricula appear to work from a different assumption: that everybody will be an expert statistician. Students' journey through statistics in school is a trajectory subdivided with this goal in mind.

Because statistics is treated as a mathematical discipline, progression is based on increasingly complex calculations and diagrammatic representations. Only superficial levels of interpretation are necessary and inference is treated as a higher-order skill encountered only during further study. Recent research (Makar & Rubin 2014) into informal inferential reasoning has found that students at primary level are capable of discussing likelihood based on simple data in familiar contexts and, with carefully planned tasks, can make reasoned decisions despite few sophisticated techniques being available to them.

So, what do students need to know? Cobb (Steen 1992) argued that any introductory course should explore the following ideas.

- 1 The need for data.
- 2 The importance of data production.
- 3 The omnipresence of variability.
- 4 The quantification and explanation of variability.

This provides us with a good starting point for considering the aims of our curriculum, but to decide on the focus and content a little more detail is needed. Building on this work, Garfield (1995) suggested that a college-level initial course in statistical literacy should be based on the following foundations.

- 1 The idea of variability of data and summary statistics.
- 2 Normal distributions are useful models though they are seldom perfect fits.
- 3 The usefulness of sample characteristics (and inference made using these measures) depends critically on how sampling is conducted.
- 4 A correlation between two variables does not imply cause and effect.
- 5 Statistics can prove very little conclusively, although it may suggest things, and therefore statistical conclusions should not be blindly accepted.

The most striking thing about many research papers on statistics education is the lack of discussion of the merits of training students to perform standardised statistical calculations and produce formalised graphical representations; yet most curricula rely on this type of learning to form the backbone of their content. So why is this still the case when expert opinion has been pointing to a different approach for several decades, especially when the underlying ideas are accessible to young students if technical expertise is not introduced as a barrier?

There are several things to consider. Firstly, statistics is a relatively new discipline compared to much of school mathematics. While the cutting edge of mathematics is constantly advancing, the bulk of the mathematical content taught in modern secondary classrooms can trace a direct line back to luminaries like Euclid and Pythagoras over 2000 years ago. Statistics on the other hand has developed much more recently, and the fundamental principles have advanced rapidly along with the rise of computing power. Taking the study of probability (a key building block of statistical inference) as an example, while there is historical evidence that the ancient Egyptians enjoyed dice games around 3500 BCE, it was not until Blaise Pascal and Pierre de Fermat began investigating problems of probability in the 1600s that a formal mathematical theory began to develop (Lightner 1991). Modern computing methods allow the creation of probabilistic simulations in the classroom unimaginable just a few decades ago. New technology has transformed the interaction between probability and statistics and allowed powerful techniques of analysis to be applied to the growing discipline of statistical inference from so-called ‘big data’, the collection of vast databases of stellar observations, purchasing patterns and even search histories.

Modern computing allows the almost instantaneous generation of graphical representations of data, some of which conform to none of the standardised charts and graphs that people have become familiar with during schooling. This means that unless a specific pedagogical value can be attributed to training students to construct by hand bar charts, stem-and-leaf diagrams, and all the other graphs that most school leavers will currently be familiar with, it is hard to justify spending large amounts of time doing so.

Most research into the development of graphing techniques advocates modelling with data (Lehrer 2007) to answer simple questions posed by students, with time spent inventing meta-representations (diSessa 2004) – charts and tables devised by the students that do not conform to any pre-defined method. Over an extended period of time, students discuss and refine their models, eventually moving towards more standard representations appropriate to the type of data they are working with. While this process necessarily begins at primary level as students move from object-based representations (sorting objects into groups and arranging them physically to create ‘object graphs’) to iconic representations (using drawings and stickers to represent data before moving towards crosses and bars), it should continue into secondary school as students begin to explore new kinds of data such as continuous grouped data and work towards developing the more formal representations against which they will be assessed.

A far more important skill to focus on is selecting *useful* representations when a computer can dumbly generate multiple diagrams without consideration of their appropriateness for the data. On top of this,

without the ability to interpret unfamiliar representations in the form of infographics and the like, useful information will be obscured. It is incredibly difficult for school curricula to keep up with advances in the field, and any changes that are made require substantial investment in training for the teaching workforce.

A second reason for the disconnect between expert recommendations and practice is due to the status of statistics as an aspect of mathematics. In one key way statistical literacy is fundamentally different to other areas of applied mathematics, especially at secondary level: it is completely reliant on context and interpretation. The study of statistical literacy can be thought of as the study of numbers in context. It is the context in which data exists that makes sense of statistical results, and ignoring the context reduces the study of statistics to an exercise in following a set of recipes that can be mechanically repeated without understanding their purpose. Unfortunately, as statistical literacy occupies a place in the mathematics curriculum, the focus of study is on the mathematical characteristics that can be refined to formulae and processes, rather than the more difficult-to-define interpretive aspect that relies on discussion and a broad understanding of the wider context. Defining the study of statistical literacy as a purely mathematical discipline is akin to defining the construction of furniture as the assembly of a flat-pack kit. Most people can create a serviceable cabinet by following instructions and putting together some pre-made pieces with standard tools, but the end product lacks the quality of a handmade item and almost always has to fit in a space whose dimensions are at odds with the size of the finished article.

The third reason is assessment. Education is obsessed with the measurement of progress, and it is convenient to create a hierarchy of ever more complicated calculations on which to base judgement of students' ability to 'do statistics'. This is an additional hangover from the treatment of statistics as purely mathematical objects. It is rare in formal, summative mathematics assessment that the sophistication of a student's response to a question is measured. The most prevalent form of assessment question has a single 'correct' answer with perhaps several possible methods of solution available, all receiving equal credit. With statistics, however, requiring a single correct solution may strip the context and the nuance from the data. This is peculiar to mathematics, as most other subjects in school are much more comfortable with the idea of students answering questions with differing levels of sophistication, all of which gain a reasonable amount of credit. Consequently, while the mechanics of calculating statistics may well best be taught by maths teachers, the interpretive aspect and eventual assessment would arguably be better delivered in subjects where the context is key, such as geography, biology and psychology amongst many possibilities.

1.3 The ability to interpret diagrams

Because of the increasing ease with which data can be presented using computers, students must leave school with sophisticated skills in the interpretation of data presented in diagrammatic forms. Currently curricula tend to focus on a small number of key ideas:

- types of correlation (strong/weak, positive/negative)
- identifying the modal value/class
- identifying outliers
- comparing range / interquartile range
- identifying the median and calculating the mean.

While these are straightforward to assess, they are almost useless as procedural skills when classroom practice meets the messy real world. Take scatter graphs for example: it is rare that the kind of neat and tidy examples of positive or negative correlation found in mathematics classrooms exist in data collected in the real world. Far more likely is that a scatter graph will have regions in which there seems to be a stronger correlation, and regions where the association appears less clear-cut. In these cases, the skill is in deciding whether the lack of clarity is a result of the natural variation in the data distorting a genuine relationship, subpopulations in the data with different strengths of relationship, or a coincidental pattern that appears to show association where none exists. The ability to consider these different possibilities allows individuals to make informed decisions when such data is presented alongside a specific interpretation. In many cases the perspective or bias of the person sharing the data will have an impact on the given interpretation, and it is essential that this is always considered.

1.4 An understanding of variability

Understanding variability is fundamental to understanding what is going on with statistical processes and making good judgements. Variability is present in all stages of statistical investigation and has an impact on the entire process from data collection through to any eventual interpretation. The job of a statistician is to understand the implications of variability in the data and to minimise the aspects of it over which there is a degree of control. It is important to identify the types of variability that can occur in data. Firstly, variability caused by inaccuracy in measurement can be minimised by well-designed methods of data collection but never eliminated entirely. Secondly, some variability is inherent in the object of study as most real-world processes are not deterministic. This aspect

of variability provides important information for the statistician, and the relative size of this variability in relation to the context of the data will have a big impact on any conclusions that are drawn. Finally, there is an element of variability caused by sampling from a larger population or populations, which again has an impact on the quality of the inferences that may be drawn. Many classroom activities involve comparing two related data sets, such as heights of girls and heights of boys. Students are encouraged to compare measures of average and spread, but they should also question whether they are dealing with a population, which would allow definitive statements of comparison to be made, or a sample, with an underlying variability that may make conclusions unreliable or demand further investigation.

1.5 Sampling and populations

Statistics can be used as parameters for populations, allowing us to make definitive statements about the populations' composition or to make reliable comparisons between them. Unfortunately, due to the way statistical literacy is often taught, inferences made from samples are treated as if they were solid conclusions drawn from populations. The difference in height of boys and girls in class 4B is treated as a valid indicator of the difference in height of all boys and girls in a population with little consideration of the error in the calculated statistics. This is particularly evident in assessment questions where students are asked to make specific comments on differences and similarities in data from diagrams such as boxplots, but are given no opportunity to gain credit by discussing whether the differences are statistically significant based on the sample presented.

It can be a tricky concept as the distinction between population and sample can be related to the language used in the question. 'Does the data show that girls in class 4B are generally taller than boys in class 4B?' treats the data for the class as an entire population. 'Using your data for class 4B, are girls generally taller than boys?' treats the data as a sample, meaning that any conclusions must be treated differently.

Sampling is a fundamental aspect of statistical analysis, with significant consequences for the decisions that can be made and the validity of any conclusions that are drawn.

It is not enough to learn that 'a random sample should be taken to ensure that the data are representative'. It would surprise many people to know that in a lot of cases, a random sample may result in a data set that is significantly less representative. Imagine for example that we want to test the hypothesis that 'students enjoy stats lessons'. Our population is a local secondary school in which students of different ages and abilities are taught by different teachers. The likelihood is that each

student's experience of these lessons will be dependent on both their age (affecting the content being taught) and their teacher. In a truly random sample, where every student has an equal likelihood of inclusion, it is not inconceivable that the vagaries of chance may result in proportionally more students being selected from a single class. If this is the class of a teacher whose inspirational approach to teaching, inspired by Robin Williams, has students standing on their desks and passionately reciting the formula for calculating standard deviation before undertaking a spot of extra-curricular data analysis, an overly positive assessment of the popularity of the subject may result. It is essential that experimental design takes into account variability in sampling to eliminate bias without introducing constraints that could obscure signals in the data.

A thorough understanding of sampling methods would enable students to critique the methodology behind statistical conclusions. The goal should be for students to be able to look behind the data presented and consider whether the conclusion is safe. Many advertisements now give some details of the sample used to justify any claims made, and this can be illuminating to those prepared to look critically at the sample and ask questions such as: how many people were involved? What proportion was positive? Were the participants independent of the product?

On top of this, if students left school with some understanding of the power and importance of double blind trials and randomised control trials, it would give them the great advantage of being able to recognise when poor-quality studies are being presented as definitive research and to assess the validity of the conclusions accordingly.

1.6 Distributions and their shapes

Statistics is often perceived as a method of taking a large set of data and, through a process of mathematical alchemy, turning it into a single, easy to use, numerical value. Currently students are comfortable making direct comparisons between the median values for two data sets and will receive credit in assessments for doing so. They may also make some brief comment on the range or interquartile range shown on a diagram. This process of 'reading the data' (Curcio 1987) is the least sophisticated analysis technique and does not consider the global perspective (the complete data set). Students need to be able to go beyond the values they read from the graph and consider the implications of the shape of the data. How does the spread of the data affect the validity of the median value? How does the scale of the diagram affect the perception of the total spread? Does the data set have a large skew or long tails? The implications of the answers to these questions must be considered when deciding if the calculated statistics are meaningful. If the data is a

sample from a larger population, the effect of sampling variation must be considered too.

1.7 Correlation and causation

Data in the modern world are a genuine commodity, with companies trading massive data sets collected from their customers using tools such as mobile phone apps and web pages. It is not uncommon when buying goods or services on the internet to be forced to decipher the language of pre-sale tick boxes to communicate that as a customer ‘Yes, I agree to the terms and conditions’ and ‘No, I would not be happy for my details to be passed on to carefully selected partners’. This data is often used to identify correlations in order to model and predict consumer habits to target advertising in the future.

The technique of ‘exploratory data analysis’ is a relatively new one that has developed alongside the rise of large, multivariate data sets. Where statistical analysis was once based on creating a hypothesis and then collecting the necessary data to demonstrate if the hypothesis was true or not, exploratory techniques rely on the use of technology to analyse data for interesting results and then seeking to explain what the underlying reason for the results might be.

The *Gapminder world* website (www.gapminder.org) contains multivariable demographic data for countries that can be plotted using bivariate representations, allowing associations to be spotted. The ability to easily identify associations without first hypothesising that one exists makes it more important than ever that students are aware of the old mantra that ‘correlation does not imply causation’. A strong relationship identified between two variables may in fact be the result of a hidden variable that is driving both, or even reflect an underlying selection bias. Some thought must also be given to how outliers and sample/population size affect the conclusions drawn. It is straightforward to find a correlation coefficient using spreadsheet software, but such statistics can be strongly influenced by outliers or subpopulations, and it is important that both visual and numerical techniques are used to gain a clear idea of what is going on.

1.8 What’s changing?

Many of the concepts detailed so far go beyond what is currently taught in classrooms in preparation for end-of-course assessments. This is largely because the current style of assessment does not reflect the experience of statistics that people will have outside the school system and focuses narrowly on the mathematical aspects of the discipline while neglecting the contextual and philosophical aspects.