The Foundations of Chance & Data

Jane Watson

explains how the two fundamental ideas of variation and expectation underpin our teaching of chance and data.

The fortunes of chance and data have fluctuated in the mathematics curriculum in Australia since their emergence in the *National Statement* (Australian Education Council [AEC], 1991) in the early 1990s. Their appearance in Australia followed closely on similar moves in the United States (National Council of Teachers of Mathematics [NCTM], 1989). In both countries the topics, taken together, were given a section status equal to other areas of the mathematics curriculum, such as space or algebra. In Australia this was reflected in the state curricula of many states (e.g., Department of Education and the Arts [DEA], 1993; Curriculum Council, 1998) but not all (e.g., Board of Studies NSW, 1989). In recent years chance and data have held their place in the United States (NCTM, 2000) but in some places in Australia have been diluted by being merged with other parts of the curriculum, for example both chance and data with measurement (e.g., Victoria Curriculum and Assessment Authority, 2005) or chance with number (Board of Studies NSW, 2002). As discussions proceed on national consistency across the state mathematics curricula in Australia, the place of chance and data appears to be in further jeopardy. Whether this is due perhaps to a lack of appreciation of the need for statistical literacy skills in all students who leave school (see e.g., Rubin, 2005) or perhaps to a traditional concern to cater for the elite mathematics talent who will study mathematics at university, is beyond the scope of this article to consider.

If chance and data are to survive and flourish, however, it is essential to elaborate the underpinning foundations that give them identity that goes beyond frivolous activity. Looking behind the activities involving throwing dice and recording outcomes or surveying students in the school to see if they want shorter school days, curriculum statements are usually based on the steps in a typical statistical investigation. These steps include data collection, data representation (e.g., production of tables or graphs), data reduction (e.g., finding means or ranges), and drawing

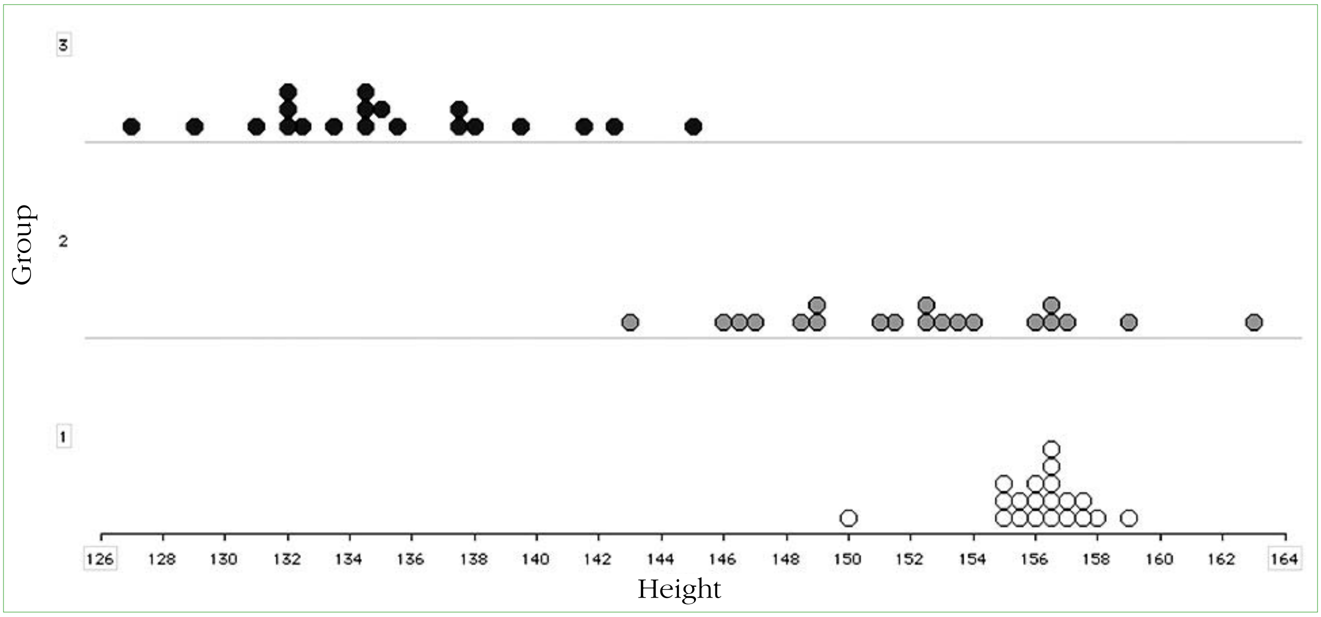

Figure 1. Graphs for height of an individual student [1], a "Grade 6" class [2], and a "Grade 3" class [3].

inferences, supplemented by theoretical and frequency approaches to chance (e.g., AEC, 1991; DEA, 1993). This structure is useful in planning, but often it is not practical to include all aspects in a single lesson or even series of lessons.

Moving one step further back from the investigation structure there are two underpinning central ideas for everything that occurs in relation to chance and data. The most fundamental idea is that of variation. Without variation not only would the world be a very dull place but also there would be no need for statistics. Variation is the raison d'être for statistics. Variation in people's feet means that shoe manufacturers have to make different sizes of shoes and hence have a great interest in knowing the distribution of sizes across ages in the population. Variation in the outcomes when two coins are tossed simultaneously and note taken of heads and tails, provides interest for punters and income for casinos that offer the game of "two-up" to their customers.

Variation occurs for different reasons in different mathematical contexts. In a measurement situation, for example measuring the height of a student, variation may occur due to measurement error. If every child in a class independently measured the height of a single person, not all of the values would be the same. The person's height would not change but there would be variation in the values obtained. This variation is usually called measurement error and the assumption is made that the true height of the person is somehow approximated by these values.

Continuing with the height example, the context might be extended to include the measurement of the heights of all students in a class. Variation in this case would be associated with the different growth rates of the different students. The variation of heights of all students within a class would be greater than the variation in the heights measured for an individual student.

One step further it would be possible to measure the heights of all students in a Grade 3 class and all students in a Grade 6 class. As well as talking about the variation of heights within each class, it would be possible to discuss the variation in heights between the two classes. One could imagine that the variation between classes would be greater than in the previous two contexts.

Imagining graphs for the three contexts is a useful way to characterise the issues of variation. Figure 1 shows some hypothetical data illustrating the variation present in the three cases. In this form students can discuss and describe differences and variation in data sets.

Usually when variation is encountered it gives rise to (or is associated with) an expectation. In the first height case, what is the actual or expected height of the person measured? In the second case, what is the typical or expected height of the class? In the third case, the question might be, is there a difference in the typical or expected heights of the two grades; or indeed is there a difference in the variation (spread) in heights of the two classes? In contexts such as these traditionally the arithmetic mean has been used to represent the expected value. As the mean is usually accessible by Grade 5 or 6, it is possible for students to base discussions on graphical representations including means and draw informal inferences about "the height of the single person measured," "the typical height in the class," or "the difference in typical height between two different classes."

Intuitively variation is "everywhere" in all data handling contexts: in sampling, in graphs, in different sample means, in chance events; and it even should be noted when conclusions are drawn. Expectation, on the other hand, can be more difficult to describe across contexts. Expectation may arise from a data-collecting situation, for example the expected height of a student discussed earlier. In a chance context, however, expectation may be the starting point; for example when tossing a coin many times there is likely to be the initial expectation that half of the outcomes will be heads and half tails. From the point of view of the accepted formal study of statistics, however, expectation is usually quantified as an arithmetic mean or proportion, whereas variation is usually quantified as the standard deviation. Because the mean is easier to calculate than the standard deviation, it has filtered further down in the school curriculum. Although it is not clear that firm ideas associated with expectation have accompanied the mean (Mokros & Russell, 1995), informal observation suggests that the importance of variation has been overlooked due to the lack of a "statistic" to calculate or measure it (see e.g., AEC, 1991; 1994).

Research on school students' intuitive understandings of aspects of the chance and data curriculum, however, suggests that variation is an idea absorbed much earlier than the idea of expectation (Watson, 2005). When asked in the context of a graph of how children come to school, whether the graph will look the same everyday, almost all students (even 6-year-olds) say that it will not and can give contextual reasons as to why, for example children being ill or the weather changing. Similarly when interviewed in a context of drawing 10 lollies from a container with 50 red, 30 green and 20 yellow lollies, young children readily suggest that there will not be the same number of red each time 10 lollies are drawn out. When initially asked how many reds they would expect to get, however, young children are more likely to choose their favourite numbers than five, reflecting the relative number of reds in the container. In most contexts, recognising the existence of variation is more straightforward than recognising the expectation that is at the heart of the variation. This is not to say that the variation described by young students would satisfy a statistician in terms of potential distributions of data; often students suggest ranges for variation that are unrealistically large. In relation to the two ideas and their relationship to each other and the process of a statistical investigation, it would appear that understanding develops in the following fashion, based initially on intuition and later influenced by the school curriculum:

- intuition about variation
- intuition about expectation
- measures of expectation (e.g., means and probabilities)
- measures of variation (e.g., standard deviation)

The importance of variation for everything that comes after (e.g., there is no point in talking about expectation if there is no variation) means that all activities from the beginning of work with chance and data need to emphasise variation in data sets.

Questions such as the following begin early.

- Although apples are the most popular fruit in our class, how many different fruits do we like? Would this change at different times of the year? Do you think it would be different if we lived in a different part of the country?
- When all of the class spun their half black-half white spinners 10 times, why did the results vary? Why would getting 10 blacks surprise you? [Or] How could you explain Jake's result when he got 10 blacks?

- In a graph of daily maximum temperatures throughout the year, plotted against time, what is interesting across the year? Are there any trends you can explain? Why is the graph not "smoother"? Would the graph look the same/similar every year? Why or why not?

The point of collecting data or conducting trials is not just to create a graph or to "prove" that a certain proportion of heads occur. All of the way through a statistical investigation issues of variation and expectation underlie the procedures followed and the decisions taken. In the final inferential step it is the balance of expectation and variation that influences the decision made. It also influences the confidence with which that decision is reported. The variation in outcomes from spinning spinners may be so great that an expectation of fair trials having taken place has to be rejected. Or the variation between the heights of Grade 3 and Grade 6 classes is so large compared to the variation within the classes that it is concluded that indeed Grade 6 students are taller.

Whatever techniques are used in statistical investigations, their purposes are to display, summarise, or compare variation and expectation. As Australia potentially moves to a nationally-based curriculum, the statistical component needs to acknowledge explicitly these two underlying features as a foundation not only for students who will go on to study statistics at the senior secondary and tertiary levels but also for all students who will take part in decision-making or questioning of decisions in social settings. To understand that a single person presented in a media report as "typical" of a social problem is only a sample-of-size-one, displaying no variation, is important in questioning or criticising claims made. To understand that odds express expectation only, not associated potential variation in outcomes, is also important for the punter.

To diminish the importance of chance and data in the curriculum or to imply that the thinking involved is not as sophisticated as that associated with algebraic reasoning, is to do irreparable damage to students who will be tomorrow's citizens of Australia.

Acknowledgement

The graph in Figure 1 was created with *Tinkerplots* (Konold & Miller, 2005).

References

Australian Education Council (1991). *A National Statement on Mathematics for Australian Schools*. Melbourne: Author.

Australian Education Council (1994). *Mathematics — A Curriculum Profile for Australian Schools*. Melbourne: Curriculum Corporation.

Board of Studies NSW. (1989). *Mathematics K–6*. Sydney: Author.

Board of Studies NSW. (2002). *Mathematics K–6 Syllabus*. Sydney: Author.

Curriculum Council (Western Australia). (1998). *Curriculum Framework Learning Statement for Mathematics*. Perth: Author.

Department of Education and the Arts. (1993). *Mathematics Guidelines K–8*. Hobart: Curriculum Services Branch.

Konold, C & Miller,C.D. (2005). *Tinkerplots Dynamic Data Exploration Software*. Emeryville, CA: Key Curriculum Press.

Mokros, J. & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education, 26*, 20–39.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

Rubin, A. (2005, Spring). Math that matters: The case for replacing the algebra/calculus track with data literacy – a critical skill for modern life. *Threshold: Exploring the Future of Education*, pp. 22–25, 31.

Victoria Curriculum and Assessment Authority. (2005). *Victorian Essential Learning Standards. Discipline-based learning strand: Mathematics*. Melbourne: Author.

Watson, J. M. (2005). Variation and expectation as foundations for the chance and data curriculum. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce & A. Roche (Eds), *Building connections: Theory, research and practice - Proceedings of the 28th annual conference of the Mathematics Education Research Group of Australasia, Melbourne*. pp. 35–42. Sydney: MERGA.

Jane Watson
University of Tasmania
<jane.watson@utas.edu.au>

APMC