Being mean ABOUT THE MEAN

by **Colin Foster**

On several occasions recently I have heard teachers say that they want their learners to understand which of 'the big three' averages (mean, mode and median) is the most suitable one to use in different circumstances. However, on none of these occasions have I seen a clear example that I have agreed with of when one average is obviously 'more appropriate' than the others.

Teachers and textbooks often point out that with categorical data (such as favourite colours) you may be able to find a mode, whereas you can't calculate a mean (because there are no numbers to add up) or a median (because the items can't be placed in order). On the other hand, they draw attention to the fact that there isn't always a mode for every set of data (e.g. 1, 2, 3, 4, 5), whereas every set of numerical data has a mean and a median. They also mention that when there is a mode it is necessarily a value that is found in the data set, whereas the mean and the median may be 'impossible' numbers, like '2.4 children'. So these observations could be framed as 'advantages' and 'disadvantages' of the mode, and I have no disagreement with them.

But I am thinking of situations in which it is possible to calculate all three averages but in which it is being argued that one of them is more or less 'appropriate' than the others. I think the arguments given are often rather dubious. For example, teachers often say that the mean is 'unduly' influenced by outliers, but is this right? Surely outliers influence the mean to a precisely proportionate degree? If the outlier is a mistake, and we don't want it to influence our conclusions, we should remove it from our data. If we leave it in, because it is real, then we should expect and *want* it to have its due influence on our results. An outlier may be the most important piece of data.

The most common example I see used in the mathematics classroom is of salaries in a small business, which might be:

£20k, £20k, £20k, £20k, £120k,

with a mean of £40k, which is greater than everyone's except for one person (perhaps the CEO). But surely the problem here is economic inequity, rather than that there is something wrong with the mean itself? Here the mean is 'misleading' in the sense that it doesn't describe what most of the people earn, but when you talk about 'most of the people' you are invoking the mode, and can we blame the mean for not being the mode? In the unlikely event that these five people were doing something together, like contributing to a charity, then their mean salary, although it might be close to nobody's individual salary, does represent the equal-shares income of them all, and might be useful. That is all the mean can ever do. Criticisms like this of the mean seem to be blaming it for not being something it never claims to be. If we only like the mean when it's close to the mode, then we should just use the mode.

Likewise, the mean gets blamed for not being a good 'measure of central tendency' for skewed data, because it is not near the middle of the distribution, but the mean was never intended to be the median! So one average seems to get criticized because it is not close to the other two. However, rather than complaining about this, it might be more helpful to note that the fact that it is not close to the other two is telling us something important about the data that the other two are missing, so in a sense making it particularly useful. We need to hear the dissenter!

If the three averages for a set of data come out almost the same, because the distribution is quite symmetrical, then it doesn't much matter which one we use. (I have seen lessons in which pupils are expected to debate which one is more appropriate when the differences between them are minute.) On the other hand, when the three averages are significantly different, it is unwise to use just one of them on some questionable grounds of 'appropriateness'. When the three averages are very different, the distribution needs to be examined, and pupils benefit from thinking about *why* the three numbers are so different and what each one is telling us. Choosing what they regard as 'the best', and discarding the other two, doesn't seem a sensible course to take. This highlights the importance of always looking at the distribution before (and perhaps, sometimes, instead of) calculating summary statistics.

It seems to me that the issue of 'appropriateness' is often confused with the deliberate misuse of statistics to misrepresent a situation to someone's advantage. Pupils are sometimes asked questions like 'Which average would you use if you were the boss of the company?' The assumption is that obviously they will be happy to mislead. I think it is extremely important for pupils to learn about how to lie with statistics (Huff, 1991), so that they can be in a strong position to critique bogus arguments presented to them, and are aware of traps to avoid themselves. However, I have sometimes been left in lessons with the impression that selecting an average that best suits your purposes is somehow normal and an acceptable use of statistics – particularly when language like 'most appropriate for the boss' is used. It is not even obvious to me what the answer is to the question of which average the boss should use if they want to mislead: they could use the mean to give the impression that they are a higher payer, but this could backfire if their employees get together and start asking, 'How come we're all earning less than the average?' Although it is lower, the median *conceals* high salaries, as though they are irrelevant to 'ordinary' people's concerns, and perhaps that is why we are so often told that median income is a more 'useful' measure – useful to whom? Nonetheless, perhaps the business of giving one average when the other two are significantly different is inherently dangerous and not something we should normalize in the classroom.

Reference

Huff, D. 1991 *How to Lie With Statistics*, Penguin Books, London.

Keywords: Averages; Mean; Measures of central tendency; Mode; Median.

Author Colin Foster, School of Education, University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB.
e-mail: **c@foster77.co.uk**
website: **www.foster77.co.uk**