MSV 34: A Small Sample

You are given a small sample
*consisting of **11** positive whole numbers.*

*You find it has an inter-quartile range of **3***
*and a variance of **5**.*

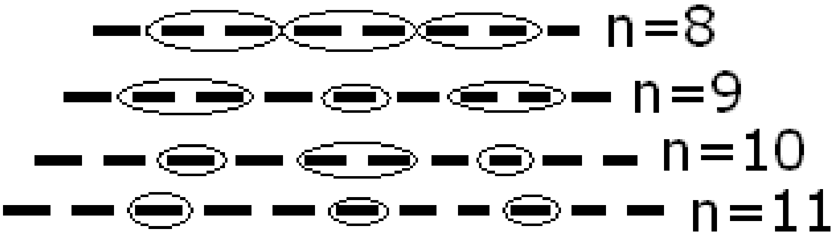
What might the sample be?

Many of the young mathematicians I teach are impressively ambitious, and a well-chosen starter rewards those who get there first and want to get stuck in. The key word, however, is well-chosen. In the hurly-burly of those few minutes pre-lesson, it is easy to improvise what looks like an innocuous bagatelle, only to discover later that it has surprisingly complex harmonies dwelling within.

Take the other day, when before a statistics lesson for AS students I wrote up the above. I hadn't worked the problem myself beforehand, but there was a little method in my possible madness.

'11 is a brilliant sample size – never any doubt as to what the quartiles are there. Now to engineer the IQR to be 3 is a piece of cake, which leaves a bit of tweaking on the top numbers to get the variance to be 5. Job done.'

I must confess that whenever I get involved with finding the quartiles for small sets of data, I start to worry. Textbooks disagree on how the quartiles should be defined, for one thing. It is possible to devise activities where the quartiles for the data set vary wildly according to what the next piece of data turns out to be. The topic is, however, on the syllabus, so it must be tackled. We are fortunate enough to do the MEI papers, and their definition for the quartiles for small data sets is admirably sensible. For MEI, the quartiles divide the data set into four parts that are as equal as possible and MEI give in effect four definitions, for the sample sizes 4n, 4n+1, 4n+2, and 4n+3.



A bubble around two values indicates ‘take the mean of these’. It can be seen that taking n = 11 keeps things nice and simple. Not simple enough, however; after ten minutes, my students were still labouring away, so I felt it was time to step in.

We agreed to start with 1, 1, 1, 1, 1, 1, 1, 1, 4, x, y as our sample, since 1 is the simplest positive integer. (On reflection, I could have use 0s to start with and then added 1 to everything afterwards, since a translation of the data does not affect its spread.) For this sample, the IQR is clearly 3, and the variance is

$$\frac{24 + x^2 + y^2 - 11\frac{(x+y+12)^2}{11^2}}{10} = 5.$$

I was starting to feel queasy. ‘A bit of tweaking on the top numbers’ was sounding rash to say the least. ‘Still, we can’t turn back now,‘ I thought.

This exercise was, in fact, proving to be helpful in lots of ways. Students were stumped as to how to multiply out $(x + y + z)^2$, a task for which GCSE algebra had not prepared them, but which comes up regularly in later maths. After multiplying out, we arrived at

$$x^2 + y^2 - 0.2xy - 2.4x - 2.4y - 43 = 0. \qquad *$$

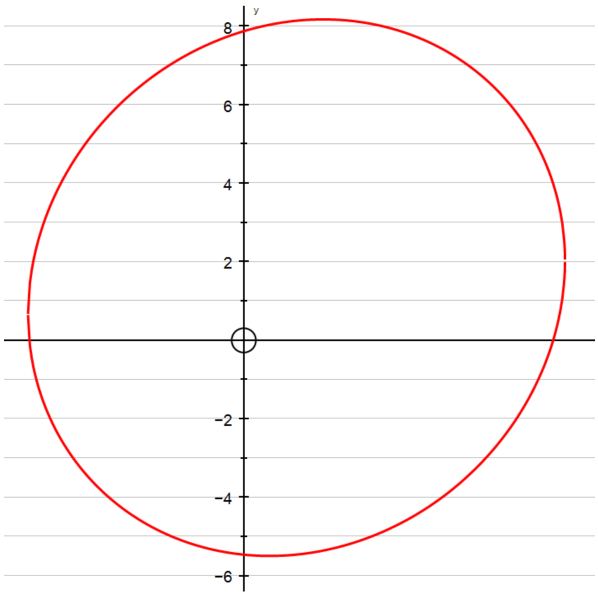
‘What kind of a curve is that?’ asked my curious students. ‘It’ll certainly be a conic,’ I said gravely. ‘What’s a conic?’ asked Ben; the others were thinking the same. I reached for the visual aid on my bookcase.

‘Cut a cone straight across, you get a circle. At a slight angle, you get…’

My wooden cone revealed an ellipse, followed by a parabola and half a hyperbola. Not on our syllabus, but surely something every A Level mathematician should see at least once in their studies. And how much better for this to arise unexpectedly from an apparently unrelated question. The level of motivation in the room was growing.

‘Now the wonderful truth is that if you take any polynomial equation for x and y of order two, that is, one involving just a mixture of x^2, y^2, xy , x, y and constants, you are guaranteed to get a conic section.’

‘So what will we get here?’ asked Emily. ‘An ellipse, I reckon,’ I said, and to my relief, Autograph backed me up.



‘Symmetrical about the line y = x; why should this be?’ I asked. This activity was leading in all sorts of directions that would be helpful to my students later in their course. ‘So we are looking for a point with integer coefficients that is on the curve,’ I said.

‘But the coordinates of the point have to be bigger than or equal to 4,’ said Amy, remembering our initial sample.

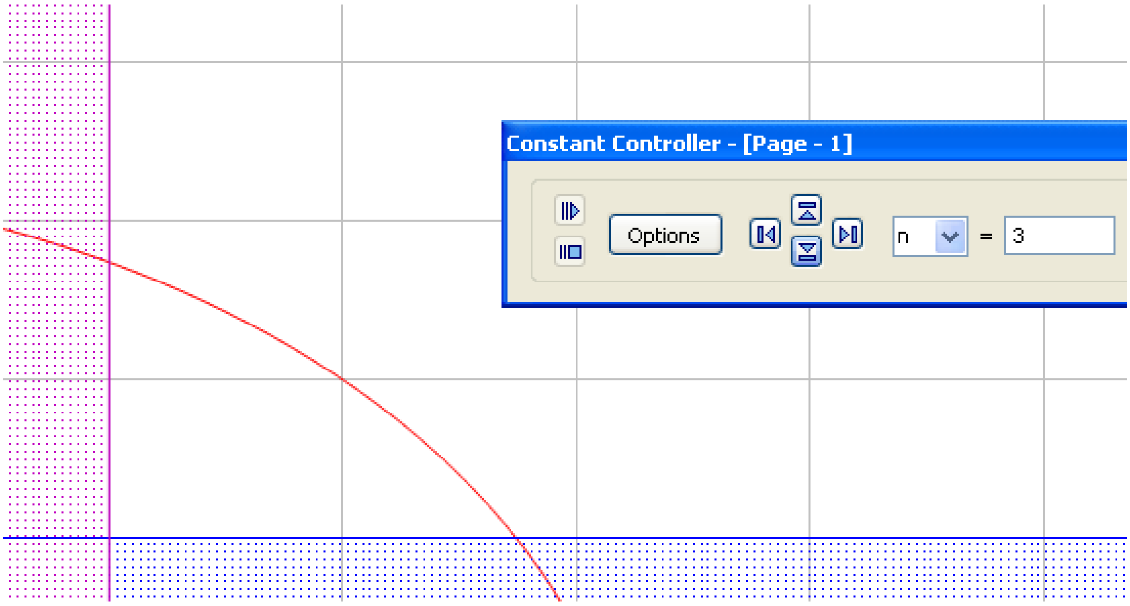
‘So we seek a lattice point on the curve in this region,’ I said, zooming in on the area in question.

‘So – there isn’t any such point,’ said Josh. The red line passed through no grid-points. I had set a question I couldn’t do. My students would usually be up in arms at this, throwing pens onto the table and sitting back in disgust, but this digression was thankfully going well enough to quell any rebellion. I gulped, and improvised again.

‘Let’s say instead,’ I offered carefully, ‘That the variance is n, where n is a whole number.’ Revising equation * we arrived at

$$\mathbf{x^2 + y^2 - 0.2xy - 2.4x - 2.4y - 11n + 12 = 0}$$

Now Autograph’s Constant Controller facility allowed us to step through values of n, looking all the while for somewhere where our red line passed through a grid-point. We didn’t have to search for too long. When n = 3, the curve passes through (5, 5).



So the sample 1, 1, 1, 1, 1, 1, 1, 1, 4, 5, 5 has an IQR of 3 and a variance of 3.

‘I’ll revise my question for next year,’ I said to my group – they looked pleased at having been part of this refinement for the future.

My guess is that real statisticians might be throwing up their hands in horror at this point. ‘What on earth have your students learnt about statistics from this?’ they might enquire. My answer would be, ‘Almost nothing.’ But what have they learnt of pure mathematics and of ways to investigate questions mathematically? A great deal. Who knows, maybe on some other day I would teach a pure lesson in which students would learn a lot along the way about statistics.

I will concede I was lucky here. The mathematics my starter threw up proved to be stuff I knew about, and the lesson had a happy ending. Another day my lack of preparation would have seen me come a cropper. But looking back, my students for this half hour could sense I was winging it, which lent a welcome unpredictability to the outcome. The important thing for me, and for my students too I hope, was that this lesson was alive.

Jonny Griffiths

www.making-statistics-vital.co.uk