Students determining the median for different data sets: a spectrum of responses

Jöran Petersson
Department of Mathematics and Science Education, Stockholm University

The present study reports on 359 students' responses to two test items on the median. Among the responses were other measures of location such as the arithmetic mean and the midrange. Moreover, among those who used a median strategy there was a spectrum of sources of confusion in the data set. In one test item the data were given in a table of signed integers and some students ignored the negative signs in the data. In the other test item the data were given in a bivariate diagram. Instead of correctly using the horizontal coordinate of the data, several students used the axis grading or the vertical coordinates as data. A conclusion is that the representation format of the data had a large effect on the achievement on the two test items.

Research question and method

The present study is a part of a larger study and aims to explore the following research question: Given different formats of data, what conceptual challenges do students meet as they determine the median as a measure of location? The research question was explored through a test.

Cooper and Shore (2008) found that some students confuse data with axis grading in univariate diagrams. In the present study data sets were a univariate table and a bivariate diagram.

A measure of location can be conceptualised as stable properties of samples and students may spontaneously tend to summarise (unimodal) data as modal clumps (Konold & Pollatsek, 2004). In the present study the median was chosen as measure of location since in school statistics it is arithmetically easy to calculate (Bakker & Gravemeijer, 2006). Despite this, the median is conceptually more difficult to grasp than the arithmetic mean (Bakker, 2003). Mayén, Díaz and Batanero (2009) found that students may confuse the median with other measures of location, such as arithmetic mean or midrange.

Results and discussion

The majority (56%) of the students gave a correct response to the test item of determining the median when the data was given in a table. Other responses were to determine the arithmetic mean (6%), the midrange (4%) as in (Mayén, Díaz & Batanero, 2009). Some students (3%) responded with the median of the absolute values, that is, they ignored the signs of negative numbers (3%). All students, who justified their midrange responses, used a number line for illustration. They marked the temperatures –9 and 7 on the number line and determined a point in the middle of this segment. In an interview where they could present their justification in more detail, one of them said: "I tried to calculate it in some way, like middle…". This justification corresponds to a confusion of an *ordinal middle* of all values (the median) with a *geometric middle* of the extremes (the midrange).

The other test item was to determine the median when the data was given in a bivariate diagram. Here a minority (28%) of the students gave a correct response despite having achieved well (81%) on a task of reading a single data point from the same diagram. Some students (6%) had reasoned correctly but omitted/ignored one or two points and thus got a wrong answer. Some students (4%) confused the median with the arithmetic mean or midrange.

For many students, the source of confusion was in the choice of data set; a confusion similar to that in (Cooper and Shore, 2008). Some (12%) responded with the median of the vertical component of the coordinates instead of the horizontal component. A large proportion (21%) determined the median of the axis grading. There were also interval answers (6%) of which some corresponded to a modal clump (Konold & Pollatsek, 2004). For example interviewed students justified their interval response with "They were really very close there" and "most were there ", referring to the points in the diagram.

A conclusion is that the students may confuse one measure of location with another, but the format of the data set is also a significant source of confusion.

References

Bakker, A. (2003). The early history of average values and implications for education. *Journal of Statistics Education, 11*(1)

Bakker, A., & Gravemeijer, K. P. E. (2006). An historical phenomenology of mean and median. *Educational Studies in Mathematics, 62*(2), 149-168. doi:10.1007/s10649-006-7099-8

Cooper, L. L. & Shore, F. S. (2008). Students' Misconceptions in Interpreting Center and Variability of Data Represented via Histograms and Stem-and-leaf Plots. *Journal of Statistics Education, 16*(2).

Konold, C., & Pollatsek, A. (2004). Conceptualizing an average as a stable feature of a noisy process. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 169-199). Boston: Kluwer Academic Publishers.

Mayén, S., Díaz, C., & Batanero, C. (2009). Students' semiotic conflicts in the concept of median. *Statistics Education Research Journal, 8*(2), 5-32.