Correspondence

Dear Editor,

Kudos to Paul Belcher for his "Matter for Debate" piece [1] on outliers in data with more than one dimension. It is indeed valuable to consider the tradeoffs of various definitions as well as what can be learned by looking at the data visually. Two additional things worth noting in the case of two-dimensional data:

1. Interpretation of outliers in a data set requires context, so readers may appreciate a concrete vignette such as [2, p. 181] "a study of the weight of children aged 7 to 17; a child aged 16 would not be an outlier, nor would a child who weighed 70 pounds ... but a 16-year-old who weighed 70 pounds would be."
2. Not every outlier (i.e., data point whose response y does not follow the data set's general trend) and not every data point with high leverage (i.e., with an extreme value of the predictor x) ends up substantially affecting the estimated regression slope coefficient, correlation coefficient, or hypothesis test result for fitting a line to the data. This can be explored by using statistical software to run linear regressions with and without the data point in question (see examples at https://online.stat.psu.edu/stat462/node/170/). This can also be explored informally via an applet such as https://www.nctm.org/Classroom-Resources/Illuminations/Interactives/Line-of-Best-Fit/.

References

1. Paul Belcher, Definitions for outliers in data in two-dimensional and higher-dimensional data, *Math. Gaz.* **108** (November 2024) pp. 507-511.
2. W. Paul Vogt, *Quantitative Research Methods for Professionals*. Pearson (2007).

10.1017/mag.2025.35
Published by Cambridge University Press
on behalf of The Mathematical Association

LARRY LESSER
UTEP Mathematical Sciences Department,
500 W University Ave.,
The University of Texas at El Paso,
TX 79968, USA
e-mail: *lesser@utep.edu*