

TL12: Outliers

Real-life data is rarely perfect and the data being considered in this Case is no different. As pupils attempt to sketch lines of best fit they will find points which sit considerably off any predicted line. These are called outliers (we are using this term a little loosely here as technically an outlier is defined as one and half times the interquartile range above the upper quartile or below the lower quartile). A range of factors could contribute to these performances including the introduction of new technology, drugs, environment of competition etc. Upon identifying the outliers in their data, pupils may take educated guesses at potential reasons for these performances or may choose to go away and research this performance following lesson three. Several reasons were listed in **TL7** as to why the 1896 Olympic results are outliers in almost all events and reasons in some other events were given in TL5. However you may let pupils try to find some of these.

800m model

When examining the scatter plot for the men's 800m, the first performance in 1896 stands out as an outlier for the reasons about the track given in **TL7**. Interestingly, the first women's 800m (1928), although much slower than the next one in 1960, is seen on the scatter plot to not be a significant outlier; the 32 year gap means it is close to any line of best fit. This is not easy to see without using a graph.

Key principles for reflection

Data rarely fits a perfect pattern and will usually contain outliers. A range of factors may cause an outlier and these should be considered when considering the point in the context of a trend.

