

Demo: Exploring the Impact of Outliers on Correlation

To explore the impact of unusual observations and outliers on the correlation coefficient, we use the demoCorr script. This script is in the JMP Sample Scripts Directory.

The demo correlation script starts with several observations and allows us to drag points or add new points and explore changes in the correlation coefficient.

The initial correlation is 0.675. X and Y are positively correlated. So, in general, as X increases, Y also increases.

When I add a new observation at a low value of Y and a high value of X, the correlation drops.

When I drag this observation to a high value of Y and a low value of X, again the correlation drops.

In both of these cases, without looking at the graph, we might conclude that there is no relationship between the two variables.

What happens if this point is an outlier for both X and Y? I'll change the axes for X and Y and add this observation at extreme values for both variables.

Do you notice what happens to the correlation now? The correlation is now strongly positive. This one outlier has caused the correlation to be inflated.

Look how important it is to plot your data before you interpret the output!