Traditional Variance

The traditional formula for variance — the square root of which is the standard deviation — looks like this:

$$v = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Problems Arise

But this formula has a problem with large data sets in that it requires us to work out the average ($\bar{x}$) first and then run back through the data to find the variance.

Problems Solved

So, we often use a reworked version of the formula as so:

$$\begin{aligned}
(n-1)v &= \sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \sum_{i=1}^{n}(x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \bar{x}^2 \sum_{i=1}^{n} 1 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}(n\bar{x}) + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 + (n-2n)\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2
\end{aligned}$$

Here, as we can see, we don't need the average until after we've summed the squares of the data items. Thus we can add the squares at the same time we are adding the values themselves and then at the end calculate the average and then variance.