Symbolic Representation, Calculating the Mean, Variance, and Standard Deviation and the Effect of Outliers

In statistics we use letter as symbolic representations for variables.
For example, X may be a variable label for a set of scores on a Chemistry Exam.

X	$X - \bar{X}$	$= (X - \bar{X})$	$(X - \bar{X})^2$
92	92 - 88	4	16
90	90 - 88	2	4
90	90 - 88	2	4
88	88 - 88	0	0
88	88 - 88	0	0
88	88 - 88	0	0
86	86 - 88	-2	4
86	86 - 88	-2	4
84	84 - 88	-4	16

$\Sigma X = 792$ $\Sigma (X - \bar{X})^2 = 48$

$\bar{X} = \frac{\sum X}{N} = 792/9 = 88$ $s_X^2 = \frac{\sum(X-\bar{X})^2}{(N-1)} = 48/8 = 6$ $s_X = \sqrt{s_X^2} = \sqrt{6} = 2.45$

To calculate the mean of X we sum every score (X) and divide by the Total (N)

To calculate Variance (s_X^2) and Standard Deviation (s_X) of X, first we subtract the mean ($\bar{X}$) from every score.

We then square these deviations then sum them and divided by (N-1) to obtain . To obtain the Standard Deviation (s_X), we take the square root of the variance.

Unfortunately, one extreme score (outlier) can bias the Mean so that it does not represent most of the scores. This asymmetry (or skewness) also inflates the Variance and Standard Deviation as measures of Dispersion. For example, if an extremely low score of 28 were added as a tenth observation of X, then the resultant mean of $\bar{X}$ = 82, is not close to many of the observations. The standard deviation is also inflated to s_X = 19.11 even though 90% of the scores are relatively close together (i.e., excluding the 28, the scores range from 84 to 92).

X	$X - \bar{X}$	$= (X - \bar{X})$	$(X - \bar{X})^2$
92	92 - 82	10	100
90	90 - 82	8	64
90	90 - 82	8	64
88	88 - 82	6	36
88	88 - 82	6	36
88	88 - 82	6	36
86	86 - 82	4	16
86	86 - 82	4	16
84	84 - 82	2	4
28	28 - 82	-54	2916

$\Sigma X = 820$ $\Sigma (X - \bar{X})^2 = 3288$

$\bar{X} = \frac{\sum X}{N} = 820/10 = 82$ $s_X^2 = \frac{\sum(X-\bar{X})^2}{(N-1)} = 3288/9 = 365.33$

$s_X = \sqrt{s_X^2} = \sqrt{365.33} = 19.11$