Variance and Standard deviation

Variance and Standard Deviation are the two important measurements in statistics.

Variance According to layman's words, the variance is a measure of how far a set of data are dispersed out from their mean or average value. It is denoted as 'σ^2'.

Standard deviation is the measure of the distribution of statistical data.

The basic difference between both is standard deviation is represented in the same units as the mean of data, while the variance is represented in squared units.

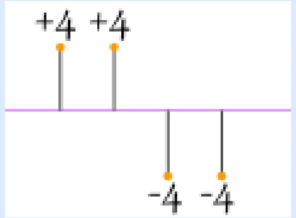
Properties of Variance

•It is always non-negative since each term in the variance sum is squared and therefore the result is either positive or zero.

•Variance always has squared units. For example, the variance of a set of weights estimated in kilograms will be given in kg squared. Since the population variance is squared, we cannot compare it directly with the mean or the data themselves.

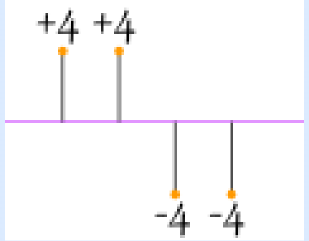
Why *square* the differences?

If we just add up the differences from the mean ... the negatives cancel the positives:



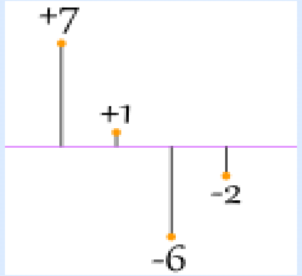
$$\frac{4 + 4 - 4 - 4}{4} = \boxed{0}$$

So that won't work. How about we use absolute values?



$$\frac{|4| + |4| + |-4| + |-4|}{4} = \frac{4 + 4 + 4 + 4}{4} = \boxed{4}$$

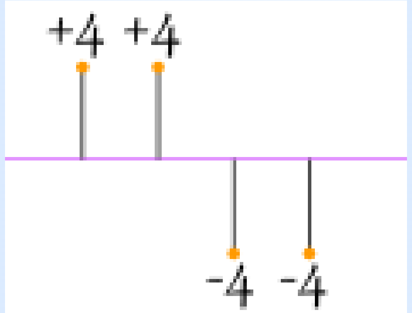
That looks good (and is the Mean Deviation), but what about this case:



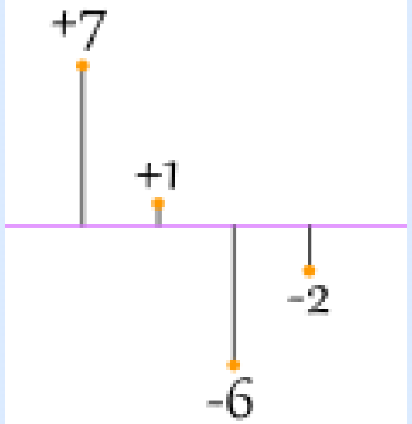
$$\frac{|7| + |1| + |-6| + |-2|}{4} = \frac{7 + 1 + 6 + 2}{4} = \boxed{4}$$

Oh No!
It also gives a value of 4, Even though the differences are more spread out.

So let us try squaring each difference (and taking the square root at the end):



$$\sqrt{\left(\frac{4^2 + 4^2 + (-4)^2 + (-4)^2}{4}\right)} = \sqrt{\left(\frac{64}{4}\right)} = 4$$



$$\sqrt{\left(\frac{7^2 + 1^2 + (-6)^2 + (-2)^2}{4}\right)} = \sqrt{\left(\frac{90}{4}\right)} = 4.74...$$

That is nice! The Standard Deviation is bigger when the differences are more spread out ... just what we want.

Standard Deviation

The spread of statistical data is measured by the standard deviation. Distribution measures the deviation of data from its mean or average position. The degree of dispersion is computed by the method of estimating the deviation of data points. It is denoted by the symbol, ‘σ’.

Properties of Standard Deviation

It describes the square root of the mean of the squares of all values in a data set and is also called the root-mean-square deviation.

The smallest value of the standard deviation is 0 since it cannot be negative.

When the data values of a group are similar, then the standard deviation will be very low or close to zero. But when the data values vary with each other, then the standard variation is high or far from zero.

Variance and Standard Deviation Formula

As discussed, the variance of the data set is the average square distance between the mean value and each data value. And standard deviation defines the spread of data values around the mean.

	Population	Sample
Variance	$\sigma^2 = \dfrac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$	$S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}$
Standard deviation	$\sigma = \sqrt{\dfrac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$	$S = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}}$

The sample variance formula is given as:

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

Here,

s^2 = Sample variance

n = Number of observations in sample

x_i = ith observation in the sample

$\overline{x}$

Variance Formula:

The population variance formula is given by:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$$

Here,

σ^2 = Population variance

N = Number of observations in population

X_i = ith observation in the population

μ = Population mean

Question: If a die is rolled, then find the variance and standard deviation of the possibilities.

Solution: When a die is rolled, the possible outcome will be 6. So the sample space, n = 6 and the data set = { 1;2;3;4;5;6}.

To find the variance, first, we need to calculate the mean of the data set.

Mean, $\overline{x} = (1+2+3+4+5+6)/6 = 3.5$

We can put the value of data and mean in the formula to get;

$\sigma^2 = \Sigma (x_i - \overline{x})^2/n$

$\sigma^2 = ⅙ (6.25+2.25+0.25+0.25+2.25+6.25)$

$\sigma^2 = 2.917$

Now, the standard deviation,$\sigma = \sqrt{2.917} = 1.708$