Mean, Standard Deviation and Variance

Deviation just means how far from the normal

Standard Deviation: The Standard Deviation is a measure of how spread out numbers are.

Its symbol is **σ** (the greek letter sigma)

The formula is easy: it is the **square root** of the **Variance.** So now you ask, "What is the Variance?"

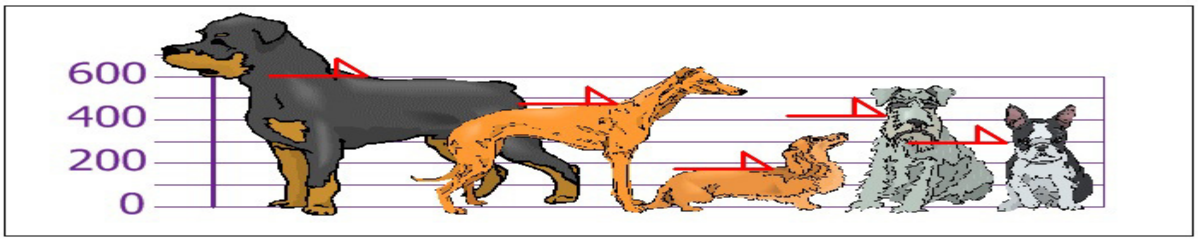
Variance: The Variance is defined as:

The average of the **squared** differences from the Mean.

To calculate the variance follow these steps:

- Work out the <u>Mean</u> (the simple average of the numbers)
- Then for each number: subtract the Mean and square the result (the *squared difference*).
- Then work out the average of those squared differences

Example: You and your friends have just measured the heights of your dogs (in millimeters):



The heights (at the shoulders) are: 600mm, 470mm, 170mm, 430mm and 300mm.

Find out the Mean, the Variance, and the Standard Deviation.

Your first step is to find the Mean:

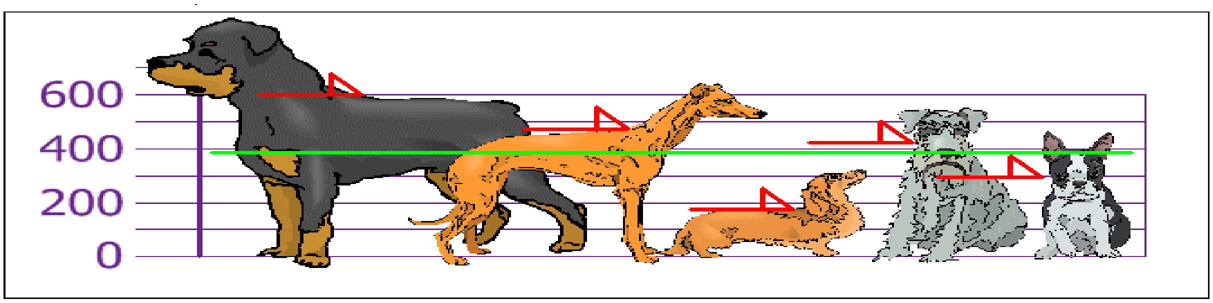
Answer:

Mean = 600 + 470 + 170 + 430 + 3005

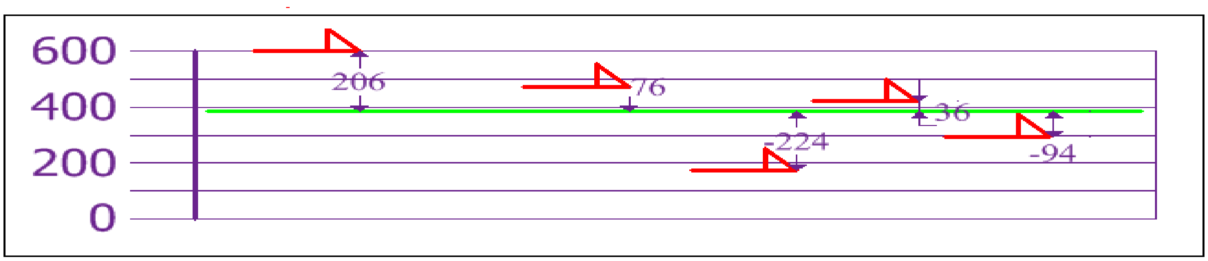
= 19705

= 394

so the mean (average) height is 394 mm. Let's plot this on the chart::

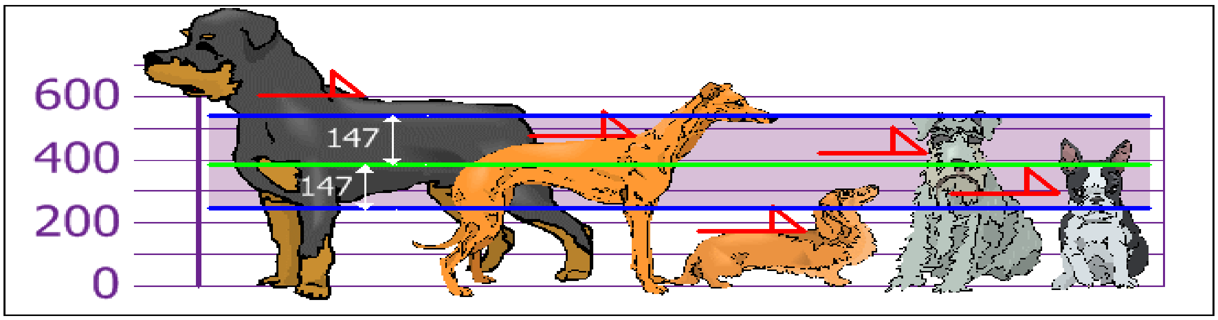


Now we calculate each dog's difference from the Mean



To calculate the Variance, take each difference, square it, and then average the result:
Variance
σ2 = 2062 + 762 + (−224)2 + 362 + (−94)25
= 42436 + 5776 + 50176 + 1296 + 88365
= 1085205
= 21704
So the Variance is 21,704
And the Standard Deviation is just the square root of Variance, so:
Standard Deviation
σ = √21704
= 147.32...
= 147 (to the nearest mm)

And the good thing about the Standard Deviation is that it is useful. Now we can show which heights are within one Standard Deviation (147mm) of the Mean:



So, using the Standard Deviation we have a "standard" way of knowing what is normal, and what is extra large or extra small.

Rottweilers **are** tall dogs. And Dachshunds **are** a bit short ...!

Our example has been for a **Population** (the 5 dogs are the only dogs we are interested in).

But if the data is a **Sample** (a selection taken from a bigger Population), then the calculation changes!

When you have "N" data values that are:

- **The Population**: divide by **N** when calculating Variance (like we did)
- **A Sample**: divide by **N-1** when calculating Variance

All other calculations stay the same, including how we calculated the mean.

Example: if our 5 dogs are just a **sample** of a bigger population of dogs, we divide by **4 instead of 5** like this:

Sample Variance = 108,520 / **4** = **27,130**

Sample Standard Deviation = √27,130 = **164** (to the nearest mm)

Think of it as a "correction" when your data is only a sample.

The "**Population** Standard Deviation":

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