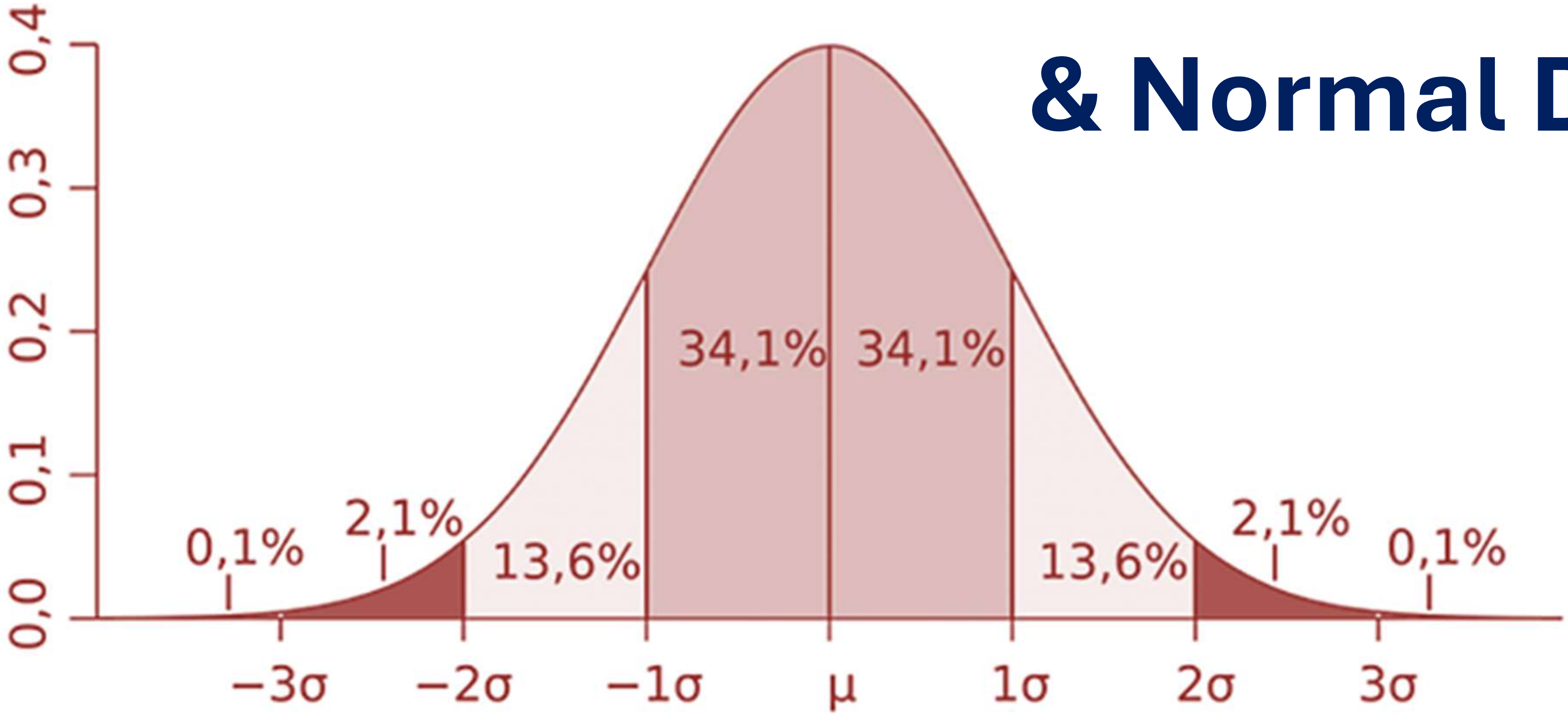
STATISTICS

@DiligentEdu

2024

Standard Deviation

& Normal Distribution



DiligentEdu©

Standard Deviation

- **Measures of Dispersion**

An **average** (*mean*, *median* and *mode* are the most popular averages) can *represent a series* only as best as ***a single figure*** can, but it certainly ***cannot reveal the entire story*** of any phenomenon under study. Specially it ***fails*** to give any ***idea about the scatter of the values*** of items of a variable in the series around the true value of average.

DiligentEdu©

Standard Deviation

▪ Measures of Dispersion

In order to measure this ***scatter***, statistical devices called **measures of dispersion** are calculated.

Important measures of dispersion are:

(a) range

(b) mean deviation

(c) standard deviation (**SD** or symbol '**σ**' sigma)

DiligentEdu©

Standard Deviation

▪ Standard Deviation

Standard Deviation tells about **how much** data values are ***deviated*** (spread, dispersion, spread) from the **mean** value.

- The Standard Deviation is a measure of **how spread out numbers are**.
- **Deviation** just means **how far from the normal**.
- A standard deviation (or **σ**) is a measure of how **dispersed the data** is in relation **to the mean**.

DiligentEdu©

Standard Deviation

▪ Standard Deviation

If the **data** points are ***close*** **to the mean**, there is a **small variation** whereas the data points are ***highly spread out*** **from the mean**, then it has a **high variance**.

Standard deviation calculates the ***extent*** to which the values differ from the average.

DiligentEdu©

Standard Deviation

- ## Standard Deviation

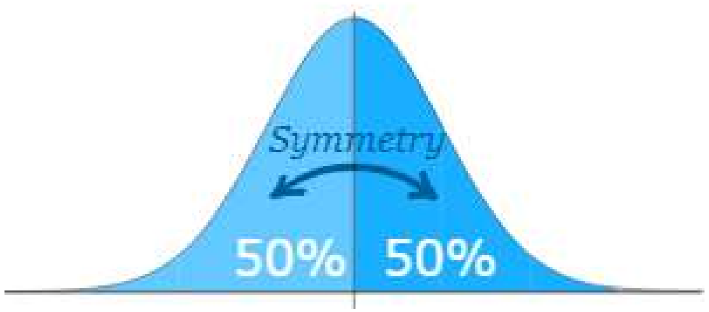
Standard Deviation, the *most widely* used measure of dispersion, is ***based on all values***.

Therefore a change in even one value affects the value of standard deviation.

DiligentEdu©

Standard Deviation

Normal Distribution



DiligentEdu©

Standard Deviation

- **Normal distribution**

Normal Distribution: a symmetrical, **bell-shaped** distribution in which **most** scores **cluster near the mean** and in which scores become *increasingly* ***rare*** as they become increasingly **divergent from this mean**.

Many things that can be quantified are Normally Distributed.

DiligentEdu©

Standard Deviation

- **Normal distribution - *Examples***

Distributions of
- height
- weight
- the age at which infants begin to walk
- *etc*.

are all examples of *approximately* **normal distributions**.

DiligentEdu©

Standard Deviation

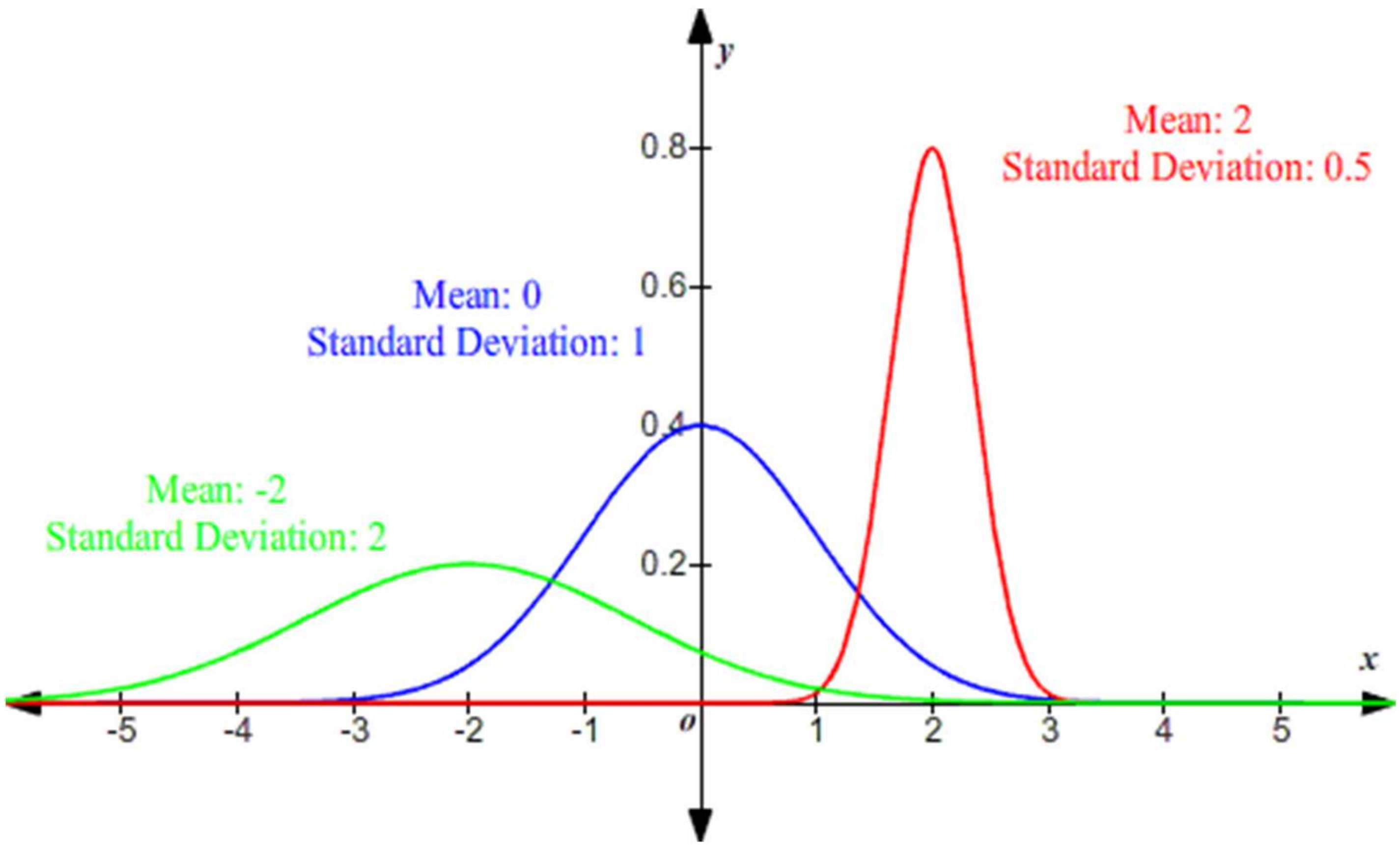
▪ Shape of Normal Distribution

The shape of a **normal distribution** is determined by the **mean** and the **standard deviation**.

The **steeper** the **bell curve**, the **smaller the standard deviation**. If the examples are **spread far apart**, the **bell curve** will be **much** **flatter**, meaning the **standard deviation** is **large**.

DiligentEdu©

Standard Deviation



Source: https://www.varsitytutors.com

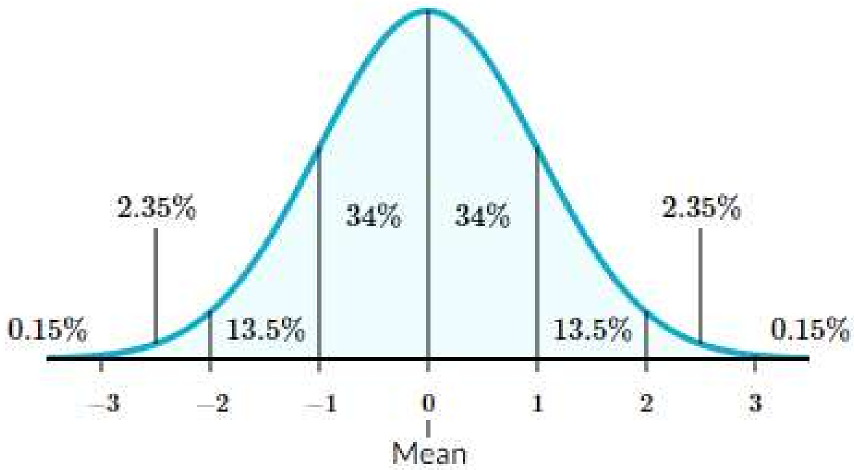
DiligentEdu©

Standard Deviation

▪ Normal distribution

Why it is called a Normal Distribution ?

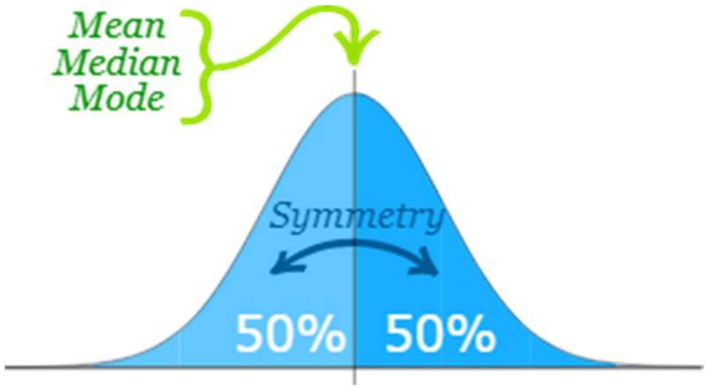
Early statisticians noticed the ***same shape coming up over and over again*** in different distributions - so they **named it** the **normal distribution**.



DiligentEdu©

Standard Deviation

- **Normal distribution**



The **Normal Distribution** has:

- mean = median = mode
- symmetry about the centre
- **50%** of values **less than the mean** and **50% greater than the mean**

DiligentEdu©

Standard Deviation

▪ Normal distributions - *Examples*

1. Height

Height is a suitable example of **normal distribution**. For instance, an individual uses this tool to evaluate the height of a random population of 2,000 other individuals. **Many** of these people have **average height**. There is a **smaller group** of people who are **taller** or **shorter** than the people who have **average height**. Here, there are more values that are near the mean and some values that are occurring at the tail ends. There are many factors, like genetic and environmental, that may affect this outcome, so **height** remains an independent variable.

DiligentEdu©

Standard Deviation

▪ Normal distributions - *Examples*

2. Intelligence quotient (IQ)

Intelligence quotient or **IQ** is another appropriate example of normal distribution. For instance, an individual intends to assess the IQ levels of a specific population. This population may be **students in a classroom** or **teachers in a school**. After this assessment, the individual finds that **most people have an IQ that is within the normal range**. There are only a **few persons** who are within the **deviate region**, possessing an IQ level that is **above** the average or **below**.

DiligentEdu©

Standard Deviation

▪ Normal distributions - *Examples*

3. A newborn's weight

A healthy newborn has a weight that typically ranges from two to four kilograms. This average may develop a normal curve. The weight of most newborns is within the range and there are only some newborns whose weights may deviate or even occur below or above the average. Examining a population of newborns may help an individual determine the average weight of the sample population.

DiligentEdu©

Standard Deviation

▪ Normal distributions - *Examples*

4. Distribution of income

Income ranges largely depend on the economy of a country or its government's financial policies. The middle class makes up a larger population, compared with the rich and poor. The income of the middle class forms the mean on a normal distribution curve. An economy's income distribution may produce a bell-shaped graph, as this class has a dense population that is occurring towards the expected value.

DiligentEdu©

Standard Deviation

▪ Normal distributions - *Examples*

5. Clothing size

Clothing manufacturers may conduct a survey to determine which sizes of clothing are most in demand. The size of shoes and clothes has a normal curve. This happens because there are more people who are within the average range of size and fewer people who are below or above this average. This factor may have an influence on the number of clothes that the manufacturer is making for a particular size range. It may also get challenging for people to find clothes in their sizes when these people are outside the range.

DiligentEdu©

Standard Deviation

▪ Normal distributions - *features*

A normal distribution has some interesting properties: it has a **bell shape**, the ***mean*** and ***median*** are equal, and 68% of the data falls within 1 standard deviation.

- symmetric bell shape
- mean and median are equal; both located at the center of the distribution
- ≈ 68% of the data falls within 1 standard deviation of the mean
- ≈ 95%of the data falls within 2 standard deviations of the mean
- ≈ 99.7% of the data falls within 3 standard deviations of the mean

DiligentEdu©

Standard Deviation

▪ Normal distributions - *Examples*

Example 1:

A set of data is normally distributed with a mean of 5 . What percent of the data is less than 5 ?

A normal distribution is symmetric about the mean. So, half of the data will be less than the mean and half of the data will be greater than the mean.

Therefore, 50% percent of the data is less than 5 .

DiligentEdu©

Standard Deviation
(SD / σ)

... *continued*

DiligentEdu©

Standard Deviation

- **Standard Deviation (SD / σ)**

Formula: it is the **square root** of the **Variance**.

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

DiligentEdu©

Standard Deviation

- **Steps to calculate the Variance:**

- Work out the Mean (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.

DiligentEdu©

Standard Deviation

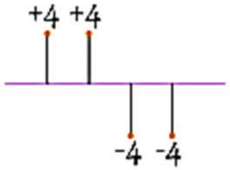
- **Steps to calculate the Variance:**

- Work out the Mean (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. (Why Square?)

DiligentEdu©

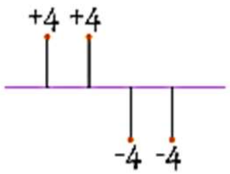
Why *square* the differences?

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



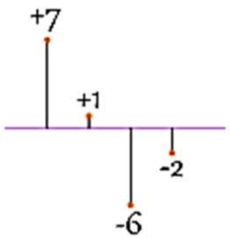
$$\frac{4+4-4-4}{4} = 0$$

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



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

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



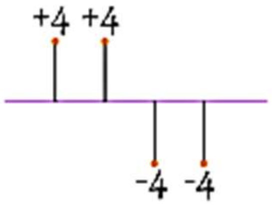
$$\frac{|7|+|1|+|-6|+|-2|}{4} = \frac{7+1+6+2}{4} = 4$$

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

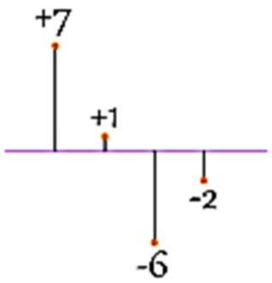
DiligentEdu©

Standard Deviation

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.

In fact this method is a similar idea to distance between points, just applied in a different way.

And it is easier to use algebra on squares and square roots than absolute values, which makes the standard deviation easy to use in other areas of mathematics.

DiligentEdu©

Standard Deviation

- **Standard Deviation (SD / σ) - *Example***

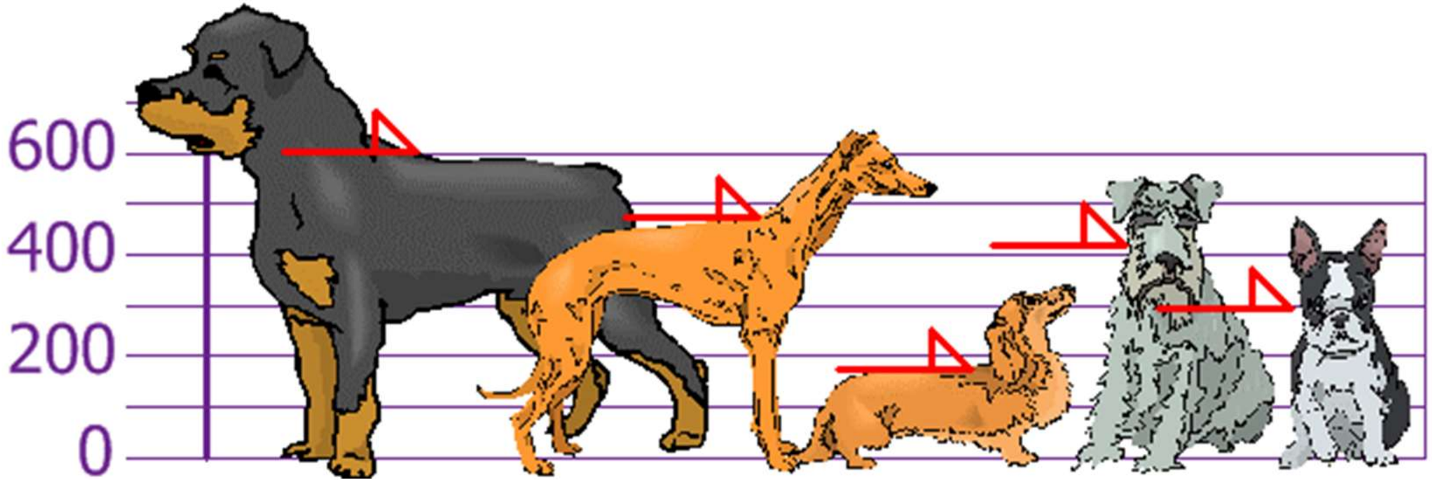
Question/Case - You and your friends have just measured the heights of your dogs (in millimetres):

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

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

DiligentEdu©

Standard Deviation



DiligentEdu©

Standard Deviation

Your first step is to find the Mean:

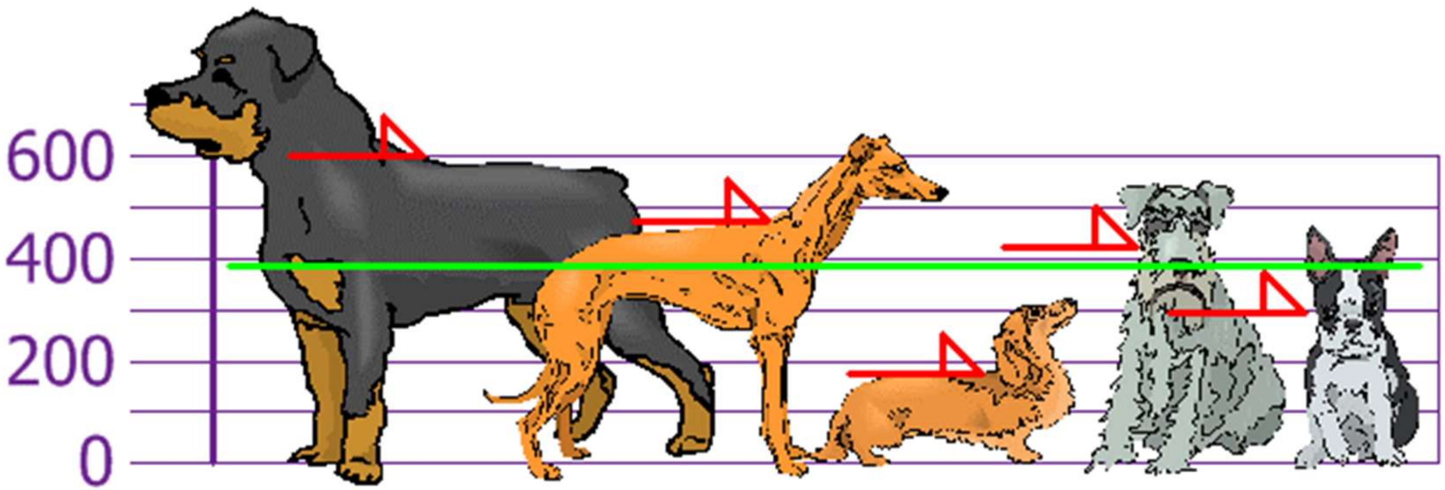
Answer:

$$\text{Mean} = \frac{600 + 470 + 170 + 430 + 300}{5}$$

$$= \frac{1970}{5}$$

$$= 394$$

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



DiligentEdu©

Standard Deviation

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

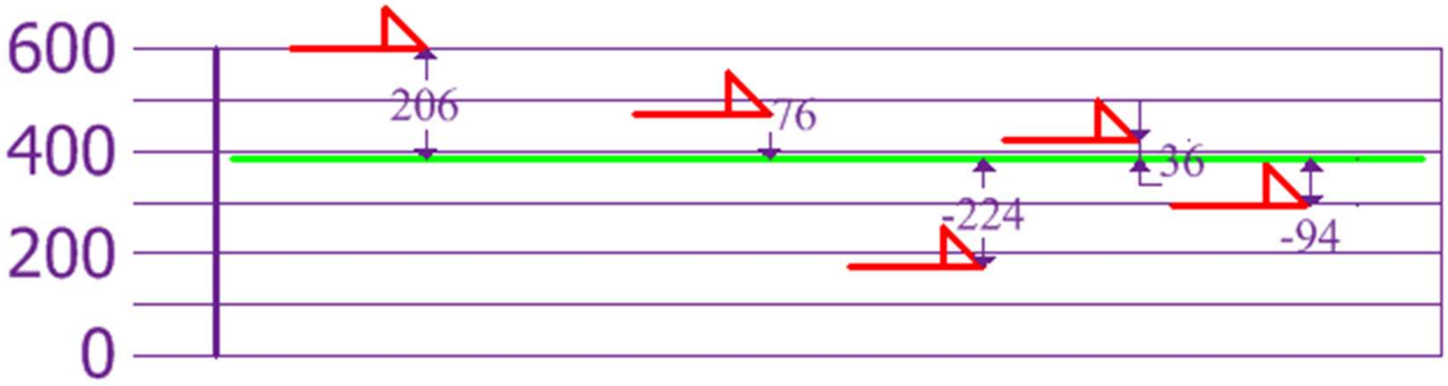
$$\sigma^2 = \frac{206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2}{5}$$

$$= \frac{42436 + 5776 + 50176 + 1296 + 8836}{5}$$

$$= \frac{108520}{5}$$

$$= 21704$$

So the Variance is **21,704**



DiligentEdu©

Standard Deviation

And the Standard Deviation is just the square root of Variance, so:

Standard Deviation

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

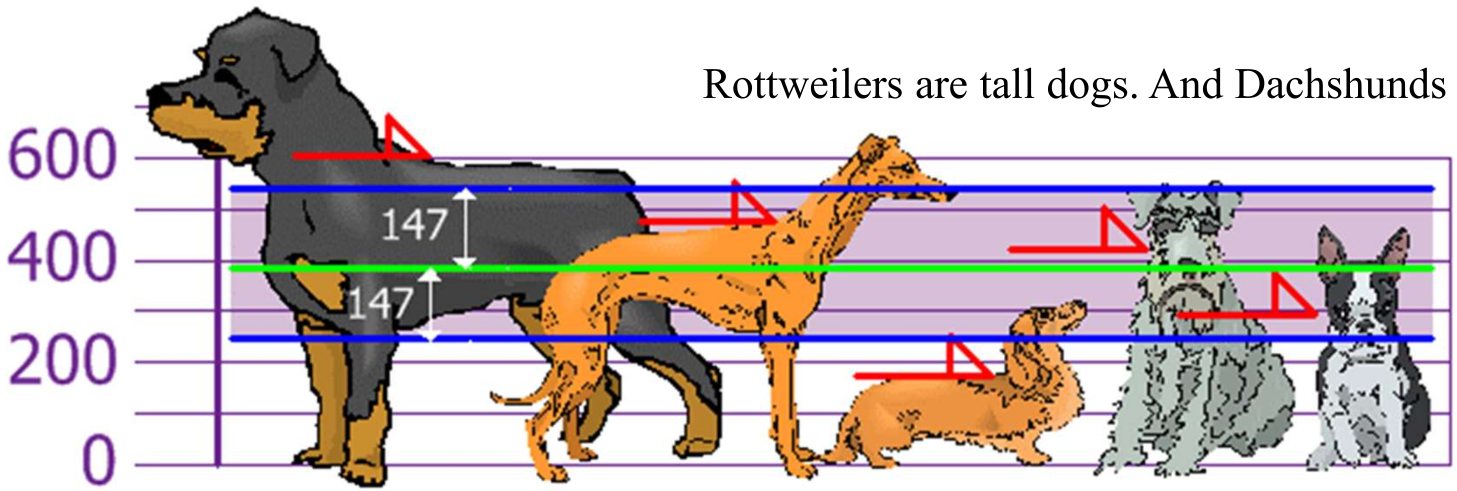
$\sigma = \sqrt{21704}$

$= 147.32...$

$=$ **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:

Rottweilers are tall dogs. And Dachshunds are a bit short, right?



DiligentEdu©

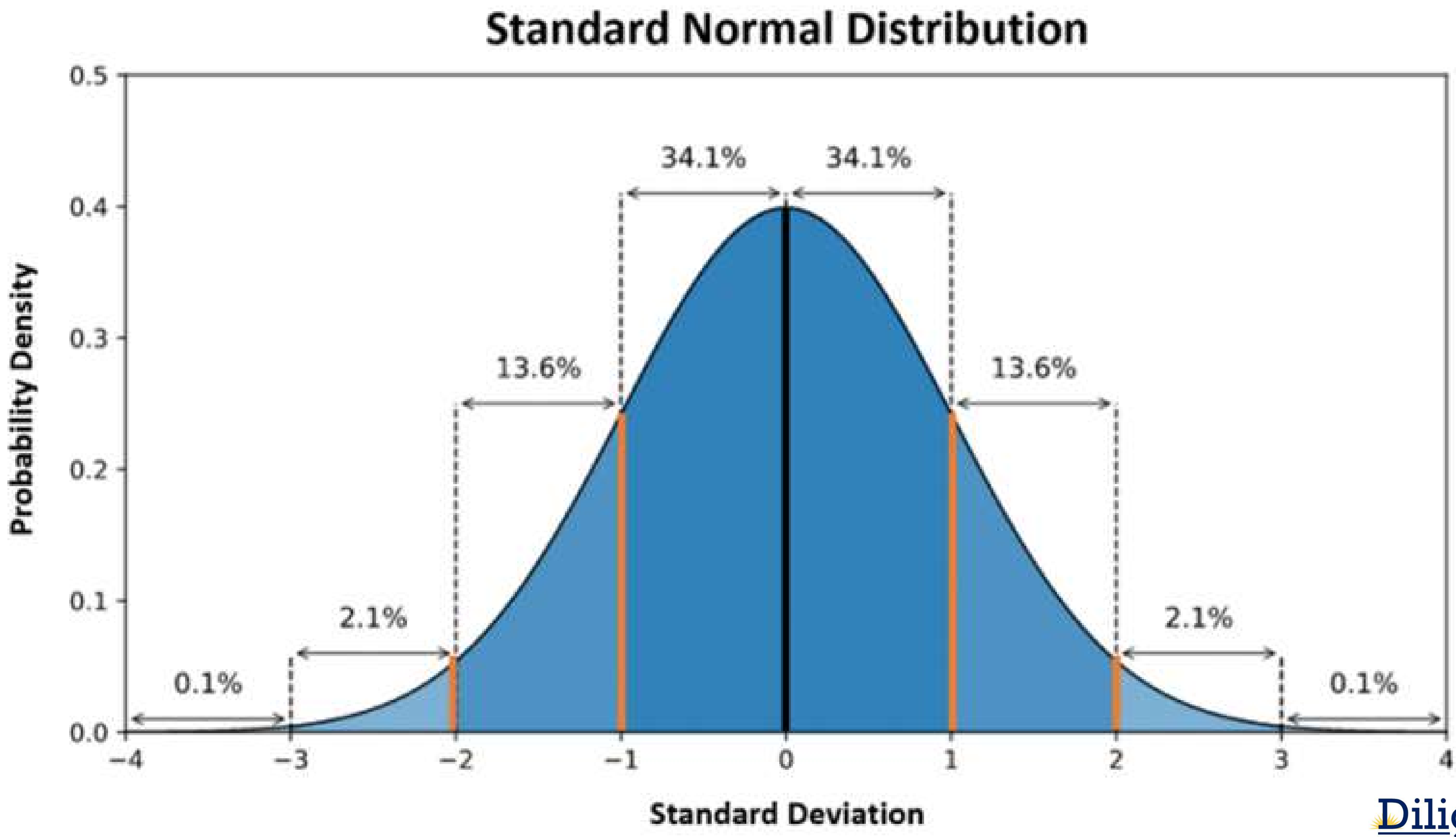
Standard Deviation

- **Standard Deviation (SD / σ)**

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

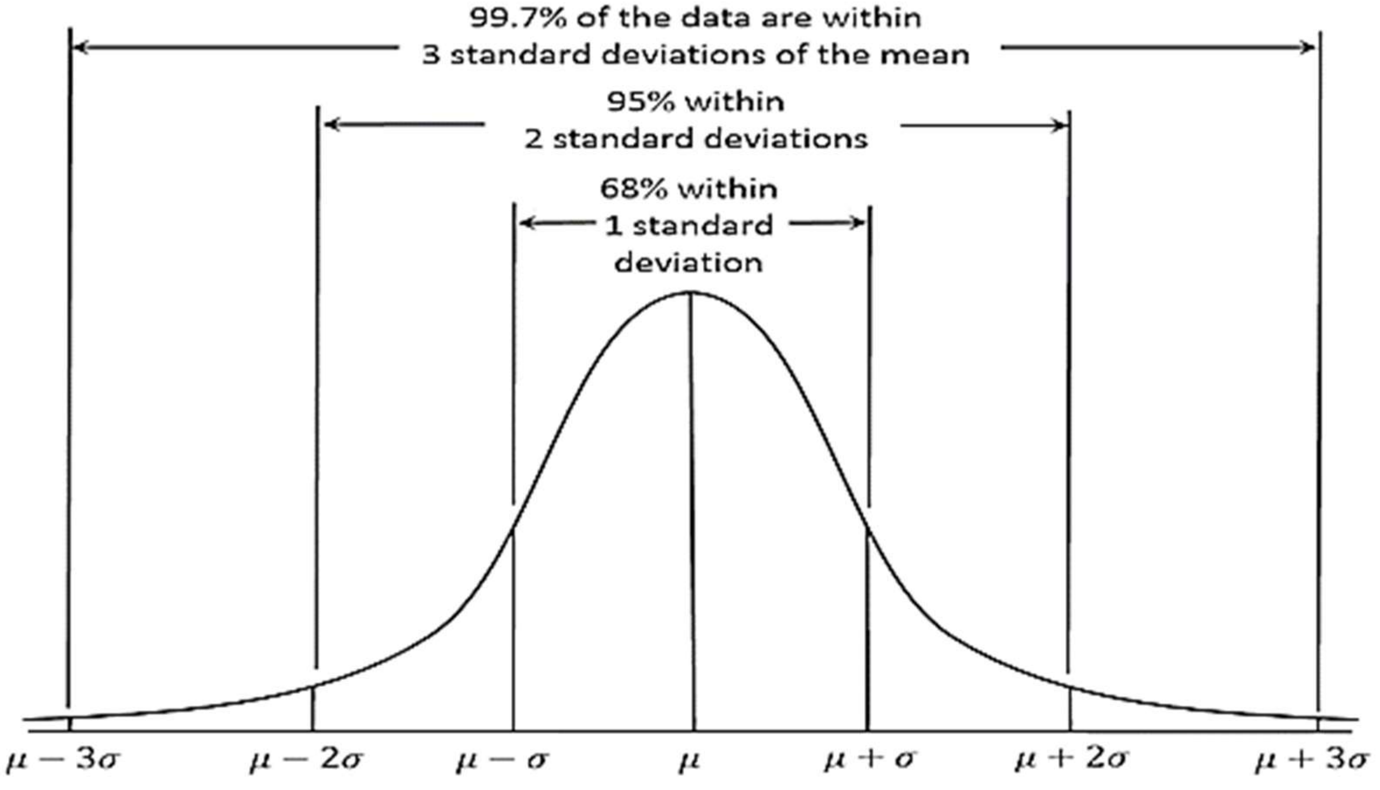
DiligentEdu©

Standard Deviation



DiligentEdu©

Standard Deviation

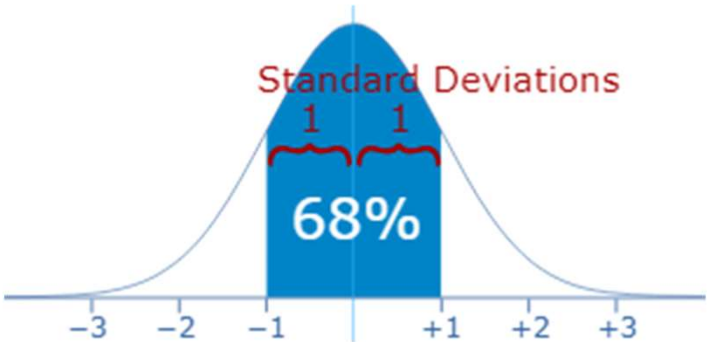


DiligentEdu©

Standard Deviation

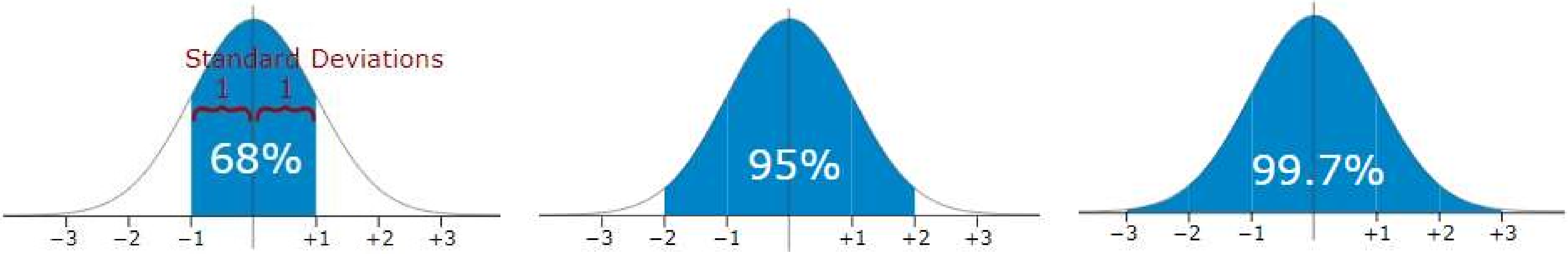
- **Standard Deviation (SD / σ)**

We can *expect* about **68%** of values to be within plus-or-minus **1** standard deviation.



DiligentEdu©

Standard Deviation



DiligentEdu©

Standard Deviation

- **Standard Deviation**

  Population & Sample Data

In the previous example – it was 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 …

DiligentEdu©

Standard Deviation

▪ Standard Deviation

Population & Sample Data

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.

DiligentEdu©

Standard Deviation

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 = **165** (to the nearest mm)

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

DiligentEdu©

Standard Deviation

▪ Standard Deviation (SD / σ)

Formulas

- The **Population** Standard Deviation:

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

- The **Sample** Standard Deviation:

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

The important change is to divide by **N-1** (instead of **N**) when calculating a **Sample Standard Deviation**.

DiligentEdu©

Standard Deviation

Main Sources / References

- https://www.varsitytutors.com
- https://www.mathsisfun.com
- https://www.math.net
- https://byjus.com
- https://www.statsdirect.com
- https://www.khanacademy.org
- Link - sph.bu.edu - Boston University
- Others …

DiligentEdu©

Thank you!

DiligentEdu©

www.DiligentEdu.in

www.YouTube.com/@DiligentEdu

DiligentEdu©